

WISTA

Wirtschaft und Statistik

Sonderrubrik

Malte Hansen

Sarah Redlich

Martin Beck, Luisa Baumgärtner,
Katja-Verena Bürk, Matthias Redecker

Martin Beck, Luisa Baumgärtner,
Katja-Verena Bürk, Matthias Redecker

Dr. Markus Zimmermann

Nadine Gühler,
Dr. Oda Schmalwasser

Tim Hochgürtel, Julia Weinmann

Jörg Feuerhake, Kerstin Lange,
Annelen Siegismund, Elsa Vigneau

Deutsche EU-Ratspräsidentschaft im zweiten Halbjahr 2020

Dynamische Preissetzung im Onlinehandel: zur langfristigen
Anwendung von automatisierter Preiserhebung

Web Scraping zur Gewinnung von Testdaten für
administrative Register

Einführung des EU-Unternehmensbegriffs:
Konzept und Umsetzung

Auswirkungen der Einführung des EU-Unternehmensbegriffs

Immaterielles Kapital und Produktivität im Verarbeitenden
Gewerbe

Anlagevermögen, Abschreibungen und Abgänge in den
Volkswirtschaftlichen Gesamtrechnungen

Haushalte in der Berichterstattung des Mikrozensus
ab 2020

Kodierung des Geburtsstaats in der Wanderungsstatistik

3 | 2020

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 hinterlegt.

INHALT

Englischer Titel geändert
am 12.11.2020

3	Editorial
4	Kennzahlen — Wichtige Indikatoren zur Corona-Krise
8	Kurznachrichten
11	Sonderrubrik Deutsche EU-Ratspräsidentschaft Deutsche EU-Ratspräsidentschaft im zweiten Halbjahr 2020 <i>German EU Council Presidency in the second half of 2020</i>
14	Malte Hansen Dynamische Preissetzung im Onlinehandel: zur langfristigen Anwendung von automatisierter Preiserhebung <i>Dynamic pricing in online trade: the long-term application of automated price collection</i>
24	Sarah Redlich Web Scraping zur Gewinnung von Testdaten für administrative Register <i>Web scraping used to obtain test data for administrative registers</i>
35	Martin Beck, Luisa Baumgärtner, Katja-Verena Bürk, Matthias Redecker Einführung des EU-Unternehmensbegriffs: Konzept und Umsetzung <i>New EU definition of enterprise as statistical unit: theory and implementation</i>
49	Martin Beck, Luisa Baumgärtner, Katja-Verena Bürk, Matthias Redecker Auswirkungen der Einführung des EU-Unternehmensbegriffs <i>Impact of introducing the EU enterprise concept</i>
61	Dr. Markus Zimmermann Immaterielles Kapital und Produktivität im Verarbeitenden Gewerbe <i>Intangible capital and productivity in manufacturing</i>

INHALT

76	Nadine Gühler, Dr. Oda Schmalwasser Anlagevermögen, Abschreibungen und Abgänge in den Volkswirtschaftlichen Gesamtrechnungen <i>Stock of fixed assets, consumption of fixed capital and retirements in national accounts</i>
89	Tim Hochgürtel, Julia Weinmann Haushalte in der Berichterstattung des Mikrozensus ab 2020 <i>Households in the microcensus from 2020</i>
98	Jörg Feuerhake, Kerstin Lange, Annelen Siegismund, Elsa Vigneau Kodierung des Geburtsstaats in der Wanderungsstatistik <i>Coding the country of birth in migration statistics</i>

EDITORIAL

Dr. Georg Thiel



LIEBE LESERIN, LIEBER LESER,

die Corona-Pandemie trifft Wirtschaft und Gesellschaft in einem bisher unbekannten Ausmaß. Die amtliche Statistik Deutschlands stellt vermehrt Daten am aktuellen Rand der Entwicklung zur Verfügung. Digitalisierung und neue Produktionswege ermöglichen dies zunehmend.

Unsere Sonderseite [▶ Corona-Statistiken](#) bündelt die verfügbaren aktuellen statistischen Informationen. Eine Auswahl relevanter Indikatoren finden Sie auf der erweiterten Kennzahlen-Übersicht in diesem Heft. Unser [▶ Krisenmonitor](#) vergleicht zudem die aktuellen Entwicklungen mit denen der Finanzmarktkrise 2009.

Erstmals informiert eine Sonderrubrik zur deutschen EU-Ratspräsidentschaft im zweiten Halbjahr 2020 über Inhalte, Prioritäten und Arbeitsprogramm der Ratsarbeitsgruppe Statistik.

Lesen Sie weitere interessante Aufsätze in dieser Ausgabe und bleiben Sie gesund!



Präsident des Statistischen Bundesamtes

Auswirkungen der COVID-19-Pandemie auf Wirtschaft und Gesellschaft

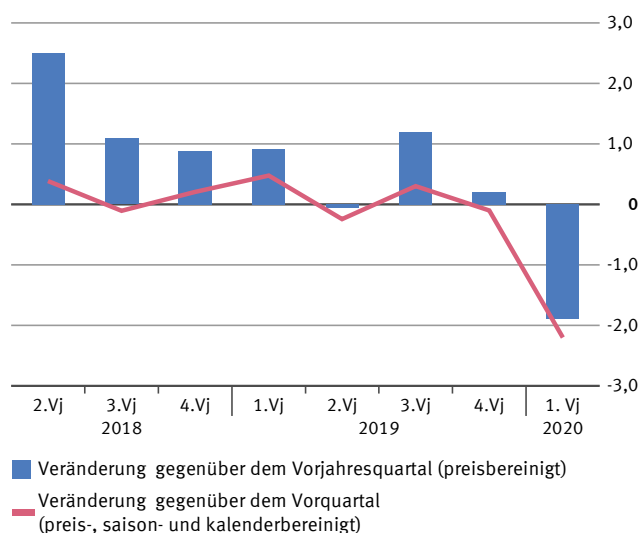
Die globale Ausbreitung des SARS-CoV-2-Virus und der dadurch verursachten Erkrankung COVID-19 im Frühjahr 2020 trifft Gesellschaft und Wirtschaft stark. Die langfristigen Folgen der Corona-Pandemie sind noch nicht überschaubar.

Die Kennzahlen auf dieser und den folgenden Seiten zeigen die Entwicklung einer Auswahl wichtiger Indikatoren, bei denen Auswirkungen sich bereits zeigen oder zu erwarten sind.

Umfangreiche weitere Informationen, gegliedert nach wirtschaftlichen Auswirkungen, einem Vergleich der Corona- mit der Finanzmarktkrise ([Krisenmonitor](#)) und gesellschaftlichen Auswirkungen, bietet die Sonderseite [Corona-Statistiken](#). Die Ergebnisse werden in Infografiken abgebildet und laufend aktualisiert.

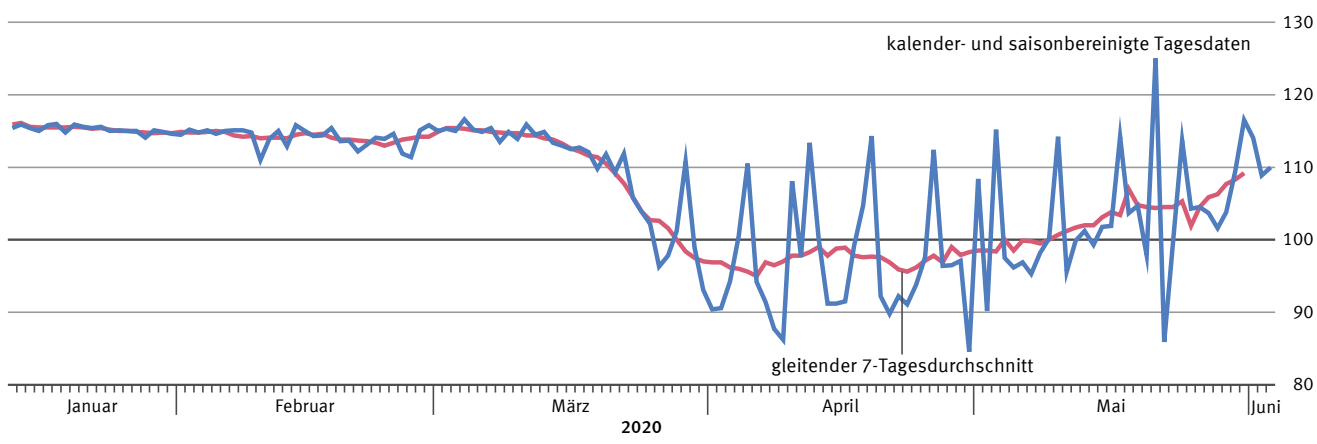
Als zusätzlichen Service verweist die Sonderseite auf die Pressemitteilungen des Statistischen Bundesamtes mit Fakten rund um Corona und informiert Unternehmen über die elektronischen Meldewege. Über den Menüpunkt „Weitere Angebote“ ist das wöchentlich aktualisierte [Dossier](#) des Statistischen Bundesamtes mit Statistiken zur COVID-19-Krise erreichbar, ebenso die thematischen Angebote der Bundesagentur für Arbeit, des Robert Koch-Instituts und weltweiter Datenanbieter.

Bruttoinlandsprodukt
in %



In einer [Online-Presskonferenz](#) Mitte Mai 2020 hat das Statistische Bundesamt erste vorläufige Ergebnisse für das Bruttoinlandsprodukt im ersten Quartal 2020 sowie weitere aktuelle Ergebnisse zur gesamtwirtschaftlichen Lage in Deutschland bis einschließlich April 2020 vorgestellt.

Lkw-Maut-Fahrleistungsindex ab Januar 2020
2015 = 100

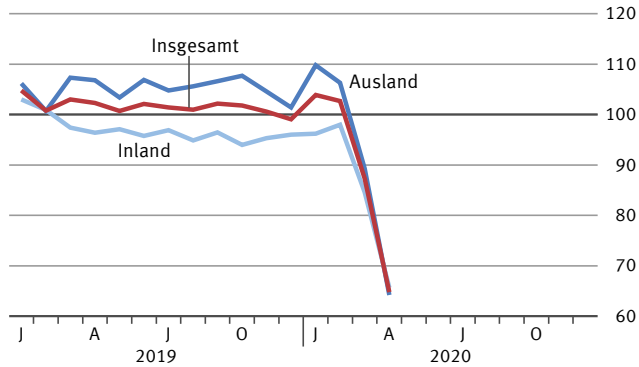


Quellen: Bundesamt für Güterverkehr, Deutsche Bundesbank, Statistisches Bundesamt

Stand: 9.6.2020

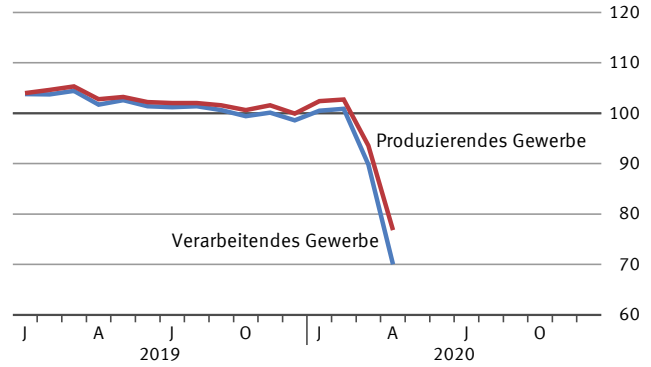
Kennzahlen – Wichtige Indikatoren zur Corona-Krise

Auftragseingang im Verarbeitenden Gewerbe
Volumenindex 2015 = 100



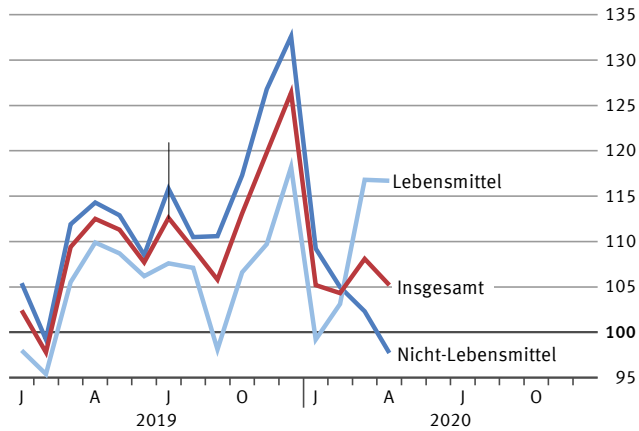
Kalender- und saisonbereinigter Wert nach dem Verfahren X13 JDemetra+. – Vorläufiges Ergebnis.

Produktion im Produzierenden und Verarbeitenden Gewerbe
Index 2015 = 100

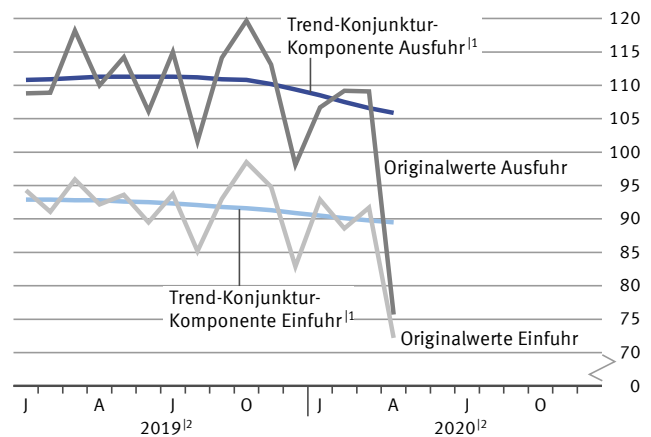


Kalender- und saisonbereinigter Wert nach dem Verfahren X13 JDemetra+. – Vorläufiges Ergebnis.

Umsatz im Einzelhandel
2015 = 100



Außenhandel
in Mrd. EUR



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

Stand: 9.6.2020

Kennzahlen – Wichtige Indikatoren zur Corona-Krise

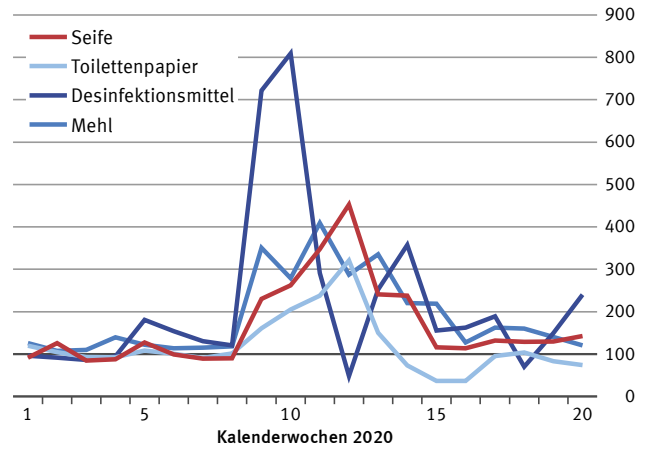
Verbraucherpreisindex 2015 = 100

2019		2020	
Januar	103,4	Januar	105,2
Februar	103,8	Februar	105,6
März	104,2	März	105,7
April	105,2	April	106,1
Mai	105,4		
Juni	105,7		
Juli	106,2		
August	106,0		
September	106,0		
Oktober	106,1		
November	105,3		
Dezember	105,8		

Veränderung
zum Vorjahresmonat

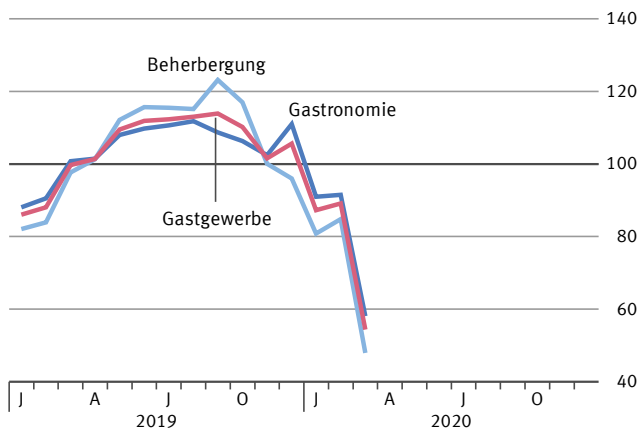
106,1 ↑ 0,9 %

Kaufverhalten: Absatz von ausgewählten Verbrauchsgütern 32. KW 2019 = 100

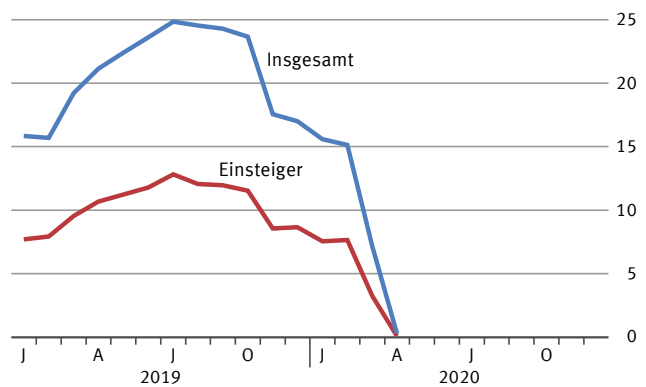


Experimentelle Daten (www.destatis.de/exdat). Ausgewertet wurden digital verfügbare Kassendaten (sogenannte Scannerdaten).

Umsatz des Gastgewerbes in konstanten Preisen (real), Originalwert, 2015 = 100

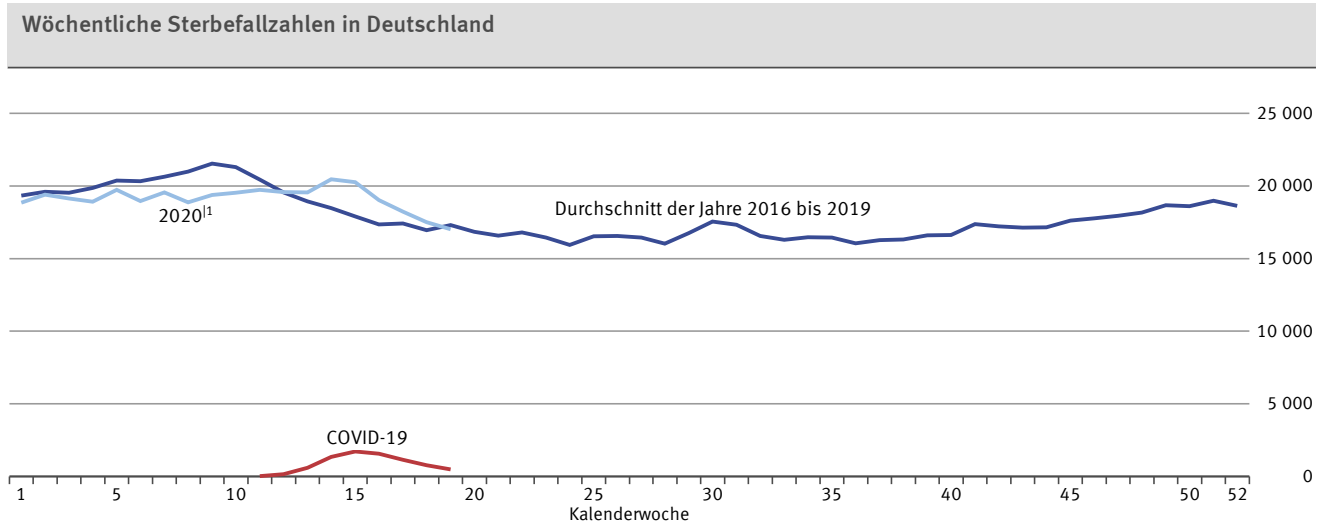
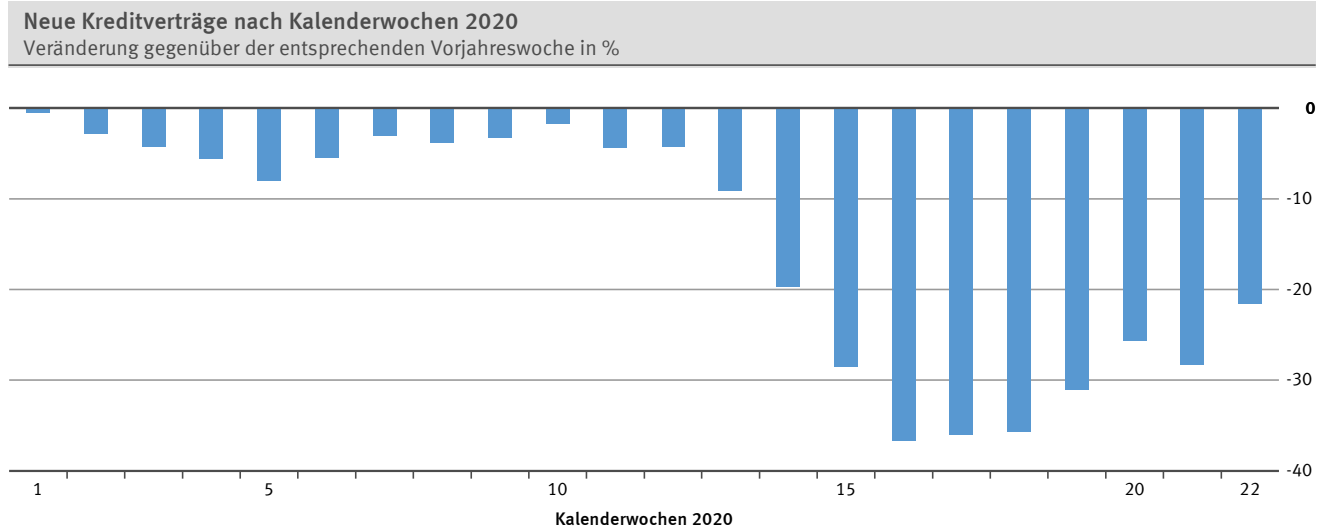


Flugpassagiere auf deutschen Flughäfen in Mill.



Stand: 9.6.2020

Kennzahlen – Wichtige Indikatoren zur Corona-Krise



¹ Sonderauswertung der vorläufigen Sterbefallzahlen.
Quelle: Statistisches Bundesamt (Sterbefallzahlen insgesamt), Robert Koch-Institut (COVID-19-Todesfälle)

Ankündigung in eigener Sache: Die nächste WISTA-Ausgabe wird über neue – zum Teil hier dargestellte – Statistiken und weiterentwickelte Methoden berichten, die die Datenbedarfe in Zeiten von Corona durch mehr Aktualität und Datenverfügbarkeit decken.

Stand: 9.6.2020

KURZNACHRICHTEN

INFORMATIONEN ZUR CORONA-KRISE

Neues Internetangebot: Krisenmonitor

Um die wirtschaftlichen Folgen der Corona-Pandemie mit der globalen Finanzmarkt- und Wirtschaftskrise 2008/2009 vergleichen zu können, hat das Statistische Bundesamt in seinem Internetauftritt wichtige Konjunkturindikatoren in einem „Krisenmonitor“ gebündelt. Das Grafikset ermöglicht den direkten Vergleich der konjunkturellen Entwicklung während beider Krisen. Der Beginn der jeweiligen Krise wird anhand des ersten Rezessionsquartals bestimmt: Der Startpunkt der Corona-Krise im ersten Quartal 2020 wird dem zweiten Quartal 2008 gegenübergestellt. Die Zahlen für das Jahr 2020 werden fortlaufend aktualisiert.

Der Krisenmonitor ergänzt die Sonderseite Corona-Statistiken (www.destatis.de/corona), die seit Anfang April 2020 statistische Informationen zu den wirtschaftlichen und gesellschaftlichen Folgen der Corona-Pandemie zusammenführt. Die Rohdaten der in diesen Angeboten enthaltenen Grafiken stehen im offenen CSV-Format zum Download bereit, auch die Grafiken selbst sind im PNG- und SVG-Format verfügbar.

Hintergrundinformationen zum Krisenmonitor und zu den dargestellten Konjunkturindikatoren runden dieses neue Internetangebot ab.

➤ www.destatis.de

Dossier mit Statistiken zur COVID-19-Pandemie

Ein umfassendes, vierzehntägig aktualisiertes Dossier des Statistischen Bundesamtes mit aktuellen Statistiken zur COVID-19-Krise steht neu auf der Sonderseite Corona-Statistiken zur Verfügung. Zu den Themengebieten „Gesellschaft und Umwelt“, „Konjunktur, Wirtschaft, Arbeitsmarkt“ sowie „Branchen und Unternehmen“ informieren Tabellen und Grafiken, zum Teil wird auch die Situation in Europa einbezogen. Hinweise und Links auf aktuelle Pressemitteilungen und die Datenquellen sowie auf weitere Webangebote ergänzen die Informationen.

➤ www.destatis.de

IN EIGENER SACHE

WISTA-Kennzahlen neu ausgerichtet

Ausgewählte statistische Kennzahlen wurden in dieser Zeitschrift bereits bisher in der Rubrik „Kennzahlen“ grafisch aufbereitet. Im Zuge der Corona-Krise haben wir das Angebot inhaltlich angepasst und im Umfang ausgeweitet: Auf zwei Doppelseiten finden Sie nun eine Zusammenstellung wichtiger Indikatoren mit direktem Bezug zur Corona-Krise.

Förderung von wissenschaftlichen Abschlussarbeiten

Zur Förderung von wissenschaftlichen Abschlussarbeiten an deutschen Hochschulen erstattet das Statistische Bundesamt seit dem 1. Juni 2020 Studierenden und Promovierenden die Kosten der Nutzung der Daten der Forschungsdatenzentren der Statistischen Ämter des Bundes und der Länder.

Eine Richtlinie auf einer Subsite des Statistik-Campus erläutert, welche Voraussetzungen dazu vorliegen müssen und wie eine Kostenerstattung beantragt werden kann. Die Richtlinie sieht unter anderem vor, dass dem Statistischen Bundesamt einfache Nutzungsrechte an den Ergebnissen der Abschlussarbeit eingeräumt werden, insbesondere die Veröffentlichung in einer seiner Publikationen.

➤ www.destatis.de

NEUERSCHEINUNGEN

Erklärvideo zum Schalenkonzept der Finanzstatistiken

Ausgaben und Einnahmen sowie Schulden und Finanzvermögen der öffentlichen Hand werden in den Finanzstatistiken für den „Öffentlichen Gesamthaushalt“ angegeben.

Was das ist, erklärt das sogenannte Schalenkonzept: Es liegt den Finanzstatistiken zugrunde und hilft, die Daten vergleichbar zu machen. Wie es funktioniert, erklärt ein neues Video des Statistischen Bundesamtes, das unter www.destatis.de/Schalenkonzept und auf YouTube zur Verfügung steht. Es erklärt, was ein Kernhaushalt und was ein Extrahaushalt ist, nennt Beispiele und schafft ein tieferes Verständnis für das, was finanzstatistisch zur „öffentlichen Hand“ gehört.

➤ www.destatis.de/Schalenkonzept

➤ www.youtube.com

NEUE INTERAKTIVE ANWENDUNGEN

Schuldnerberatungsatlas: Wie nah ist die nächste Schuldnerberatungsstelle?

Der interaktive Schuldnerberatungsatlas zeigt die Erreichbarkeit der nächsten Beratungsstelle von verschiedenen Standorten in ganz Deutschland. Die Erreichbarkeiten werden in Minuten angegeben und beziehen sich auf die Fahrtzeit in einem Pkw bei ungestörter Verkehrslage.

➤ schuldnerberatungsatlas.destatis.de/

Energieverbrauch in der Industrie

Diese interaktive Karte stellt den Energieverbrauch der Industrie im Jahr 2018 nach Kreisen dar. Die Detailinformationen zu den Kreisen zeigen zusätzlich eine Aufteilung der Energieträger in Kohle, Heizöl, Erdgas, Erneuerbare Energien (zum Beispiel Photovoltaik), Strom, Wärme und sonstige Energieträger (andere Mineralöl-erzeugnisse, sonstige Gase und Abfälle).

➤ service.destatis.de

DEUTSCHE EU-RATSPRÄSIDENTSCHAFT IM ZWEITEN HALBJAHR 2020

📌 **Schlüsselwörter:** EU-Ratspräsidentschaft – Ratsarbeitsgruppe Statistik – Trio-Präsidentschaft – Europäisches Statistisches System

ZUSAMMENFASSUNG

Deutschland übernimmt am 1. Juli 2020 für ein halbes Jahr den Vorsitz im Rat der Europäischen Union (EU). Während der deutschen EU-Ratspräsidentschaft wird das Statistische Bundesamt für die Bundesregierung die Sitzungen der Ratsarbeitsgruppe Statistik vorbereiten und leiten. Darüber hinaus verfolgt das Statistische Bundesamt das Ziel, das europäische Statistikprogramm aktiv mitzugestalten. In dieser und den kommenden Ausgaben dieser Zeitschrift informiert eine Sonderrubrik zur deutschen EU-Ratspräsidentschaft 2020 über deren Inhalte, Prioritäten und das Arbeitsprogramm.

📌 **Keywords:** German EU Council Presidency – Council Working Party on Statistics – Trio Presidency – European Statistical System

ABSTRACT

On 1 July 2020, Germany will take over the presidency of the Council of the European Union (EU) for a period of six months. During the German EU Council Presidency, the Federal Statistical Office will prepare and chair the meetings of the Council Working Party on Statistics on behalf of the Federal Government. In addition to that, the Federal Statistical Office intends to play a part in shaping the European Statistics Programme. Starting with the present issue of this journal, a special section on the German EU Council Presidency 2020 will provide information on its content, priorities and the work programme.

Deutsche EU-Ratspräsidentschaft im Bereich Statistik

Die Ratsarbeitsgruppe Statistik ist das zentrale Gremium für die Beratung von EU-Rechtssetzungsakten auf dem Gebiet der Statistik. Sie ist vorbereitend für den Rat der Europäischen Union tätig, indem sie die EU-Gesetze im Bereich der amtlichen Statistik verhandelt und diese dem Rat dann zur Annahme vorlegt. Der Rat der EU wiederum übt zusammen mit dem Europäischen Parlament die Rechtsetzung der Europäischen Union aus. Er repräsentiert die Regierungen der EU-Mitgliedstaaten.

Der Vorsitz im Rat der EU (EU-Ratspräsidentschaft) rotiert nach einem gleichberechtigten Turnus zwischen den EU-Mitgliedstaaten. Alle sechs Monate wechselt die Ratspräsidentschaft zwischen den EU-Mitgliedsländern gemäß einer festgelegten Reihenfolge. Im ersten Halbjahr 2020 tagt er unter dem Vorsitz Kroatiens und im zweiten Halbjahr 2020 unter dem Vorsitz Deutschlands. Anschließend folgen Portugal (1. Halbjahr 2021) und Slowenien (2. Halbjahr 2021). Den Vorsitz in der Ratsarbeitsgruppe Statistik während der deutschen EU-Ratspräsidentschaft wird der Präsident des Statistischen Bundesamtes, Dr. Georg Thiel, übernehmen. Der Vorsitzende ist dafür verantwortlich, die Beratungen des Rates über EU-Rechtsvorschriften voranzubringen und für die Kontinuität der Agenda der EU, den ordnungsgemäßen Verlauf der Gesetzgebungsverfahren und die Zusammenarbeit der Mitgliedstaaten zu sorgen.

Ziel des Statistischen Bundesamtes ist es, im Zeitraum der deutschen Ratspräsidentschaft die Gesetzesvorhaben im Bereich der Unionsstatistik möglichst effizient voranzutreiben. Die zu verhandelnden Gesetzesinitiativen werden vom Statistischen Amt der Europäischen Union, Eurostat, als zuständiger Generaldirektion der Europäischen Kommission vorgelegt. Über die Arbeit in der Ratsarbeitsgruppe Statistik hinaus verfolgt das Statistische Bundesamt das Ziel, aktiv das europäische Statistikprogramm mitzugestalten. Nur durch grenzüberschreitende Zusammenarbeit und in einem kohärenten Statistiksistem kann die amtliche Statistik belastbare Antworten auf zukunftsweisende Fragen geben und damit faktenbasierte Entscheidungen in der EU unterstützen.

EU-Ratspräsidentschaft vor dem Hintergrund der Corona-Pandemie

Die Corona-Pandemie stellt eine außerordentliche Herausforderung für die EU dar und erzeugt neuen Informationsbedarf. Zur Bewältigung der Pandemie und ihrer Folgen sind aktuelle statistische Informationen fundamental. Qualitativ hochwertige Statistiken sind eine wichtige Grundlage für faktenbasierte Entscheidungen und politische Maßnahmen in der EU – gerade in Krisenzeiten. Auch die Relevanz der amtlichen Statistik für nachhaltige aber auch zeitnahe Entscheidungsfindung ist durch die Krise noch stärker als bisher in den Fokus gerückt. Vor diesem Hintergrund gilt es für die amtliche Statistik, Lehren aus der Krise zu ziehen, um auf künftige ähnliche Situationen angemessen vorbereitet zu sein. Die Erkenntnisse aus der Krise bilden daher die Grundlage für die Schwerpunkte und das Programm der deutschen EU-Ratspräsidentschaft im Bereich Statistik.

So gilt es beispielsweise, die Statistikproduktion sowie Koordinations- und Kommunikationsprozesse anzupassen und damit eine effiziente und flexible Reaktion auf Krisensituationen zu ermöglichen. Dazu gehört etwa, die Chancen der Digitalisierung voll auszuschöpfen, um Informationen auch kurzfristig bereitstellen zu können. Wichtig ist dabei, Datenbedarfe zur Messung und Gegensteuerung der Pandemie zu decken. Die deutsche EU-Ratspräsidentschaft im Bereich Statistik wird sich daher im Europäischen Statistischen System für einen koordinierten Ansatz einsetzen, um die Bereitstellung verlässlicher Statistiken von hoher Priorität sicherzustellen.

Trio-Präsidentschaft mit Portugal und Slowenien

Zusammen mit Portugal und Slowenien bildet Deutschland eine sogenannte Trio-Präsidentschaft¹. Mit der Zusammenarbeit in Trio-Präsidentschaften wird das Ziel verfolgt, trotz des turnusmäßigen halbjährlichen Vorsitzwechsels Kontinuität zu schaffen und den erzielten Ergebnissen mehr Nachhaltigkeit zu verleihen. Als Trio werden Deutschland, Portugal und Slowenien für die Zeit ihrer 18-monatigen Trio-Präsidentschaft eng zusam-

¹ Jeweils drei Mitgliedstaaten, die formal nacheinander die EU-Ratspräsidentschaft innehaben, bilden eine Trio-Präsidentschaft.

menarbeiten. Dafür haben die Trio-Partner ein gemeinsames Arbeitsprogramm verfasst, in dessen Mittelpunkt die Bereitstellung hochwertiger europäischer Statistiken steht. Hochwertige, objektive und verlässliche Daten beschreiben die Entwicklungen in Gesellschaft, Wirtschaft und Umwelt präzise und sind daher von zentraler Bedeutung für fundierte Entscheidungen. Damit die europäische amtliche Statistik auf neue, durch Globalisierung und Digitalisierung entstehende Nutzerbedarfe angemessen reagieren kann, unterstützt das Trio die Modernisierung des Europäischen Statistischen Systems. Hierfür sollen neue Datenquellen und Technologien stärker genutzt sowie innovative Methoden und Werkzeuge entwickelt werden. Gleichzeitig setzt sich die Trio-Präsidentschaft dafür ein, dass die zentralen Werte der amtlichen Statistik nicht in den Hintergrund rücken: Daten sollen transparent, verfügbar und verlässlich sein, die statistische Geheimhaltung soll sichergestellt werden.


Präsidentschaftsteam und weitere Informationen

Für die Vorbereitung und Durchführung der deutschen EU-Ratspräsidentschaft im Bereich Statistik sorgt das Präsidentschaftsteam des Statistischen Bundesamtes und des Bundesministeriums des Innern, für Bau und Heimat:

- › Dr. Georg Thiel – Vorsitzender der Ratsarbeitsgruppe Statistik (Statistisches Bundesamt)
- › Dr. Daniel Vorgrimler – Stellvertretender Vorsitzender der Ratsarbeitsgruppe Statistik (Statistisches Bundesamt)
- › Yvonne Vatter – Koordinatorin des Präsidentschaftsteams (Statistisches Bundesamt)
- › Annette Dürkop (Bundesministerium des Innern, für Bau und Heimat)
- › Sabrina Töpfer (Statistisches Bundesamt)
- › Christina Schliffka (Statistisches Bundesamt)
- › Nadine Stechmann (Statistisches Bundesamt)
- › Andreas Schmitt (Statistisches Bundesamt)

Zu erreichen ist das Team der deutschen EU-Ratspräsidentschaft im Bereich Statistik unter der E-Mail-Adresse eu2020@destatis.de.

Ab dem 1. Juli 2020 wird die speziell eingerichtete Internetseite www.destatis.de/eu2020 alle Informationen zur deutschen EU-Ratspräsidentschaft im Bereich Statistik bereitstellen. Dort stehen dann alle wichtigen Dokumente der Sitzungen der Ratsarbeitsgruppe Statistik zur Verfügung. Darüber hinaus bietet die Seite einen Überblick über die Prioritäten der deutschen EU-Ratspräsidentschaft im Bereich Statistik sowie Informationen zu geplanten Tagungen und Veranstaltungen während der deutschen EU-Ratspräsidentschaft.

Auch in den kommenden Ausgaben dieser Zeitschrift informiert die Sonderrubrik Deutsche EU-Ratspräsidentschaft über Inhalte, Prioritäten und Arbeitsprogramm der Ratsarbeitsgruppe Statistik. 



Malte Hansen

ist Entwicklungsökonom mit Abschlüssen der Universitäten Göttingen und Stellenbosch (Südafrika). Seit September 2017 arbeitet er im Statistischen Bundesamt, derzeit als wissenschaftlicher Mitarbeiter im Referat „Verbraucherpreise“. Er untersucht vorwiegend die dynamische Preissetzung im Onlinehandel zur Weiterentwicklung des Verbraucherpreisindex.

DYNAMISCHE PREISSETZUNG IM ONLINEHANDEL: ZUR LANGFRISTIGEN ANWENDUNG VON AUTOMATISIERTER PREISERHEBUNG

Malte Hansen

➤ **Schlüsselwörter:** Verbraucherpreisindex – dynamische Preissetzung – Web Scraping – automatisierte Preiserhebung – Onlinehandel

ZUSAMMENFASSUNG

Durch die Digitalisierung im Onlinehandel ist seit einigen Jahren ein neues Preissetzungsverhalten der Anbieter zu beobachten: dynamische Preissetzung. Dabei verwenden Anbieter maschinelle Algorithmen, um Preise von Produkten laufend an veränderte Marktsituationen anzupassen. Für die Verbraucherpreisstatistik stellt sich damit die Frage, inwieweit Anpassungen in der Datenerhebung für den Verbraucherpreisindex erforderlich sind. Das Statistische Bundesamt arbeitet deshalb seit einiger Zeit an der Automatisierung der Preiserhebung mittels Web Scraping. In diesem Zusammenhang wurde eine Studie durchgeführt, um optimale Handlungsempfehlungen für die langfristige Nutzung des Web-Scraping-Programms zu ermitteln. Der Aufsatz stellt das methodische Vorgehen und die Ergebnisse dieser Studie vor.

➤ **Keywords:** consumer price index – dynamic pricing – web scraping – automated price collection – online trade

ABSTRACT

As the result of digitalisation in online trade, a new price setting behaviour has been observed for several years: dynamic pricing. Vendors use automatic algorithms in order to adapt product prices continuously to changing market situations. Regarding consumer price statistics, the question arises whether adjustments are required in the process of collecting data for the consumer price index. For several years, the Federal Statistical Office has been working on the automation of price collection, using web scraping. A study has been conducted to provide recommendations for the optimal and long-term use of the web scraping program. This article presents the methodical approach and the findings of the study.

1

Einleitung

Seit vielen Jahren ist die im Tagesablauf variable Preissetzung von Treibstoffen an Tankstellen bekannt: Zu bestimmten Stoßzeiten sind die Preise höher als zu Standardzeiten. Würde an einem Tag beziehungsweise in einem Monat der Standardpreis erhoben und bei der nächsten Erhebung der erhöhte Preis zur Stoßzeit, würde die Preisstatistik fälschlicherweise einen Preisanstieg melden. Eine ähnliche Problematik ergibt sich durch die dynamische Preissetzung im Onlinehandel. Um diese und weitere Entwicklungen sowie ihre Auswirkungen auf die Preisstatistik zu untersuchen, führte das Statistische Bundesamt eine umfangreiche Studie durch. Deren Aufbau und Ergebnisse stellt der folgende Beitrag vor.

Durch die zunehmende Digitalisierung verlagert sich das Kaufverhalten in Deutschland verstärkt zum Onlinehandel. Die Wirtschaftsauskunftei Creditreform Boniversum GmbH (Boniversum) und der Bundesverband E-Commerce und Versandhandel Deutschland (bevh) führten im Jahr 2017 eine Verbraucherumfrage durch. Demnach stieg der Umsatz im Onlinehandel im Jahr 2016 um über 12 % gegenüber dem Vorjahr auf rund 52,7 Milliarden Euro. Sein Anteil am Gesamtumsatz im Einzelhandel belief sich damit auf 10,4 % (Statistisches Bundesamt, 2020). Im Jahr 2017 erhöhte sich dieser Anteil auf 11,3 %. Und immer mehr Käuferinnen und Käufer erwerben vermehrt im Internet die für den Verbraucherpreisindex relevanten Artikel. So berichtet der Spiegel (Wahnbaek, 2019), dass im Vorweihnachtsgeschäft 2019 mehr als ein Drittel der Deutschen beabsichtigten, ihre Geschenkkäufe online zu tätigen.

Der Verbraucherpreisindex (VPI) wird monatlich berechnet und veröffentlicht. Seine Veränderungsrate zum Vormonat gibt die durchschnittliche Preisentwicklung aller Konsumausgaben von privaten Haushalten im Inland an. Diese Veränderungsrate wird allgemein als Inflationsrate bezeichnet. Der Verbraucherpreisindex umfasst etwa 300 000 Artikel, die repräsentativ für alle in Deutschland zu erwerbenden Konsumgüter stehen (Statistisches Bundesamt, 2018). Etwa 10 000 dieser Artikel repräsentieren den Onlinehandel. Für die korrekte Berechnung der Entwicklung des Verbraucherpreisindex wird je Artikel mindestens einmal im Monat

ein Preis erhoben. Dabei wird angenommen, dass dieser Preis repräsentativ für den Monatspreis des Artikels ist. Vor dem Hintergrund der dynamischen Preissetzung im Onlinehandel muss diese Annahme hinterfragt und die Datenerhebung gegebenenfalls angepasst werden.

Darüber hinaus kann eine angepasste Datenerhebung die Genauigkeit der Preisreihen sowie ihren Informationsgehalt erhöhen. Die adäquate Darstellung der monatlichen Preisentwicklung durch den Verbraucherpreisindex bleibt damit nicht nur gewährleistet, sondern wird sogar noch verbessert. Für die Studie zur dynamischen Preissetzung im Onlinehandel wurden Preise über einen Zeitraum von zwölf Monaten erfasst. Mit 8 624 Artikeln umfasst sie einen Großteil der im Verbraucherpreisindex hinterlegten 10 000 Artikel. Die Erkenntnisse führen zu Handlungsempfehlungen für die langfristige Nutzung des Web-Scraping-Programms, mit dem die automatisierte Datenerhebung durchgeführt wird.¹

Neben dem Statistischen Bundesamt experimentieren auch andere nationale Statistikämter in Europa mit Web-Scraping-Daten oder haben diese bereits in ihre VPI-Berechnungen integriert. Das Statistikamt des Vereinigten Königreichs (ONS) benutzt Web-Scraping-Daten für mehrere Produktgruppen (Flower/Karachalias, 2019). Für die meisten Produktgruppen werden wöchentliche Erhebungen durchgeführt. Für einige Ausnahmen finden tägliche Erhebungen statt. Auch für den belgischen Verbraucherpreisindex werden die Preise wöchentlich oder täglich erhoben (van Loon/Roels, 2018).

Der weitere Aufsatz gliedert sich wie folgt: In Kapitel 2 wird die dynamische Preissetzung definiert und der Aufbau der Studie vorgestellt. Im Anschluss werden die Rohdaten auf ihre Verwertbarkeit analysiert und die angewandten Aufbereitungsschritte erläutert. Kapitel 3 beschreibt die Ergebnisse hinsichtlich der Anzahl der Preisänderungen, deren zeitliches Auftreten und gibt Handlungsempfehlungen zur automatisierten Preiserhebung. Dabei wird insbesondere auf Kalender- und Saisoneffekte eingegangen. Zudem werden die Ergebnisse zusammengefasst und die Umsetzbarkeit ihrer Implikationen diskutiert. Die Ausführungen schließen mit einem Fazit und Ausblick in Kapitel 4.

¹ Für eine detaillierte Beschreibung der automatisierten Datenerhebung mit dem Web-Scraping-Programm und dessen Funktionsweisen siehe Bladow/Seeger (2019).

2

Methodisches Vorgehen

2.1 Definition und Aufbau der Studie

Dynamische Preissetzung wird in dieser Studie wie folgt definiert: Dynamische Preissetzung liegt bei einem Artikel vor, wenn 3 % oder mehr der Preise eines Artikels über dessen Betrachtungszeitraum Änderungen zeigen. Dabei wurde von 8736 erhobenen Preisen ausgegangen, sofern der Preis eines Artikels über alle 52 Wochen eines kompletten Kalenderjahrs zu jedem Tag und jeder Stunde erhoben werden konnte. Ändern sich 3 % dieser Preise, liegen 262 Preisänderungen vor. Das ergibt – auf einzelne Wochen umgerechnet – fünf Preisänderungen im arithmetischen Mittel.

Einige Produktgruppen stehen zu Sonder- und Feiertagen im Fokus der Konsumenten und werden häufiger durch Rabattaktionen im Preis heruntergesetzt. Dazu gehören beispielsweise elektronische Unterhaltungsartikel rund um den Black Friday, den Tag nach Thanksgiving in den Vereinigten Staaten. Dadurch zählen diese Produktgruppen relativ viele Preisänderungen zu bestimmten Zeiten im Jahr. Im restlichen Verlauf des Jahres sind die Preiserien jedoch stabil. Solch eine Dynamik soll nicht als dynamische Preissetzung verzeichnet werden, da in der manuellen Preiserhebung die Sonder- und Feiertage schon speziell betrachtet werden. Mit der 3-%-Regel wird versucht, dies abzudecken.

Neue Technologien wie der Onlinehandel vereinfachen das größtenteils automatisierte Preissetzungsverhalten der Warenanbieter. Die dynamische Preissetzung ist damit zu einem Phänomen geworden, welches Anpassungen in der Preiserhebung erfordern kann. Vor diesem Hintergrund wird in der Preisstatistik an der Einführung eines automatisierten Preiserhebungsprogramms gearbeitet. Durch die Untersuchung des Preissetzungsverhaltens kann festgestellt werden, welche Anbieter ihre Artikelpreise dynamisch setzen. Deren Preise sollten in erhöhter Frequenz abgefragt werden. Ziel dieser Studie ist es, diese Anbieter zu identifizieren und für sie entsprechende Erhebungsintervalle zu bestimmen. Darüber hinaus soll das Web-Scraping-Programm durch die hier gewonnenen Erkenntnisse zu bestimmten Tages-

und Wochenzeiten auch langfristig statistisch validierte Artikelpreise für den Verbraucherpreisindex liefern können. Für ein möglichst detailliertes Bild war deshalb die Erhebungsfrequenz, anders als bei der traditionellen, manuellen Preiserhebung, von monatlich auf stündlich erhöht.

Für die stündliche Erhebung wurde in der Inputdatenbank des Web-Scraping-Programms für jeden Konsumartikel genau eine URL (Uniform Resource Locator) zugewiesen.¹² Bei jedem einzelnen Scraping-Vorgang passiert dann Folgendes: Automatisiert wird an jedem Tag und in jeder Stunde je Artikel eine URL in einem Web-Browser-Fenster abgefragt. Wird die URL erreicht, liest das Web-Scraping-Programm den jeweiligen Artikelpreis, dessen Name und die Artikelnummer heraus. Beim Artikelpreis wurde zwischen Normal- und Sonderpreis unterschieden. War für einen Artikel ein reduzierter Sonderpreis ausgewiesen, so wurde versucht, den entsprechenden Normalpreis ebenfalls zu erfassen, um anschließend eine Plausibilitätsprüfung durchzuführen.¹³

Ein potenzielles Problem bei der Online-Datenerhebung von Artikelpreisen ist das Auftreten von individualisierter Preissetzung. Dabei versuchen Anbieter, den einzelnen, potenziellen Kundinnen und Kunden aufgrund ihres Verhaltens auf der Internetseite einen möglichst hohen, noch von ihnen akzeptierten und individuellen Preis anzuzeigen. Das Verhalten der Kundinnen und Kunden ergibt sich unter anderem aus der Anzahl der Seitenaufrufe beim jeweiligen Onlinehändler, der Verweildauer auf der jeweiligen Seite und das Aufrufen der einzelnen Artikel beim jeweiligen Händler. Um diesen Effekt für die Studie ausschließen zu können, wurden drei Maßnahmen ergriffen: 1. Die Anfragen an die Artikelseiten wurden nicht von einer einzelnen IP(Internet-Protocol)-Adresse

- 2 Auf diese Weise konnte das Programm die einzelnen Artikelseiten direkt ansteuern und musste nicht durch die Internetseiten der einzelnen Onlinehändler navigieren. Ansonsten hätten die einzelnen Scraping-Vorgänge für eine stündliche Preiserhebung zu lange gedauert. Der Einsatz des Web-Scraping-Programms wurde somit vollständig auf dessen Geschwindigkeit optimiert.
- 3 Internetseiten sind meistens im sogenannten HTML-Format aufgebaut. Die einzelnen Elemente einer Internetseite im HTML-Format lassen sich über sogenannte XPath's genau ansteuern. Ein XPath gibt die Position und Funktion eines Objekts innerhalb eines HTML-Dokuments an. So ist zum Beispiel ein Artikelpreis durch seinen eigenen XPath als ein bestimmtes Objekt definiert und lässt sich auf der Internetseite genau eingrenzen und auslesen. Zur Lokalisierung der Positionen auf der Internetseite mussten für jeden Onlinehändler XPath's in der Inputdatenbank hinterlegt werden. Die XPath's konnten für alle Artikel eines Onlinehändlers genutzt werden, da die einzelnen Artikelseiten je Onlinehändler gleich aufgebaut sind.

gestartet, sondern es wurde zu jeder Stunde eine neue IP aus einer Liste von mehreren IPs zufällig ausgewählt. Auf diese Weise erschien das Web-Scraping-Programm nicht als einzelner Kunde mit einer IP-Adresse. 2. Vor jedem stündlichen Scraping-Vorgang wurden die Cookies im Internetbrowser gelöscht, um hinterlegte Daten zu Seitenaufrufen zu entfernen. 3. Die einzelnen URLs wurden zu jeder Stunde in zufälliger Reihenfolge aufgerufen. Der Internetbrowser öffnete mehrere Browserfenster, auf die die Seitenaufrufe aufgeteilt waren.

2.2 Datenanalyse und Datenaufbereitung

Der Rohdatensatz umfasst mehr als 90% der für den Verbraucherpreisindex relevanten 10 000 Artikel. Die Artikel werden zur Berechnung des Verbraucherpreisindex zu Produktgruppen der Classification of Individual Consumption by Purpose (COICOP) zusammengefasst. Die Artikel sind auf 324 COICOP-10-Steller, die unterste Gliederungsebene, aufgeteilt. Der Verbraucherpreisindex umfasst insgesamt etwas mehr als 600 COICOP-10-Steller (Statistisches Bundesamt, 2018). In der VPI-Datenbank sind rund 300 Onlinehändler vertreten. Zu 273 Onlinehändlern wurde mindestens ein Preis erhoben. Der Erhebungszeitraum sollte das komplette Jahr 2019 umfassen. Die Preiserhebung begann bereits Ende 2018, damit der reibungslose Ablauf der Datenerhebung möglichst zu Beginn des Jahres 2019 erreicht werden konnte. Aufgrund der dennoch eingetretenen

Anlaufschwierigkeiten in den ersten Wochen 2019 wurden auch Preise für die erste Januarwoche 2020 erhoben, um die Preisentwicklung nach Weihnachten und Neujahr mit besserer Datenlage analysieren zu können.

▸ Tabelle 1

Die erhobenen Daten sind Rohdaten, die aufbereitet und plausibilisiert werden müssen. Für die Studie wurden zwei Gruppen von Problemfällen identifiziert, die getrennt zu beachten sind: zu editierende Preise und unplausible Preisänderungen. Zu editierende Preise sind zum Beispiel solche, bei denen festgestellt wurde, dass das Web-Scraping-Programm einen Onlinepreis fehlerhaft ausgelesen hat. Dieser Fehler ließ sich durch die Korrektur der Kommastelle beheben. In anderen Fällen lag der Normalpreis unter dem Sonderpreis, sodass der Normalpreis anstatt des Sonderpreises eingesetzt wurde.

Unplausible Preisänderungen wurden durch zwei Berechnungsmethoden identifiziert. Für die erste Methode wurden für jeden Artikel prozentuale Preisänderungen zum vorigen Preis ermittelt. Unplausible Preisänderungen sind in dieser Studie definiert als Änderungen, die um 400 % über oder um 80 % unter dem vorangegangenen Preis liegen. Für die zweite Berechnungsmethode wurde für jede Preisreihe ein 95 %-Konfidenzintervall um das arithmetische Mittel der Preisänderungen über den gesamten Zeitraum errechnet. Liegt eine Preisänderung um 400 % über der Obergrenze oder um 80 % unter der Untergrenze des Konfidenzintervalls, wird diese als unplausibel markiert und aussortiert.

Nach der Datenaufbereitung waren für die verbliebenen 8 624 unterschiedlichen Artikel fast 42 Millionen Artikelpreise von 241 Onlinehändlern verwendbar. Die Preise wurden zur Indexberechnung auf 324 COICOP-10-Steller aufgeteilt. Zur Untersuchung des Preissetzungsverhaltens wurden ausschließlich erhobene Preise betrachtet. Fehlende Preise wurden nicht imputiert, um die Ergebnisse nicht zu verfälschen.⁴ Aufgrund der vorliegenden Massendaten stellen fehlende Werte mit einem Anteil von über 25 % keinen Informationsverlust dar, der die Analyse der Zeitreihen beeinträchtigen würde. Die Ausfälle der Preise sind zufällig und verzerren deshalb nicht die vorhandenen Daten des Parameters von Interesse:

Tabelle 1

Erhobene Preise nach dem Erhebungsmonat

	Anzahl der Preise
Dezember 2018	342 562
Januar 2019	2 289 648
Februar 2019	2 668 073
März 2019	3 442 491
April 2019	2 697 063
Mai 2019	3 571 108
Juni 2019	3 214 571
Juli 2019	3 693 373
August 2019	4 025 614
September 2019	3 789 103
Oktober 2019	3 836 901
November 2019	3 473 676
Dezember 2019	3 457 040
Januar 2020	1 167 409
Insgesamt	41 668 632

4 In dieser Analyse werden vorrangig Veränderungen und Veränderungsraten von Artikelpreisen betrachtet. Je nach Wahl der Imputationsmethode wären diese Parameter „mutwillig“ verändert worden und hätten keine rein objektiven Beobachtungen mehr dargestellt.

den Artikelpreis. Generell gibt es keine Regel, ab welchem Anteil fehlender Werte eine Imputation notwendig ist und lückenhafte Zeitreihen nicht berücksichtigt werden dürfen (Little/Rubin, 2002).

↘ **Tabelle 2** zeigt den Aufbau des Datensatzes. Zu sehen sind unter anderem die Anzahl der plausiblen und unplausiblen

Preise und deren Anteile an allen Beobachtungen. Sowohl die berücksichtigten, editierten Preise als auch die aussortierten, unplausiblen Preise haben mit 1,25 % beziehungsweise weniger als 1 % jeweils nur einen sehr geringen Anteil an den insgesamt erfassten Preisen. Aufgrund der großen Datenmengen wurde davon abgesehen, die Abschneidegrenzen für unplausible Preisänderungen enger zu definieren – es war nicht abzuschätzen, wie viele Preise zu Unrecht durch diese Maßnahme aussortiert worden wären.¹⁵

3

Ergebnisse

Die folgende Ergebnisanalyse berücksichtigt zwei Aspekte: die Anzahl der Preisänderungen (Abschnitt 3.1) und das zeitliche Auftreten der Preisänderungen (Abschnitt 3.2).

3.1 Anzahl der Preisänderungen

↘ **Grafik 1** gruppiert die Onlinehändler über den gesamten Erhebungszeitraum in sechs Volatilitätsgruppen nach Anzahl der anteiligen Preisänderungen. Volatili-

Tabelle 2
Übersicht über die Datensatzstruktur

	Beobachtungen	Anteil an den Rohpreisen	Filter/Kommentar
	1 000	% ¹	
Alle Scraping-Vorgänge	rund 63 800		
Rohdaten	57 523		ohne leere Zellen beziehungsweise fehlende URL
Fehlende Preise	15 087		nicht berücksichtigt
Verbleibende Rohpreise	42 435	100	
Plausible und editierte Preise	41 669	98,19	verwendeter Datensatz; um unplausible Preise und Ausreißer bereinigt
Editierte Preise	528	1,25	
Unplausible Preisänderungen	260	0,63	nicht berücksichtigt

¹ Differenz zu 100 % rundungsbedingt.

tätsgruppe 1 umfasst die Händler, die überhaupt keine Preisänderungen vornahmen, während die Händler in Volatilitätsgruppe 6 mehr als 3 % ihrer Preise änderten. Einen ersten Überblick über die Anzahl der Preisänderungen auf der Ebene der Onlinehändler gibt die linke Seite von Grafik 1. Für einen Großteil der Onlinehändler wurden keine oder fast keine Preisänderungen (Gruppen 1 und 2) beobachtet. Diese haben einen Anteil von bis zu 0,2 % der Preisänderungen an ihren jeweils erhobenen Preisen. Etwa 96 % aller Onlinehändler zeigen nach der genannten Definition keine dynamische Preissetzung. Demnach scheint das derzeitige Erhebungsintervall von einmal im Monat ausreichend zu sein, um die monatliche Preisentwicklung genau darzustellen. Mehr als 3 % anteilige Preisänderungen wurden bei zehn Onlinehändlern (4 % der Gesamtheit) beobachtet, sie setzen Preise damit dynamisch.

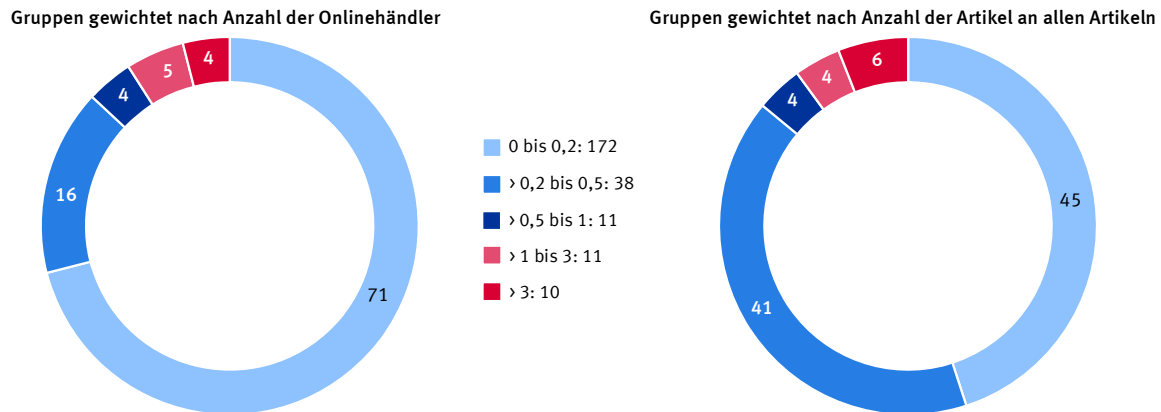
Der rechte Teil von Grafik 1 zeigt die Händler in den gleichen Volatilitätsgruppen über den gesamten Erhebungszeitraum gewichtet nach dem Anteil ihrer Artikel im Verhältnis zur Gesamtanzahl der Artikel. Die Händler der Gruppen 1 und 2, ohne oder mit sehr wenigen Preisänderungen, haben nun einen deutlich geringeren Anteil von weniger als 50 anstatt über 70 % in der Gewichtung nach ihrer Anzahl. Die Gruppe 6, mit dynamischer Preissetzung, hat jetzt einen erhöhten Anteil von 6 anstatt 4 %. Die Anpassung der Scraping-Intervalle wird für diese Händler vor dem Hintergrund der Datenqualität für den Verbraucherpreisindex empfohlen. Der Anteil ihrer Artikel an der Gesamtanzahl an Artikeln ist nunmehr statistisch signifikant. Mit der wachsenden Bedeutung des Onlinehandels kann davon ausgegangen werden, dass die Anzahl der VPI-relevanten Onlinehändler weiter stei-

⁵ Zum Vergleich der gewählten Abschneidegrenzen haben Blaudow/Burg (2018) ihre prozentualen Veränderungen deutlich weiter definiert. Eine Preisänderung gilt als auffällig, wenn ein gegebener Preis 500 % über oder fast 100 % unter dem vorigen Preis liegt. Außerdem beschreiben Blaudow/Seeger (2019), wie in Zukunft das Web-Scraping-Programm selbst Datenvalidierung beim Scraping-Vorgang betreiben soll, um die Erhebung unplausibler Preise zu reduzieren.

Dynamische Preissetzung im Onlinehandel: zur langfristigen Anwendung von automatisierter Preiserhebung

Grafik 1

Onlinehändler nach Volatilitätsgruppen¹ der Preisänderungen über den gesamten Erhebungszeitraum in %



1 Aus Darstellungsgründen wurden die Gruppen 1 (keine Preisänderungen) und 2 (bis 0,2 %) zusammengefasst.

2020 - 01 - 0211

gen wird, die mittels automatisierter Algorithmen ihre Artikelpreise ändern.

Zur genaueren Betrachtung der Onlinehändler mit den größten Anteilen an Preisänderungen sind diese in [Tabelle 3](#) anonymisiert gelistet. Die Onlinehändler sind nach der Anzahl abgedeckter Artikel absteigend sortiert. Die Tabelle listet alle Händler, die nach der 3%-Definition dynamisches Preissetzungsverhalten zeigen. Von den 241 erfassten Händlern erfüllen zehn diese Definition. Es fällt auf, dass acht der zehn Onlinehändler die 3%-Marke deutlich überschreiten. Von besonderem Interesse ist außerdem Händler 1, bei dem mehr als 1,2 Millionen Preise erhoben wurden. Die Arti-

kel von Händler 1 haben einen Anteil von über 4 % an der Gesamtanzahl an Artikeln. Sie sind in 83 der 324 COICOP-10-Steller vertreten, wodurch ihre Preisvolatilität potenziell einen signifikanten Einfluss auf die Indexberechnung haben kann.

3.2 Zeitliches Auftreten der Preisänderungen

Die Analyse des zeitlichen Auftretens der Preisänderungen erfolgt in den zwei Dimensionen Wochentag und Stunde. Zunächst wird die Anzahl der Preisänderungen gruppiert nach Wochentagen betrachtet. Es wird deutlich, dass – im Gegensatz zu den Arbeitstagen Montag bis Freitag – die Anteile der Preisänderungen an Samstagen und Sonntagen niedriger sind. Am Sonntag ist der Anteil der Preisänderungen mit 0,24 % am niedrigsten. In der Tendenz scheinen die Onlinehändler die Preise am Wochenende eher konstant zu halten. Die Anteile bleiben über die Arbeitswoche hinweg relativ konstant, liegen aber

Tabelle 3

Onlinehändler mit einem Anteil an Preisänderungen von jeweils mehr als 3 % an den erhobenen Preisen

	Preise		Preisänderungen		COICOP-10-Steller	Artikel	
	Anzahl		Anteil in %			Anzahl	Anteil in %
Onlinehändler 1	1 230 803	57 908	4,70		83	369	4,28
Onlinehändler 2	103 259	4 153	4,02		3	22	0,26
Onlinehändler 3	96 711	5 412	5,60		5	19	0,22
Onlinehändler 4	103 348	4 385	4,24		1	16	0,19
Onlinehändler 5	59 687	1 909	3,20		1	16	0,19
Onlinehändler 6	68 861	3 712	5,39		4	15	0,17
Onlinehändler 7	63 981	2 170	3,39		2	13	0,15
Onlinehändler 8	19 755	1 023	5,18		1	6	0,07
Onlinehändler 9	27 922	1 121	4,01		1	5	0,06
Onlinehändler 10	333	25	7,51		1	1	0,01

Tabelle 4

Preisänderungen nach Wochentagen

	Preis- änderungen	Preise	Preis- änderungen	Preis- steigerungen	Preis- senkungen	Verhältnis von Preissteigerungen zu Preissenkungen
	Anzahl		Anteil in %	Anzahl		
Montag	18 826	5 657 627	0,33	14 125	4 701	3,00
Dienstag	20 039	5 907 274	0,34	14 603	5 436	2,69
Mittwoch	21 343	6 043 378	0,35	15 652	5 691	2,75
Donnerstag	19 977	6 059 859	0,33	14 487	5 490	2,64
Freitag	21 843	6 246 878	0,35	15 855	5 988	2,65
Samstag	16 513	6 043 319	0,27	12 323	4 190	2,94
Sonntag	13 495	5 710 293	0,24	10 184	3 311	3,08

insgesamt auf einem höheren Niveau als am Wochenende. [↗ Tabelle 4](#)

An allen Wochentagen gibt es deutlich mehr Preissteigerungen als Preissenkungen. Im Median kommen auf eine Preissenkung 2,75 Preissteigerungen. Besonders Montage und Sonntage überschreiten diesen Wert deutlich. Anteilig gibt es am Wochenende deutlich weniger Preisänderungen im Vergleich zur Arbeitswoche. Bei diesen liegt aber das Verhältnis der Preissteigerungen zu -senkungen immer deutlich über dem Median der Woche.

Für eine optimierte Nutzung des Web Scrapers sollten Preise an unterschiedlichen Wochentagen erhoben werden. Es zeigen sich deutliche Unterschiede im Preissetzungsverhalten zwischen Wochenendtagen und Tagen innerhalb der Arbeitswoche. Werden Preise wöchentlich erhoben, sollte dies berücksichtigt und Preise abwechselnd in diesen Zeiträumen erhoben werden.

Abgesehen von den Wochentagen sollte die Erhebung auch zu unterschiedlichen Stunden erfolgen. Gerade zu Beginn

des Tages werden viele Preisanpassungen vorgenommen: 0,94 % in Stunde null, 0,48 % in Stunde vier. Ab Stunde 17 dagegen liegen die Anteile der Preisänderungen deutlich unter dem arithmetischen Mittel von 0,32 %. Zwar sind keine Umsatzdaten für die täglichen Uhrzeiten verfügbar, es ist aber davon auszuge-

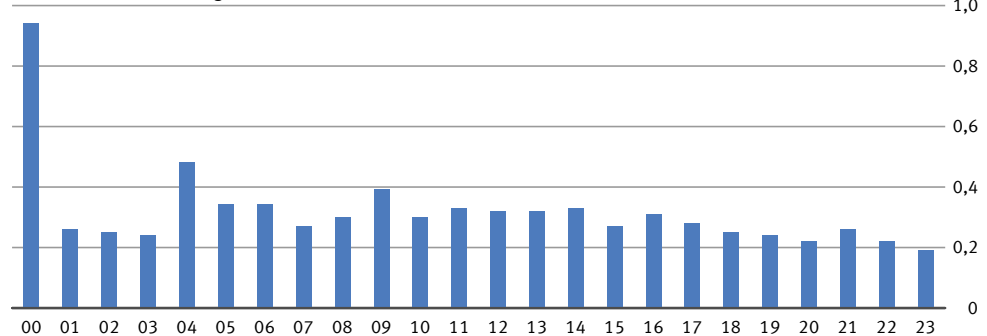
hen, dass die Konsumenten vermehrt abends (nach der Arbeit) online gehen, um Artikel zu kaufen. Auch die bereits erwähnte Verbraucherumfrage (Boniversum/bevh, 2017) untermauert dies: Die Befragten gaben an, dass sie 61 % ihrer online getätigten Einkäufe zwischen 18 und 24 Uhr durchgeführt haben. [↗ Grafik 2](#)

Neben den generellen Preisveränderungen ist auch die Anzahl der Preissteigerungen in den Morgenstunden

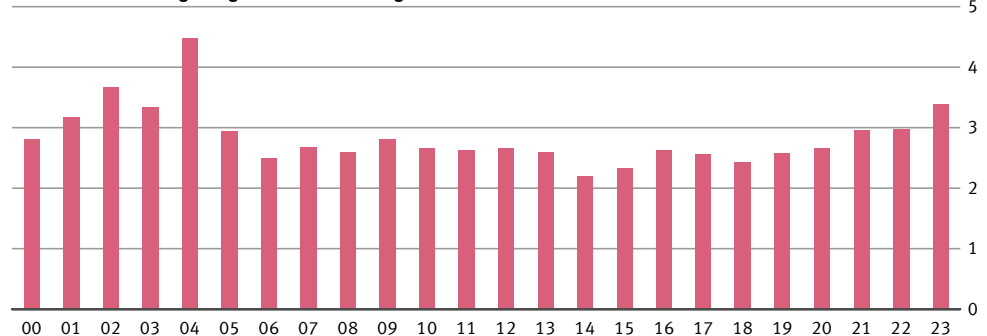
Grafik 2

Preisänderungen nach Tagesstunden

Anteil der Preisänderungen in %



Verhältnis Preissteigerungen zu Preissenkungen



2020 - 01 - 0213

hoch. Sie erreicht ihren höchsten Wert in der vierten Stunde mit durchschnittlich 4,47 Preissteigerungen je Preissenkung. Das arithmetische Mittel des Verhältnisses der Preisänderungen liegt über alle Tagesstunden hinweg bei 2,84. Von der vierzehnten bis zur zwanzigsten Stunden dagegen ist die Anzahl an Preissteigerungen vergleichsweise niedrig. Auch das lässt darauf schließen, dass Anbieter die Preise ihrer Onlineartikel tendenziell konstant halten, wenn mehr Konsumenten auf ihren Internetseiten einkaufen.

3.3 Handlungsempfehlungen zur automatisierten Preiserhebung

Insgesamt zeigt die Studie, dass die dynamische Preissetzung nach vorliegender Datenlage in einem überschaubaren Rahmen stattfindet und keine gravierenden Änderungen in der Datenerhebung für den Verbraucherpreisindex erfordert. Dynamische Preissetzung war bei zehn Onlinehändlern festzustellen; diese haben einen Artikelanteil von etwas mehr als 6 % an der Gesamtstichprobe.

Für die Datenerhebung ist vor allem der zeitliche Aspekt entscheidend. Speziell in den Morgenstunden, bis etwa 5 Uhr, werden seitens der Anbieter besonders viele Preise geändert. Ab 17 Uhr dagegen sind die Preisreihen sehr stabil. Nach Wochentagen betrachtet bleiben die Preise an den Wochenenden am stabilsten. Diese beiden Informationen kombiniert ergeben, dass vor allem Samstag- und Sonntagabende tendenziell geeignete Zeiträume für die Preiserhebung sind. Damit nicht nur die erhöhten Preise des Wochenendes erfasst werden, sollte das Scraping-Programm zumindest an einem zweiten Tag der Woche, zum Beispiel Mittwoch oder Donnerstag, zusätzlich Preise erheben. Dies sollte dann ebenfalls in den Abendstunden stattfinden, wenn die Preisreihen tendenziell unverändert bleiben.

Der Verbraucherumfrage (Boniversum/bevh, 2017) zufolge gaben 61 % der Befragten an, ihre Onlinekäufe bevorzugt zwischen 18 und 24 Uhr zu tätigen. Erfolgte die Preiserhebung also in diesem Zeitraum, würden vorrangig die Preise erfasst, die auch von den Verbraucherinnen und Verbrauchern bezahlt werden. In Bezug auf die Wochentage sind vor allem Samstag (30 %) und Freitag (16 %) beliebte Tage für Onlineeinkäufe, gefolgt von Sonntagen (15 %). Somit decken sich die identifi-

zierten Zeiten der Studie weitgehend mit den Erkenntnissen der Verbraucherumfrage.

Eine wöchentliche Erhebung könnte alle oben genannten Aspekte berücksichtigen und die Vorteile der automatisierten Datenerhebung nutzen, ohne unnötige Datenmengen zu erzeugen. Bei den zehn Onlinehändlern mit dynamischer Preissetzung könnte eine tägliche Erhebung sinnvoll sein. Die vorgestellten Ergebnisse sind Aggregationen über alle Onlinehändler hinweg. Es ist nicht auszuschließen, dass mehrere Onlinehändler leicht unterschiedliche Preissetzungsverhalten zeigen. Bei den täglichen Erhebungen könnten Scraping-Vorgänge deshalb zu unterschiedlichen Uhrzeiten angesetzt werden. Dies würde nicht nur die leichten Unterschiede der derzeit identifizierten Onlinehändler berücksichtigen, sondern wäre auch für eine sich verändernde Liste mit Anbietern geeignet. Wie die Daten gezeigt haben, sollte bei diesem Vorgehen zwischen null Uhr und fünf Uhr aufgrund der erhöhten Volatilität keine Preiserhebung erfolgen. Stattdessen sollten die täglichen Erhebungen ab fünf Uhr versetzt bis Mitternacht stattfinden.

4

Fazit und Ausblick


Mit der voranschreitenden Digitalisierung und dem damit verbundenen Wachstum des Onlinemarkts steigt der Modernisierungsdruck auf die Preisstatistik. Die Einführung der automatisierten Preiserhebung erlaubt es, die Erhebungsfrequenz zu erhöhen sowie die Anzahl der erfassten Artikel und dadurch die Präzision des Verbraucherpreisindex und dessen Teilindizes zu verbessern. Zudem lassen sich durch ihren Einsatz Ressourcen einsparen und Fehlerquellen, die mit der manuellen Preiserhebung einhergehen, minimieren. Das Ziel der beschriebenen Studie war es, die optimale Nutzung des Web-Scraping-Programms zu gewährleisten und die Erhebungsfrequenz entsprechend zu bestimmen.

Es ergibt sich also, dass Preiserhebungen an bestimmten Tagen und zu gegebenen Uhrzeiten sinnvoll sind. Die Onlinehändler zeigen ein klares Preissetzungsverhalten, das nach Wochentagen und Tagesstunden zu unterscheiden ist. Preise sollten erhoben werden, wenn diese möglichst konstant gehalten werden und die Konsumenten vermehrt die Produkte nachfragen. Das Wochenende

und die Abende an jedem Tag der Woche sind zur Erhebung besonders geeignet. Neben diesen Wochentagen und Uhrzeiten sollten aber auch zu anderen Zeiträumen Erhebungen stattfinden, um verfälschte Preisänderungen zu vermeiden, wie am Beispiel der Treibstoffpreise an Tankstellen verdeutlicht. Onlinehändler haben zudem ein besonderes Preissetzungsverhalten vor und um Sonder- und Feiertage gezeigt, das allerdings bereits durch die traditionelle Preiserhebung erfasst wird.

Daraus folgt, dass für die Mehrheit der Preisreihen wöchentliche Erhebungen empfohlen werden. Bei den Onlinehändlern mit sehr volatilen Preisreihen könnten tägliche Erhebungen die Präzision des Verbraucherpreisindex weiter erhöhen. Allerdings sollte das Web-Scraping-Programm möglichst ressourceneffizient eingesetzt werden. Übermäßige Datenmengen könnten sonst zu einer Überforderung der Datenanalysten führen.

Im Zuge dieser Analysen wurde auch das Auftreten der dynamischen Preissetzung untersucht. Ein vorläufiges Ergebnis der Studie ist, dass die derzeitige Praxis der monatlichen Preiserhebung eine adäquate Darstellung der monatlichen Preisentwicklung durch den Verbraucherpreisindex zu gewährleisten scheint. Abschließende Untersuchungen zu diesem Thema stehen aber noch aus und werden in einem folgenden Aufsatz vorgestellt werden.

Der Onlinehandel ist im stetigen Wandel begriffen und so sollte sich auch die Methodenforschung für den Verbraucherpreisindex entsprechend weiterentwickeln und die automatisierte Datenerhebung stärken. Der Einsatz des Web-Scraping-Programms erlaubt es, die Entwicklung auf dem Onlinemarkt gezielter zu beobachten und Onlinehändler zu identifizieren, die ihre Preise mittels Algorithmen automatisiert anpassen. 

LITERATURVERZEICHNIS

Blaudow, Christian/Burg, Florian. *[Dynamische Preissetzung als Herausforderung für die Verbraucherpreisstatistik](#)*. In: WISTA Wirtschaft und Statistik. Ausgabe 2/2018, Seite 11 ff.

Blaudow, Christian/Seeger, Daniel. *[Fortschritte beim Einsatz von Web Scraping in der amtlichen Verbraucherpreisstatistik – ein Werkstattbericht](#)*. In: WISTA Wirtschaft und Statistik. Ausgabe 4/2019, Seite 19 ff.

Creditreform Boniversum GmbH (Boniversum)/Bundesverband E-Commerce und Versandhandel Deutschland (bevh). *Verbraucherumfrage – Bevorzugte Tageszeiten und Wochentage für Online-Einkäufe*. 2017. [Zugriff am 13. Mai 2020]. Verfügbar unter: www.boniversum.de

Flower, Tanya/Karachalias, Eleftherios. *Using alternative data sources in consumer price indices: May 2019*. 2019. [Zugriff am 13. Mai 2020]. Verfügbar unter: www.ons.gov.uk

Little, Roderick J.A./Rubin, Donald B. *Statistical Analysis With Missing Data*. Wiley Series in Probability And Statistics. 2. Auflage. New Jersey 2002.

Statistisches Bundesamt. *Qualitätsbericht – Verbraucherpreisindex für Deutschland*. 2018. Verfügbar unter: www.destatis.de

Statistisches Bundesamt. *GENESIS-Online*. 2020. Verfügbar unter: <https://www-genesis.destatis.de>

van Loon, Ken/Roels, Dorien. *Integrating big data in the Belgian CPI*. Vortrag beim UNECE-ILO Meeting of the Group of Experts on Consumer Price Indices. 2018. [Zugriff am 13. Mai 2020]. Verfügbar unter: www.unece.org

Wahnbaeck, Carolin. *Ein Drittel der Deutschen kauft Geschenke vor allem online*. 2019. [Zugriff am 2. Dezember 2019]. Verfügbar unter: www.spiegel.de



Sarah Redlich

hat Survey Methodology (M.A.) an der Universität Duisburg-Essen studiert. Seit 2017 ist sie dort am Lehrstuhl für empirische Sozialforschung als wissenschaftliche Mitarbeiterin tätig und promoviert seit 2018 an der Fakultät für Gesellschaftswissenschaften. Für ihre Masterarbeit „Web Scraping zur Gewinnung von Testdaten für administrative Register“, die sie in diesem Artikel vorstellt, wurde sie 2019 mit dem Gerhard-Fürst-Preis des Statistischen Bundesamtes ausgezeichnet.

WEB SCRAPING ZUR GEWINNUNG VON TESTDATEN FÜR ADMINISTRATIVE REGISTER

Sarah Redlich

🔗 **Schlüsselwörter:** Traueranzeigen – Mortalität – Big Data – Web-Daten – Web Scraping

ZUSAMMENFASSUNG

Web Scraping verspricht eine kosten- und zeiteffiziente Nutzung von öffentlich zugänglichen Informationen aus dem Internet. Um Testdaten für die Erstellung eines bundesweiten Mortalitätsregisters zu generieren, könnten online gestellte Traueranzeigen mittels Web Scraping genutzt werden. Aus diesem Grund wurden für die Jahre 2015 und 2016 für eine Stadt die online extrahierten Traueranzeigen mit amtlichen Daten verglichen. Bereits bei der Konstruktion von Web Scrapern zeigten sich erhebliche Probleme. Ein Vergleich der extrahierten Informationen mit amtlichen Daten zeigen Unterschiede hinsichtlich Gesamtzahl, Geschlecht, ethnischer Zugehörigkeit und Alter der Verstorbenen.

🔗 **Keywords:** obituaries – mortality – big data – online data – web scraping

ABSTRACT

Web scraping promises cost and time-efficient use of open access information from the world wide web. To generate test data for creating a national mortality register, information could be extracted from online obituaries by means of web scrapers. The data scraped from a city's online obituaries of the years 2015 and 2016 were therefore compared with official data. Even building web scrapers caused major problems. A comparison of the information extracted and official data shows differences regarding the total number of deceased, their sex, ethnicity and age.

1

Einleitung

Die Durchführung standardisierter Befragungen wird zunehmend durch den Antwortausfall (Nonresponse) und den Beantwortungsaufwand sowie die Belastung der Befragten (Response Burden) erschwert (Schnell und andere, 2018; Wallgren/Wallgren, 2014). Als alternative Datenquelle dient das Internet, jedoch wäre eine manuelle Extraktion der dort gefundenen Informationen zeitaufwendig und mitunter fehleranfällig (Hoekstra und andere, 2010). Mit Erhebungstechniken, wie dem sogenannten Web Scraping, können öffentlich zugängliche Informationen aus dem Internet nutzbar gemacht werden (Cooley und andere, 1997; Massimino, 2016).

Web Scraping wird bereits erfolgreich in der amtlichen Statistik sowie in den Bereichen Data Science, Journalismus und Marketing genutzt (Landers und andere, 2016; Blaudow/Seeger, 2019). Ein weiteres potenzielles Anwendungsgebiet von Web Scraping ist die Generierung von Testdaten für den Aufbau von administrativen Registern. Beispielsweise wurde bisher in Deutschland kein bundesweites Mortalitätsregister eingeführt. Testdaten sollten Informationen wie Namen, Geburtsdatum und Sterbedatum umfassen. Webb und andere (1966) schlugen vor, Informationen von Grabsteinen zu nutzen. Eine aktuelle Alternative dazu sind online gestellte Traueranzeigen, die mittels Web Scraping in einen nutzbaren Datensatz gebracht werden können. Traueranzeigen als Informationsquelle wurden bereits von Boak und anderen (2007), Soowamber und anderen (2016) und Sylvestre und anderen (2018) erfolgreich erschlossen.¹ Nachfolgend wird untersucht, ob Web Scraping von online verfügbaren Traueranzeigen zu brauchbaren Testdaten führt.

Zunächst wird erläutert, was unter der Datenerhebungstechnik Web Scraping zu verstehen ist (Kapitel 2) und welche ethischen und rechtlichen Grundlagen für Web Scraping gelten (Kapitel 3). Kapitel 4 gibt einen kurzen Überblick zur Entstehung und zum Aufbau von Traueranzeigen und deren Veröffentlichung in Trauerportalen. Kapitel 5 behandelt die Probleme bei der Entwicklung

von Web Scrapern zur Extraktion von Traueranzeigen. Die Datensatzbeschreibung und der Vergleich der extrahierten Informationen mit amtlichen Daten folgen in den Kapiteln 6 und 7, bevor im letzten Kapitel eine abschließende Betrachtung folgt.

2

Web Scraping

Web Scraping bezeichnet die Extraktion von Information aus dem World Wide Web (Glez-Peña und andere, 2013; Cooley und andere, 1997). In der Literatur wird Web Scraping auch als Web Harvesting, Web Crawling und Web Mining bezeichnet, wobei lediglich der Begriff Web Harvesting synonym verwendet werden kann (Gatterbauer, 2009; Bharanipriya/Kamakshi Prasad, 2011; Najork, 2009). Web Scraper ahmen die Interaktion zwischen Mensch und Server nach (Glez-Peña und andere, 2013). Dabei wird zunächst auf die Seite zugegriffen. Mittels regulärer Ausdrücke² wird dann die gewünschte Information gesucht, anschließend extrahiert und in ein gewünschtes Datenformat gebracht.

Zu den Vorteilen von Web Scraping gegenüber anderen Datenerhebungstechniken zählen die Zeiteffizienz, die häufigere Datensammlung und die Kostenreduzierung. Letztere gilt jedoch nur, sofern der Aufwand der Datenaufbereitung gering ist und Web Scraper nie bis selten überarbeitet werden müssen (Hoekstra und andere, 2010; Landers und andere, 2016). Im Vergleich zu standardisierten Befragungen (Surveys) werden ebenfalls der Response Burden sowie Antwortverzerrungen umgangen (Hoekstra und andere, 2010; Landers und andere, 2016). In der Literatur wird auch betont, dass sich die Qualität von Statistiken verbessert (Hoekstra und andere, 2010; Blaudow/Seeger, 2019).

Von Nachteil ist, dass Web Scraping zunächst viele Ressourcen benötigt, da Personen mit Wissen und Erfahrungen im Bereich Web Scraping eingestellt (oder ausgebildet) werden müssen (Hoekstra und andere, 2010). Zudem können durch sich stark verändernde Webseiten Web Scraper schnell unbrauchbar werden und eine zeitaufwendige Anpassung erfordern (Hoekstra und andere,

¹ Die Idee zu dieser Arbeit stammt von Prof. Dr. Rainer Schnell. Teile dieser Arbeit und zusätzliche statistische Tests zur Signifikanz der Ergebnisse finden sich in Schnell/Redlich (2019).

² Mit einem regulären Ausdruck kann eine bestimmte Zeichenabfolge über syntaktische Regeln in einer Zeichenkette identifiziert werden.

2010). Hinzu kommen durch Web Scraper überladene Server und Kosten aufseiten der Betreiber der Webseite durch das Benutzen der Bandbreite. Die letztgenannten Probleme dürften mittlerweile jedoch zu vernachlässigen sein (Koster, 1993b; Thelwall/Stuart, 2006). Trotzdem ist der Ressourcenaufwand zur Aufbereitung der extrahierten Informationen nicht zu vernachlässigen (Hoekstra und andere, 2010). Die größten Probleme bei der Verwendung von Web Scrapern sind jedoch ethischer und rechtlicher Natur.

3

Ethische und rechtliche Grundlagen

Neben allgemeinen ethischen Richtlinien für Datenerhebungen³ gilt es zu betonen, dass mittels Web Scraping Informationen von Personen ohne deren Wissen und Zustimmung erhoben werden können (van Wel/Royakkers, 2004). Daher wurde das sogenannte Robots Exclusion Protocol (robots.txt) eingeführt, welches eine Extraktion von Informationen blockiert (Koster, 1993a; Thelwall/Stuart, 2006). Trotz der Möglichkeit, auch dieses zu umgehen, sollte eine Extraktion nur erfolgen, sofern das Robots Exclusion Protocol dies gestattet (Kouzis-Loukas, 2016; Landers und andere, 2016).

Die nachfolgende Darstellung der rechtlichen Grundlagen bezieht sich auf den Stand vom Oktober 2017 und wurde im Hinblick auf die Extraktion von Informationen von online gestellten Todesanzeigen betrachtet. Für eine aktuellere Anwendung oder eine Anwendung in einem anderen Bereich ist eine individuelle Bewertung der Rechtslage zwingend erforderlich.⁴

Web Scraping ist weder explizit erlaubt noch verboten, weshalb es grundsätzlich als legal angesehen wird (Black, 2016). Durch die immer häufigere kommerzielle Nutzung von Web Scrapern werden zunehmend Regelungen getroffen und Urteile gefällt, die nicht immer eine wissenschaftliche Nutzung der Daten berücksichtigen (Black, 2016). Ob die Extraktion von Daten mittels Web

Scrapern legal ist, hängt davon ab, ob der Inhalt oder die Anordnung der Daten urheberrechtlich geschützt sind. Zentrale Gesetze sind hierbei § 87 des Urheberrechtsgesetzes sowie die Richtlinie 96/9/EG des Europäischen Parlaments und des Rates über den rechtlichen Schutz von Datenbanken (hier: Kapitel zu Urheberrecht, Schutz Sui Generis). Eine Reihe von Urteilen⁵ in den vergangenen Jahren lassen dabei den Schluss zu, dass Web Scraping legal ist – und zwar auch dann, wenn die Nutzungsbedingungen dies verbieten, diesen aber (zum Beispiel durch die Erstellung eines Benutzerkontos) nicht aktiv zugestimmt wurde.

Ferner ist Web Scraping unproblematisch, wenn nur ein unwesentlicher Teil der Datenbank extrahiert wird und keine Veröffentlichung der Daten erfolgt. Zudem erklären sowohl § 87c Absatz 1 Urheberrechtsgesetz als auch die Richtlinie 96/9/EG (in Artikel 6 Absatz 2 und Artikel 9), dass es Ausnahmen für den wissenschaftlichen Gebrauch gibt. Für die Extraktion von Informationen von Traueranzeigen ist zudem von Interesse, dass personenbezogene Daten von Verstorbenen extrahiert werden. Diese unterliegen nicht mehr dem Datenschutz (§ 3 Absatz 1 Bundesdatenschutzgesetz). Allerdings dürfen keine Informationen von lebenden Personen verwendet werden, beispielsweise Namen und Kontaktdaten von Angehörigen (Löwer, 2010).

4

Traueranzeigen als potenzielle Datenquelle

Bereits seit dem 18. Jahrhundert wird der Tod eines Menschen mittels einer Traueranzeige bekannt gemacht; sie ersetzte die aufwendigere Form der sonntäglichen Kanzelabkündigung (Grüner/Helmrich, 1994). In Deutschland wurde 1753 die erste (belegte) Traueranzeige veröffentlicht, erst rund zehn Jahre später war in Zeitungen eine eigene Rubrik zu Traueranzeigen zu finden (Grüner/Helmrich, 1994). Seit Anfang des 19. Jahrhunderts werden Traueranzeigen regelmäßig in eigenen Rubriken in Zeitungen abgedruckt (Grüner/Helmrich, 1994).

3 Beispielsweise der American Association for Public Opinion Research (American Association for Public Opinion Research, 2015).

4 So erlaubt zum Beispiel das Preisstatistikgesetz durch die Änderung im Dezember 2019 die Erhebung von Preisen mit automatischen Abrufverfahren zur Erstellung benötigter Statistiken (§ 7 Absatz 2 Preisstatistikgesetz).

5 Zu den Urteilen zählen 6 U 221/08 des OLG Frankfurt am Main, 308 O 162/09 des LG Hamburg und 3 U 191/08 des OLG Hamburg sowie I ZR 159/10 des Bundesgerichtshofs und Urteil C-30/14 des Europäischen Gerichtshofs.

Obwohl es keinen aufgezwungenen Standard (mit Ausnahme der Größe, die durch den Preis bedingt wird) seitens der Zeitungen hinsichtlich des Aussehens und des Informationsgehalts von Todesanzeigen gibt, hat sich im Lauf der Jahre eine Standardform etabliert. Inhaltlich lassen sich in der Regel folgende Informationen finden: Vorspruch, Symbol, persönliche Daten der verstorbenen Person (Name, Titel, Geburtsname, Geburtsdatum, Sterbedatum), Trauertext, Namensliste der Hinterbliebenen und Hinweise religiöser, organisatorischer oder sonstiger Art (Grüner/Helmrich, 1994). [↗ Grafik 1](#)

Grafik 1

Aufbau von Traueranzeigen

Symbol	Vorspruch
	Name, Titel Geburtsname Geburtsdatum, Sterbedatum
	Trauertext
Überleitung Angehörige	
Sonstige Hinweise	

Geänderte Darstellung aus Grüner/Helmrich, 1994.

2020 - 01 - 0206

Durch die zunehmende Technisierung und die zunehmende Nutzung des Internets finden sich immer mehr sogenannte Trauerportale, zumeist von Zeitungsverlagen betrieben, die eine Online-Veröffentlichung von Traueranzeigen anbieten. Es gibt auch Trauerportale, die unabhängig von Zeitungen arbeiten, jedoch ist hier nicht ersichtlich, ob ein Beleg über den Tod in Form einer Sterbeurkunde vorgelegt werden muss, wie es bei Zeitungen verlangt wird. Ein Vorteil von Trauerportalen für die Nutzung von Traueranzeigen als Datenquelle ist, dass sowohl die gedruckte Anzeige als Bild vorhanden ist, als auch die interessierenden Angaben (Name, Geburts- und Sterbedaten) aufgeführt werden. Dies erleichtert eine Extraktion der Informationen.

Eine genaue Angabe zur Anzahl an Trauerportalen gibt es nicht. Daher wurde eine Liste aller online vertretenen (Tages-)Zeitungen in Deutschland erstellt; über deren Internetpräsenzen konnten 152 Trauerportale von Zeitungen identifiziert werden, zudem ein Trauerportal, welches mehr als nur eine Zeitung/Zeitungsgruppe umfasst.⁶ Fünf dieser Trauerportale sind registrierungs-

oder kostenpflichtig und daher für das Web Scraping ungeeignet. Die restlichen konnten hinsichtlich ihres Layouts und der zugrunde liegenden Quellcodes in drei Gruppen eingeteilt werden. Dabei befinden sich in der letzten Gruppe Trauerportale mit jeweils individuellen Quellcodes, sodass jedes Trauerportal einen eigenen Web Scraper benötigen würde. Diese Trauerportale wurden nicht berücksichtigt. Für Gruppe 1 und Gruppe 2 wurde jeweils ein Web Scraper entwickelt, somit konnten insgesamt 109 Trauerportale abgedeckt werden. Bei den nicht berücksichtigten Trauerportalen handelt es sich jedoch ausschließlich um Portale aus anderen Regionen als der in diesem Artikel betrachteten, sodass hier keine Traueranzeigen unbeachtet blieben.

5

Web Scraping von Traueranzeigen

Die Web Scraper wurden in Python 2.7 unter der Verwendung von Scrapy 1.3.3 geschrieben.⁷ Die Web Scraper sollten dabei den Vornamen und Nachnamen der verstorbenen Person einschließlich Geburtsname und Zweitnamen sowie das Geburtsdatum, das Sterbedatum, die Zeitung der Veröffentlichung und das Datum der Veröffentlichung extrahieren. Dabei wurde bereits durch die Start-URL⁸ nach Ort und Datum gefiltert, sodass nur Anzeigen für den interessierenden Ort und den Zeitraum extrahiert wurden.

Bei der Entwicklung von Web Scrapern zeigten sich einige Probleme. So gab es für einige Personen mehr als ein Bild der Traueranzeige, beispielsweise, weil eine zweite Traueranzeige als Jahresgedenken veröffentlicht wurde, oder weil mehrere Anzeigen in unterschiedlichen Zeitungen veröffentlicht wurden. Dadurch entstanden unzulässige Links für den Download des Bilds der Traueranzeige. In diesem Fall wurde immer das erste auftauchende Bild verwendet.

Ein weiteres Problem bestand darin, dass meist nur ein kleineres Vorschaubild der Anzeige gegeben war und größere Bilder mit einer besseren Auflösung in JavaScript eingebettet waren. Zu kleine Bilder machen

⁷ Eine Einführung bietet Kouzis-Loukas (2016).

⁸ Bei der Start-URL handelt es sich um die URL der Webseite des Portals, nachdem in der Suchmaske des Portals die Filter nach Ort und Datum gesetzt wurden.

⁶ Stand: Oktober 2017.

eine Texterkennung unmöglich und sind daher nicht nutzbar. Bilder, die in JavaScript eingebettet sind, sind zwar prinzipiell extrahierbar, allerdings konnte während eines angemessenen Zeitraums keine Lösung gefunden werden, um die Bilder zu extrahieren. Des Weiteren wiesen heruntergeladene Bilder einen alphanumerischen Dateinamen auf, welcher eine Zuordnung zu einer Person nicht möglich macht. Daher musste eine zusätzliche Liste erstellt werden, die den Namen des heruntergeladenen Bilds der Person zuordnet.

Bilder der Traueranzeigen können theoretisch verwendet werden, um die darauf enthaltenen Informationen zu extrahieren. Dies ist besonders von Interesse, wenn keine Auflistung der interessierenden Information stattfindet. Mittels OCR(Optical Character Recognition)-Texterkennung ist es theoretisch möglich, alle notwendigen Informationen zu extrahieren. Bei Bildern mit schlechter Auflösung, der Verwendung von Hintergrundbildern oder der Verwendung ungewöhnlicher Schriftarten (zum Beispiel Kalligraphie-Schriftarten) unterlaufen der Texterkennung viele Fehler. Versuche, dennoch Informationen aus den Bildern zu extrahieren, waren erfolglos.

Bei Trauerportalen der zweiten Gruppe sind die Datumsangaben nicht separat aufgeführt, sondern nur in einer 300-Zeichen-Vorschau des Textes der Traueranzeige eingebettet. Hier mussten also aus einem Fließtext das Geburtsdatum und das Sterbedatum extrahiert werden. Dabei erwiesen sich die unterschiedlichen Datumsformate als problematisch, ebenso fehlende Geburtsdaten und zusätzlich vorkommende Daten, wie der Tag der Beerdigung. Dieses Problem war nur durch die Annahme zu lösen, dass es sich bei den ersten zwei gefundenen Daten um das Geburtsdatum sowie das Sterbedatum handelt und – sofern nur ein Datum gefunden wurde – nur das Sterbedatum erfasst wurde.

Das gravierendste Problem beim Web Scraping ist jedoch die mögliche Veränderung des Quellcodes von Webseiten. Hierdurch werden Web Scraper im schlimmsten Fall vollkommen unbrauchbar. Zudem kann die Anzahl der gefundenen Traueranzeigen zwischen verschiedenen Zeitpunkten variieren – beispielsweise, weil ältere Traueranzeigen später hinzugefügt oder Anzeigen aus den Trauerportalen entfernt werden. Schließlich kann sich auch der öffentliche Zugang zu den Webseiten ändern, indem das Robots Exclusion Protocol der Webseite einen Zugang unterbindet.

6

Datensatzbeschreibung

Für die Jahre 2015 und 2016 wurden für eine Großstadt 6 597 Traueranzeigen gefunden. Nachdem Anzeigen außerhalb des interessierenden Zeitraums und doppelte Anzeigen gelöscht wurden, enthielt der Datensatz 3 007 Anzeigen von verstorbenen Personen. Auf Grundlage von amtlichen Daten wurden basierend auf den Vornamen der Personen das Geschlecht und die ethnische Zugehörigkeit zugewiesen. Drei Namen, für die über die amtlichen Daten kein Geschlecht zugewiesen werden konnte, wurde manuell ein Geschlecht zugeordnet. Hierbei handelte es sich um Namen mit eindeutiger Geschlechtszuweisung.

Die Traueranzeigen wurden mit amtlichen Daten verglichen. In Deutschland besteht nach § 28 Personenstandsgesetz die Verpflichtung, den Tod eines Menschen binnen drei Tagen dem zuständigen Standesamt zu melden. Damit enthalten die Daten des Standesamts alle im Einzugsbereich des Standesamts verstorbene Personen, unabhängig von deren Wohnsitz. Die Informationen des Standesamts werden allerdings nur aggregiert als Anzahl der Personen getrennt nach Jahr, Geschlecht und ethnischer Zugehörigkeit zur Verfügung gestellt. Das Standesamt ist nach § 60 Absatz 1 Personenstandsverordnung verpflichtet, andere Behörden über den Tod zu informieren. Für Mikrodaten wurde auf amtliche Daten zurückgegriffen, die nur die verstorbenen Personen mit Wohnsitz in der interessierenden Stadt enthalten. Die Mikrodaten umfassten Nachnamen, Rufnamen, Geburtsdaten und Todesdaten. Unterschiede zeigten sich bereits zwischen diesen beiden Datensätzen. Während das Standesamt 15 001 verstorbene Personen registrierte, enthielten die amtlichen Daten der Verstorbenen mit Wohnsitz in der interessierenden Stadt nur 14 003 verstorbene Personen.

7

Vergleich der Trauerportale mit amtlichen Daten

Für den Vergleich der Trauerportale mit den amtlichen Daten wurden zunächst die Datensätze hinsichtlich ihrer Geschlechterzusammensetzung, der ethnischen Zugehörigkeit und der Altersverteilung einander gegenübergestellt. Darüber hinaus wurde überprüft, ob die Datensätze übereinstimmen und inwieweit sich die übereinstimmenden Verstorbenen von den nicht übereinstimmenden Verstorbenen unterscheiden. Statistische Tests zur Signifikanz der gefundenen Unterschiede finden sich in Schnell/Redlich (2019).

7.1 Vergleich der Datensätze

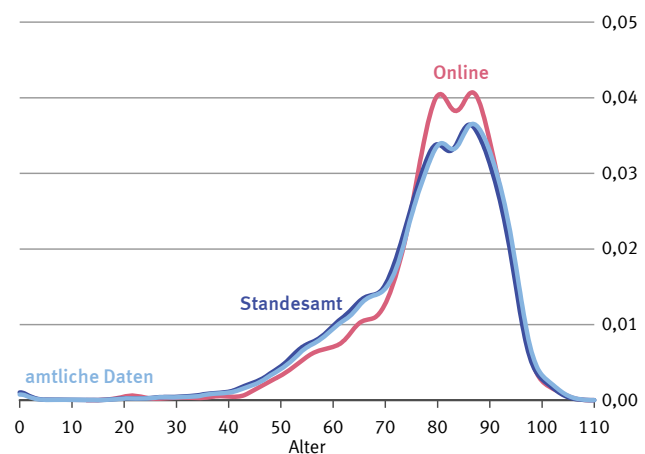
Ein Vergleich der Geschlechteranteile zeigt Folgendes: Während die amtlichen Daten einen höheren Frauenanteil bei den Verstorbenen aufweisen (52,4% nach den Standesamtsdaten; 52,9% nach den amtlichen Daten der Verstorbenen mit Wohnsitz in der interessierenden Stadt), zeigen die online extrahierten Daten einen marginal höheren Männeranteil (lediglich 49,6% der Verstorbenen waren demnach Frauen). Hinsichtlich der ethnischen Zugehörigkeit zeigen zwar alle Datensätze einen deutlich höheren Anteil an Verstorbenen mit deutscher Zugehörigkeit, jedoch waren laut amtlichen Daten 96,2% (Standesamtsdaten) beziehungsweise 97,4% (amtliche Daten der Verstorbenen mit Wohnsitz in der interessierenden Stadt) verstorbene Personen deutsch, während die online extrahierten Daten einen Anteil von 99,2% aufweisen.

↳ Grafik 2 zeigt die Altersverteilung der Verstorbenen in den drei Datensätzen. In den Online-Daten finden sich mehr verstorbene Personen im Alter von 75 bis 90 Jahren und weniger im Alter von 40 bis 70 Jahren im Vergleich zu den amtlichen Daten. Zudem lassen sich keine Personen unter 18 Jahren finden. Aus diesem Grund wurden für eine Zeitung für das Jahr 2015 die gedruckten Todesanzeigen mit den online gestellten Anzeigen verglichen. Online konnten 1 212 Anzeigen gefunden werden, während jedoch 2 766 Anzeigen abgedruckt wurden. Die Übereinstimmung zwischen den Anzeigen betrug 1 134 Personen. Dies bestätigt, dass Angehö-

rige wählen können, ob die Traueranzeige auch online gestellt wird. Durch den Vergleich zeigte sich, dass es durchaus Anzeigen von Minderjährigen gibt und diese auch online gestellt werden. Allerdings kann aufgrund von fehlenden Datumsangaben das Alter nicht berechnet werden.

Grafik 2

Altersverteilung der Online-Daten und amtlichen Daten nach Dichte



Verstorbene Personen der Online-Daten sind älter.

2020 - 01 - 0193

7.2 Übereinstimmung der Datensätze

Zusätzlich fand ein Abgleich der Datensätze statt, um den Anteil der Personen zu bestimmen, die über Traueranzeigen gefunden werden können. Dies war entsprechend nur mit den Mikrodaten möglich. Einer Person wurde der Status „Duplikat“ gegeben, wenn der Name, das Geburtsdatum und das Sterbedatum übereinstimmten. Sofern das Geburtsdatum nicht vorhanden war, musste nur eine Übereinstimmung von Name und Sterbedatum gegeben sein. Von den online extrahierten Daten konnten 70,5% der Verstorbenen in den amtlichen Daten gefunden werden, aber 84,9% der Verstorbenen laut amtlichen Daten war keine Traueranzeige zuzuordnen.

Die genauere Betrachtung der Geschlechteranteile zeigt, dass anteilig mehr Männer in beiden Datensätzen zu finden sind. Frauen sind häufiger nur in den amtlichen Daten oder nur in den Online-Daten zu finden. Hinsicht-

lich der ethnischen Zugehörigkeit ist der Anteil an nicht deutschen Verstorbenen bei den Personen, die sich in beiden Datensätzen befinden, deutlich niedriger. Dies legt erneut nahe, dass das Veröffentlichen von Traueranzeigen nicht in allen ethnischen Gruppen praktiziert wird. ➤ Tabelle 1, Tabelle 2

Tabelle 1

Geschlechteranteile unterscheiden sich zwischen Personen, die sich in beiden Datensätzen befinden, und Personen, die sich nur in einem der Datensätze befinden

	Duplikat		Einmalig	
	Anzahl	%	Anzahl	%
Weiblich	1 050	49,6	6 785	53,2
Männlich	1 066	50,4	5 961	46,8
Insgesamt	2 116	100	12 746	100

Tabelle 2

Die ethnische Zugehörigkeit zeigt, dass für nichtdeutsche Personen weniger Traueranzeigen veröffentlicht werden

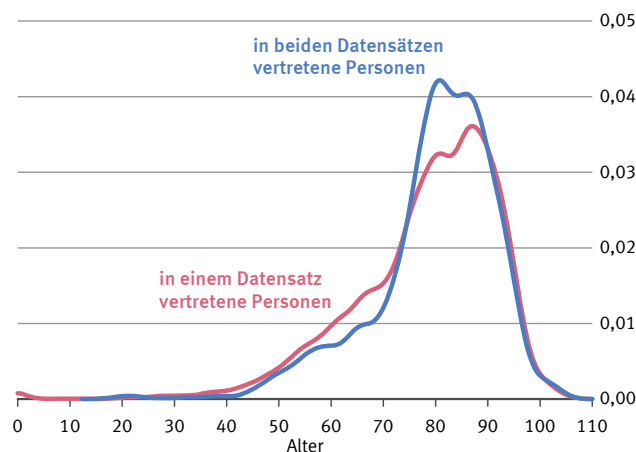
	Duplikat		Einmalig	
	Anzahl	%	Anzahl	%
Deutsch	2 103	99,4	12 376	97,2
Nichtdeutsch	12	0,6	356	2,8
Insgesamt	2 115	100	12 732	100

Eine Betrachtung der Altersverteilung zeigt, dass Personen, die sich nicht in beiden Datensätzen befinden, jünger sind. Zudem zeigt sich auch hier wieder deutlich das Problem, dass Minderjährige nicht über Traueranzeigen gefunden wurden. ➤ Grafik 3

Für Traueranzeigen, für die keine Entsprechung in den amtlichen Daten gefunden werden konnte, zeigte sich, dass die Ortseingrenzung einen unzureichenden Filter darstellte. So können Orte, deren Namen einen Teil eines anderen Orts bilden, nicht ausreichend differenziert werden; somit werden Personen eines anderen Orts berücksichtigt. Die Information des Orts ist jedoch nur aus dem Bild der Traueranzeige zu entnehmen, sodass keine Optimierung des Filters erreicht werden konnte. Zudem wurden Unstimmigkeiten in der Schreibweise von Namen gefunden sowie die Verwendung von Spitz- oder Rufnamen statt der offiziellen Vornamen. Schließlich wurden auch abweichende Geburtsdaten und Sterbedaten gefunden, die eine Zuordnung erschweren.

Grafik 3

Altersverteilung von Personen, die in beiden Datensätzen, und Personen, die in nur einem Datensatz zu finden sind nach Dichte



Jüngere Personen lassen sich nicht in beiden Datensätzen finden.

2020 - 01 - 0208

8


Schlussbetrachtung

Im Fokus der Arbeit stand die Entwicklung eines Web Scrapers zur Extraktion von Traueranzeigen. Web Scraping von öffentlich zugänglichen Daten bietet die Möglichkeit, den Response Burden von zu befragenden Personen vollständig zu reduzieren und erlaubt eine häufigere Erhebung von Informationen, wie sie in einem Survey nicht möglich wären (Hoekstra und andere, 2010; Landers und andere, 2016).

Traueranzeigen stellen eine mögliche Datenquelle für verstorbene Personen dar und könnten theoretisch als Testdaten für ein bundesweites Mortalitätsregister genutzt werden. Durch die zunehmende Zahl sogenannter Trauerportale, welche Traueranzeigen online zugänglich machen, können diese mittels Web Scraping extrahiert werden.

Obwohl für einen Großteil der Trauerportale zwei Web Scraper ausreichen, zeigten sich erhebliche Probleme bei deren Entwicklung, beispielsweise durch die Veränderung des zugrunde liegenden Quellcodes. Des Weiteren sind interessierende Informationen nicht bei allen Trauerportalen einzeln aufgeführt, sondern zwischen

anderem Text oder in JavaScript eingebettet, was eine Extraktion erschwert. Eine Texterkennung von Bildern der Traueranzeigen erwies sich als wenig erfolgreich.

Ein Abgleich der extrahierten Informationen mit amtlichen Daten zeigte einen höheren Anteil an männlichen Verstorbenen sowie einen deutlich geringeren Anteil an nicht deutschen Verstorbenen in den Online-Daten. Auch bei der Altersverteilung zeigen sich erhebliche Unterschiede. Verstorbene der extrahierten Todesanzeigen sind deutlich älter als die amtlichen Daten zeigen. Ein Abgleich hinsichtlich des Namens, des Geburtsdatums sowie des Sterbedatums zeigte, dass rund 85 % der laut amtlichen Daten verstorbenen Personen nicht über eine Traueranzeige gefunden werden konnten. Personen, die entweder nicht über eine Traueranzeige gefunden werden konnten oder zum Beispiel aufgrund von ungenauen Filtern nicht in den amtlichen Daten sind, sind eher weiblich, jünger und der Anteil nicht deutscher Personen ist vergleichsweise größer. Eine Verwendung von geeigneteren Verfahren zur Zusammenführung von Datensätzen⁹ könnte auch diejenigen Übereinstimmungen trotz Tippfehlern finden. Allerdings dürfte auch dies kaum die Unterschiede hinsichtlich der Anzahl und der demografischen Charakteristika der Verstorbenen von online gestellten Traueranzeigen zu amtlichen Daten kompensieren. 

⁹ In der Literatur finden sich diese Verfahren unter dem Stichwort „Record Linkage“.

LITERATURVERZEICHNIS

American Association for Public Opinion Research (AAPOR). *The Code of Professional Ethics and Practices (Revised 11/30/2015)*. 2015. [Zugriff am 20. April 2020]. Verfügbar unter: www.aapor.org

Bharanipriya, V./Kamakshi Prasad, V. *Web Content Mining Tools: A Comparative Study*. In: International Journal of Information Technology and Knowledge Management. Ausgabe 4(1)/2011, Seite 211 ff.

Black, Michael L. *The World Wide Web as Complex Data Set: Expanding the Digital Humanities into the Twentieth Century and Beyond through Internet Research*. In: International Journal of Humanities and Arts Computing. Ausgabe 10(1)/2016, Seite 95 ff. DOI: 10.3366/ijhac.2016.0162

Blaudow, Christian/Seeger, Daniel. *Fortschritte beim Einsatz von Web Scraping in der amtlichen Verbraucherpreisstatistik – Ein Werkstattbericht*. In: WISTA Wirtschaft und Statistik. Ausgabe 4/2019, Seite 19 ff.

Boak, Marshall B./M'ikanatha, Nkuchia M./Day, Roger S./Harrison, Lee H. *Internet Death Notices as a Novel Source of Mortality Surveillance Data*. In: American Journal of Epidemiology. Ausgabe 167(5)/2007, Seite 532 ff.

Cooley, Robert/Mobasher, Bamshad/Srivastava, Jaideep. *Web Mining: Information and Pattern Discovery on the World Wide Web*. In: 9th International Conference on Tools with Artificial Intelligence, ICTAI '97, Newport Beach, CA, USA, November 3-8, 1997. Los Alamitos 1997, Seite 558 ff. DOI: [10.1109/TAI.1997.632303](https://doi.org/10.1109/TAI.1997.632303)

Gatterbauer, Wolfgang. *Web Harvesting*. In: Liu, Ling/Özsu, M. Tamer (Herausgeber). *Encyclopedia of Database Systems*. New York 2009, Seite 3472 f.

Glez-Peña, Daniel/Lourenço, Anália/López-Fernández, Hugo/Reboiro-Jato, Miguel/Fdez-Riverola, Florentine. *Web Scraping Technologies in an API World*. In: *Briefings in Bioinformatics*. Ausgabe 15(5)/2013, Seite 788 ff. DOI: <https://doi.org/10.1093/bib/bbt026>

Grüner, Karl-Wilhelm/Helmrich, Robert. *Die Todesanzeige: Viel gelesen, jedoch wenig bekannt: Deskription eines wenig erschlossenen Forschungsmaterials*. In: *Historical Research*. Ausgabe 19(1)/1994, Seite 60 ff.

Hoekstra, Rutger/Bosch, Olav Ten/Harteveld, Frank. *Automated Data Collection from Web Sources for Official Statistics: First Experiences*. Technischer Bericht 2010-132-KOO. Statistics Netherlands 2010.

Koster, Martijn. *About /robots.txt*. 1993a. [Zugriff am 20. April 2020]. Verfügbar unter: <http://www.robotstxt.org/robotstxt.html>

Koster, Martijn. *Guidelines for Robot Writers*. 1993b. [Zugriff am 20. April 2020]. Verfügbar unter: <http://www.robotstxt.org/guidelines.html>

LITERATURVERZEICHNIS

Kouzis-Loukas, Dimitrios. *Learning Scrapy: Learn the Art of Efficient Web Scraping and Crawling with Python*. Birmingham 2016.

Landers, Richard N./Brusso, Robert C./Cavanaugh, Katelyn J./Collmus, Andrew B. *A Primer on Theory-Driven Web Scraping: Automatic Extraction of Big Data From the Internet for Use in Psychological Research*. In: Psychological Methods. Ausgabe 21(4)/2016, Seite 475 ff.

Löwer, Wolfgang. *Anhang 5: Grenzen der Erhebung und Verarbeitung von Sterbedaten durch den (postmortalen) Persönlichkeitsschutz – Überlegungen zum Schutzzumfang verstorbener Personen und der lebenden Angehörigen*. In: Dicke, Peter/Dietel, Manfred/Gawrich, Stefan/Jöckel, Karl-Heinz/Klug, Stefanie/Löwer, Wolfgang/Luttmann, Sabine/Müller, Ulrich/Schelhase, Torsten/Vennemann, Mechthild/Ziese, Thomas/Schmidt-Stolte, Martina (Herausgeber). Ein Nationales Mortalitätsregister für Deutschland: Bericht der Arbeitsgruppe und Empfehlung des Rates für Sozial- und Wirtschaftsdaten (RatSWD). 2010. Seite 31 ff.

Massimino, Brett. *Accessing Online Data: Web-Crawling and Information-Scraping Techniques to Automate the Assembly of Research Data*. In: Journal of Business Logistics. Ausgabe 37(1)/2016, Seite 34 ff. DOI: <https://doi.org/10.1111/jbl.12120>

Najork, Marc. *Web Crawler Architecture*. In: Liu, Ling/Özsu, M. Tamer (Herausgeber). Encyclopedia of Database Systems. New York 2009, Seite 3462 ff.

Schnell, Rainer/Hill, Paul B./Esser, Elke. *Methoden der empirischen Sozialforschung*. 11. Auflage. Berlin 2018.

Schnell, Rainer/Redlich, Sarah. *Web Scraping Online Newspaper Death Notices for the Estimation of the Local Number of Deaths*. In: Proceedings of the 12th International Joint Conference on Biomedical Engineering Systems and Technologies. Volume 5: HEALTHINF. 2019, Seite 319 ff.

Soowamber, Medha L./Granton, John T./Bavaghar-Zaeimi, Fatemeh/Johnson, Sindhu R. *Online Obituaries Are a Reliable and Valid Source of Mortality Data*. In: Journal of Clinical Epidemiology. Ausgabe 79/2016, Seite 167 ff.

Sylvestre, Emmanuelle/Bouzille, Guillaume/Breton, Mathias/Cuggia, Marc/Campillo-Gimenez, Boris. *Retrieving the Vital Status of Patients with Cancer Using Online Obituaries*. In: Ugon, Adrien/Karlsson, Daniel/Klein, Gunnar O./Moen, Anne (Herausgeber). Building Continents of Knowledge in Oceans of Data: The Future of Co-Created eHealth. 2018, Seite 571 ff.

Thelwall, Mike/Stuart, David. *Web Crawling Ethics Revisited: Cost, Privacy, and Denial of Service*. In: Journal of the American Society for Information Science and Technology. Ausgabe 57(13)/2006, Seite 1771 ff.

LITERATURVERZEICHNIS

van Wel, Lita /Royakkers, Lambèr. *Ethical Issues in Web Data Mining*. In: Ethics and Information Technology. Ausgabe 6(2)/2004, Seite 129 ff.

Wallgren, Anders/Wallgren, Britt. *Register-based Statistics: Statistical Methods for Administrative Data*. 2. Auflage. Chichester 2014.

Webb, Eugene J./Campbell, Donald T./Schwartz, Richard D./Sechrest, Lee. *Unobtrusive Measures: Nonreactive Research in the Social Sciences*. Chicago 1966.

RECHTSGRUNDLAGEN

Bundesdatenschutzgesetz (BDSG) vom 30. Juni 2017 (BGBl. I Seite 2097), das durch Artikel 12 des Gesetzes vom 20. November 2019 (BGBl. I Seite 1626) geändert worden ist.

Gesetz über die Preisstatistik (PreisStatG) in der im Bundesgesetzblatt Teil III, Gliederungsnummer 720-9, veröffentlichten bereinigten Fassung, das zuletzt durch Artikel 1 des Gesetzes vom 10. Dezember 2019 (BGBl. I Seite 2117) geändert worden ist.

Gesetz über Urheberrecht und verwandte Schutzrechte (Urheberrechtsgesetz) vom 9. September 1965 (BGBl. I Seite 1273), das zuletzt durch Artikel 1 des Gesetzes vom 28. November 2018 (BGBl. I Seite 2014) geändert worden ist.

Personenstandsgesetz (PStG) vom 19. Februar 2007 (BGBl. I Seite 122), das zuletzt durch Artikel 17 des Gesetzes vom 20. November 2019 (BGBl. I Seite 1626) geändert worden ist.

Richtlinie 96/9/EG des Europäischen Parlaments und des Rates vom 11. März 1996 über den rechtlichen Schutz von Datenbanken (Amtsblatt der EG Nr. L 77 Seite 20).

Verordnung zur Ausführung des Personenstandsgesetzes (Personenstandsverordnung – PStV) vom 22. November 2008 (BGBl. I Seite 2263), die zuletzt durch Artikel 5 des Gesetzes vom 18. Dezember 2018 (BGBl. I Seite 2639) geändert worden ist.

EINFÜHRUNG DES EU-UNTERNEHMENSBEGRIFFS: KONZEPT UND UMSETZUNG

Martin Beck, Luisa Baumgärtner, Katja-Verena Bürk,
Matthias Redecker

➤ **Schlüsselwörter:** Unternehmen – Unternehmensstatistik – Strukturstatistik – Imputation – Konsolidierung – Profiling

ZUSAMMENFASSUNG

Dieser Beitrag beschreibt das Vorgehen der amtlichen Statistik in Deutschland bei der Einführung des EU-Unternehmensbegriffs in der Unternehmensstrukturstatistik. Zunächst werden der Anlass und das Umsetzungskonzept dargestellt. Darauf aufbauend wird die konkrete Umsetzung beschrieben, insbesondere das Profiling sowie die neuen, zentral durchzuführenden Arbeitsschritte Imputation und Konsolidierung.

➤ **Keywords:** *enterprise – business statistics – structural business statistics – imputation – consolidation – profiling*

ABSTRACT

This article describes the approach of official statisticians in Germany in introducing the EU enterprise concept in structural business statistics. It first explains the underlying reason and the implementation design. Based on this, the concrete implementation is described, in particular the profiling and the new work steps of imputation and consolidation, which have to be carried out centrally.

Martin Beck

ist Diplom-Ökonom und leitet im Statistischen Bundesamt die Gruppe „Unternehmensregister, Klassifikationen, Verdienste“ sowie die Steuerungsgruppe zur Einführung des EU-Unternehmensbegriffs. Er befasst sich derzeit unter anderem mit der effizienteren Gestaltung der Datengewinnung und -analyse durch die Einführung neuer statistischer Methoden und der Weiterentwicklung der Unternehmensstatistiken.

Luisa Baumgärtner

ist Betriebswirtin und als Referentin im Referat „Konjunktur der Industrie, Produktionsstatistik“ im Statistischen Bundesamt tätig. Im Rahmen des Projekts zur Einführung des europäischen Unternehmensbegriffs ist sie für die Entwicklung der Konsolidierungsmethodik für komplexe Unternehmen verantwortlich.

Katja-Verena Bürk

ist Ökonomin und Referentin im Referat „Handwerk, Gewerbezeugen, Insolvenzen“ im Statistischen Bundesamt. Sie beschäftigt sich mit der zentralen Umsetzung, Evaluierung und Automatisierung von Imputations- und Geheimhaltungsverfahren in den Unternehmensstrukturstatistiken.

Matthias Redecker

ist Diplom-Statistiker und als Referent im Referat „Profiling, Unternehmensgruppen, Methodik statistischer Einheiten“ des Statistischen Bundesamtes für das Fachthema Profiling zuständig. Er koordiniert die Einführung von Profiling im Statistischen Verbund, erstellt methodische Konzepte und entwickelt diese weiter.

1

Einleitung

Die Einführung der Unternehmensdefinition der Europäischen Union (EU) in der Strukturstatistik ab dem Berichtsjahr 2018 bedeutet eine der größten methodischen Änderungen in der Unternehmensstatistik der letzten Jahrzehnte. Die amtliche Statistik Deutschlands bereitet sie seit 2016 vor, Mitte 2020 soll sie mit der Veröffentlichung der Daten für das Berichtsjahr 2018 umgesetzt werden. Seit 2016 setzten sich vier Beiträge in dieser Zeitschrift mit der Thematik auseinander. Sturm/Redecker (2016) befassen sich mit der sogenannten Einheitendiskussion auf EU-Ebene und der fachlichen und politischen Relevanz der EU-Unternehmensdefinition. Redecker/Sturm (2017) stellen das Profiling als Methode zur Identifizierung der Unternehmen nach dem Konzept der EU vor. Den Aktionsplan und das Grobkonzept zur Einführung des EU-Unternehmensbegriffs in den Strukturstatistiken beschreiben Opfermann/Beck (2018). Baumgärtner und andere (2018) vertiefen diese Beschreibung und gehen insbesondere auf die Konzepte zur Imputation und Konsolidierung als notwendige neue Arbeitsschritte ein.

Darauf aufbauend erläutert dieser Beitrag das Vorgehen der amtlichen Statistik bei der Einführung des EU-Unternehmensbegriffs in der Unternehmensstrukturstatistik und stellt zunächst Anlass und Umsetzungskonzept dar. Danach wird die konkrete Umsetzung beschrieben, insbesondere das Profiling und die neuen, zentral durchzuführenden Arbeitsschritte Imputation und Konsolidierung.

2

Umsetzung des EU-Unternehmensbegriffs in Deutschland

In der deutschen amtlichen Statistik entspricht das Unternehmen bislang der kleinsten rechtlichen Einheit, die aus handels- und/oder steuerrechtlichen Gründen Bücher führt und Geschäftsabschlüsse aufstellt beziehungsweise über ähnliche Aufzeichnungen verfügt.

Die EU-Einheitenverordnung¹ gibt eine weitergehende Definition vor: Das Unternehmen entspricht der kleinsten Kombination rechtlicher Einheiten, die eine organisatorische Einheit zur Erzeugung von Waren und Dienstleistungen bilden und hinsichtlich der Verwendung der ihnen zufließenden laufenden Mittel über eine gewisse Entscheidungsfreiheit verfügen. Ein Unternehmen übt eine Tätigkeit oder mehrere Tätigkeiten an einem Standort oder an mehreren Standorten aus. Es kann einer einzigen rechtlichen Einheit entsprechen. Rechtliche Einheiten sind juristische und natürliche Personen, die eine Wirtschaftstätigkeit selbstständig ausüben, beispielsweise Aktiengesellschaften, GmbHs, Offene Handelsgesellschaften oder auch Einzelunternehmer.

Nach der EU-Definition kann ein Unternehmen also aus mehr als einer rechtlichen Einheit bestehen. Künftig können somit in der deutschen amtlichen Statistik Unternehmen auch weiterhin einer rechtlichen Einheit entsprechen (einfache Unternehmen) oder aus einer Kombination rechtlicher Einheiten bestehen (komplexe Unternehmen).

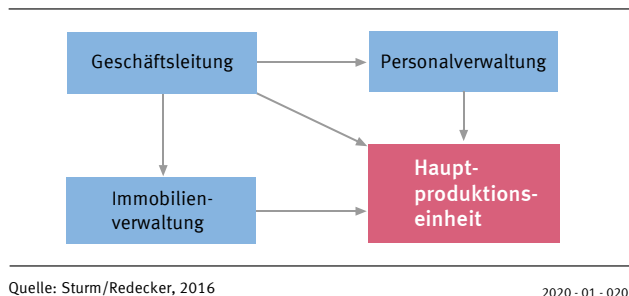
Der europäische Ansatz trägt der Entwicklung Rechnung, dass Unternehmen Teile ihres Tätigkeits- und Faktorenbündels rechtlich verselbstständigen („Outsourcing“), ohne dass dahinter ein tatsächlicher Strukturwandel steht. Das ist dann der Fall, wenn die ausgegliederten Einheiten in einem Unternehmensverbund weiter für die rechtliche Muttereinheit tätig sind, sie nicht oder kaum am Markt agieren und nicht autonom sind. Motive für derartige Ausgliederungen sind beispielsweise Steuern und Abgaben zu minimieren, die Haftung zu beschränken oder auch tarifliche Bindungen zu vermeiden. Die EU-Unternehmensdefinition zielt darauf ab, das vollständige Faktoren- und Funktionenbündel in einer Definition zusammenzuführen und damit auf europäischer Ebene vergleichbare Ergebnisse zu erhalten (Sturm/Redecker, 2016). Dahinter steht die Überlegung, dass Unternehmen aus für Branchen charakteristischen Tätigkeits-, Funktionen- und Faktorenbündeln bestehen (beispielsweise aus einer Kombination aus Produktions-, Handels-, Transport-, Lagerhaltungsfunktionen, Geschäftsleitung, Buchführung). Sturm/Redecker (2016) zeigen anschaulich ein schematisches Beispiel für ein solches

¹ Verordnung (EWG) Nr. 696/93 des Rates vom 15. März 1993 betreffend die statistischen Einheiten für die Beobachtung und Analyse der Wirtschaft in der Gemeinschaft (Amtsblatt der EG Nr. L 76, Seite 1).

komplexes Unternehmen, auf das hier zur Illustration zurückgegriffen wird. ➔ **Grafik 1**

Grafik 1

Schematisches Beispiel für ein komplexes Unternehmen



In diesem Fall weisen die deutschen Strukturstatistiken aktuell vier Unternehmen (= rechtliche Einheiten) mit unterschiedlichem wirtschaftlichem Schwerpunkt in verschiedenen Wirtschaftsbereichen nach: eine Produktionseinheit, die Geschäftsleitung, die Immobilienverwaltung und die Personalverwaltung (die drei zuletzt Genannten alle im Dienstleistungsbereich). Auf dem Markt tritt allerdings nur ein Akteur (Unternehmen) auf. Die anderen Leistungen werden lediglich intern bereitgestellt. Dies kommt einem Unternehmen gleich, bei dem diese Funktionen alle unter einem Dach gebündelt sind.

Mit der Anwendung der erweiterten Definition des Unternehmens lässt sich die Aussagekraft der Statistik verbessern. Das (komplexe) Unternehmen wird als vollständiges Bündel an Produktionsfaktoren und Hilfsfunktionen einer Branche zugeordnet und dort nachgewiesen. Dieses Vorgehen lässt eine realitätsnähere Beobachtung des Strukturwandels in der Wirtschaft zu, sowohl national als auch im europäischen Vergleich. Auch verbessert es beispielsweise die Analyse von kleinen und mittleren Unternehmen sowie die Messung der Unternehmenskonzentration. Aus diesem Grund ist die Anwendung der EU-Unternehmensdefinition nicht nur zwingende Folge verbindlicher europäischer Vorgaben, sondern verbessert auch die Qualität und Relevanz der Ergebnisse.

Für die in der Verordnung über die strukturelle Unternehmensstatistik² kodifizierte europäische Strukturstatistik

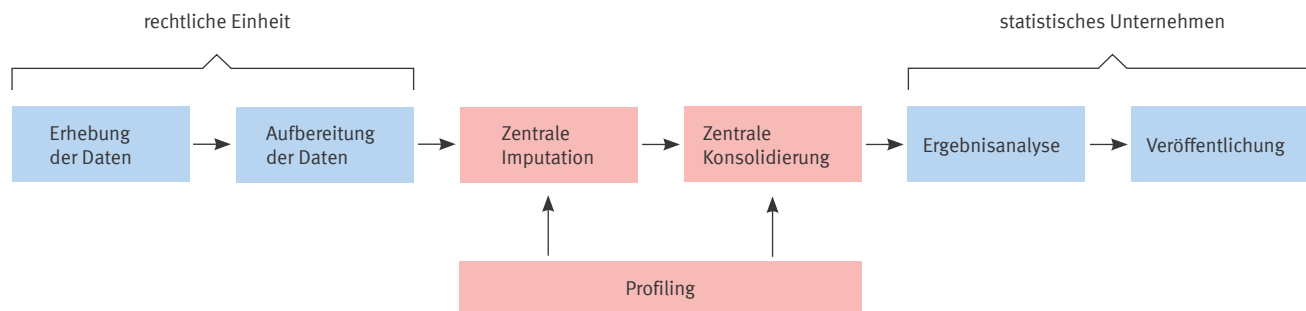
ist das Unternehmen in der Definition der EU-Einheitenverordnung („statistisches Unternehmen“) die den Ergebnissen zugrunde zu legende Darstellungseinheit. Die Praxis in Deutschland, aber auch in vielen anderen EU-Mitgliedstaaten, weicht bislang aus unterschiedlichen Gründen von der EU-Definition ab. Das Statistische Amt der Europäischen Union (Eurostat) verfolgt das Ziel, dies im Sinne einer relevanten und kohärenten europäischen Unternehmensstrukturstatistik zu überwinden und hat daher die Mitgliedstaaten zur Vorlage entsprechender Aktionspläne aufgefordert. Die Leitungen der statistischen Ämter in Deutschland haben im November 2015 einen Aktionsplan zur Anwendung des EU-Unternehmensbegriffs in der deutschen Strukturstatistik ab dem Berichtsjahr 2018 beschlossen.

Das in Deutschland entwickelte Umsetzungskonzept sieht weiterhin die rechtliche Einheit als Auswahl-, Befragungs- und Beobachtungseinheit vor. In einer späteren Aufbereitungsphase in den statistischen Ämtern werden dann Ergebnisse für die statistische Darstellungseinheit Unternehmen abgeleitet. Das Konzept behält die bisherigen Prozesse und zum Teil kodifizierten Konzepte der Statistikerstellung im Wesentlichen bei; es ermöglicht den statistischen Ämtern, für die Statistikerstellung wie bereits bislang Verwaltungsdaten zu verwenden. Zudem stehen wie bisher valide Ergebnisse für rechtliche Einheiten zur Verfügung, die von den Volkswirtschaftlichen Gesamtrechnungen des Bundes und der Länder sowie von anderen Datennutzern benötigt werden. Das neue Konzept führt absehbar zu keiner zusätzlichen Belastung der Auskunft gebenden Wirtschaft. Allerdings sind – aufbauend auf Informationen aus dem Profiling – die beiden neuen Arbeitsschritte Imputation und Konsolidierung inhaltlich, organisatorisch und zeitlich in die bestehenden Aufbereitungsprozesse der betroffenen zentralen und dezentralen Strukturstatistiken zu integrieren. ➔ **Grafik 2** auf Seite 38

² Verordnung (EG) Nr. 295/2008 des Europäischen Parlaments und des Rates vom 11. März 2008 über die strukturelle Unternehmensstatistik (Amtsblatt der EU Nr. L 97, Seite 13).

Grafik 2

Zusätzliche Arbeitsschritte in der Unternehmensstrukturstatistik: Profiling, Imputation und Konsolidierung



Die zusätzlichen Arbeitsschritte sind notwendig, um rechtliche Einheiten zu statistischen Unternehmen zusammenzufassen und ihre Merkmalswerte zu berechnen.

2020 - 01 - 0223

3

Von der rechtlichen Einheit zum Unternehmen

Auf Basis der Erhebungsergebnisse für rechtliche Einheiten sollen Daten für Unternehmen generiert werden, indem im Rahmen der Konsolidierung die internen Transaktionen identifiziert und herausgerechnet werden. Dazu ist es zum einen erforderlich, dass für alle rechtlichen Einheiten, die zu komplexen Unternehmen gehören, Einzeldaten vorliegen. Das ist jedoch aufgrund von Stichprobenerhebungen in den nationalen Strukturstatistiken nicht in jedem Fall gewährleistet, die fehlenden Angaben müssen daher durch Imputation geschätzt werden. Zum anderen werden Informationen darüber benötigt, welche rechtliche Einheiten überhaupt ein Unternehmen bilden und welchem Wirtschaftszweig es zugerechnet wird. Dies ist Aufgabe des Arbeitsschritts Profiling. Das Profiling liefert darüber hinaus weitere für die Konsolidierung relevante Informationen.

Die folgenden Abschnitte beschreiben die Weiterentwicklungen bezüglich Imputation, Profiling und Konsolidierung seit Ende 2018.³

3.1 Vorbereitung der Daten für rechtliche Einheiten

Imputation der rechtlichen Einheiten in komplexen Unternehmen

Die Methode der Imputation von Daten für rechtliche Einheiten in komplexen Unternehmen wurde bereits ausführlich beschrieben (Baumgärtner und andere, 2018, hier: Seite 38 ff.).

Die statistischen Ämter nutzen Imputationsverfahren, um fehlende Angaben in Erhebungen (meist für einzelne Merkmale bei sogenanntem Item-Nonresponse) zu vervollständigen. Bei der Einführung des EU-Unternehmensbegriffs wird der Einsatz der Imputation erweitert, indem die nicht erhobenen rechtlichen Einheiten in komplexen Unternehmen als Antwortausfall (Unit-Nonresponse) betrachtet und ihre relevanten Merkmale imputiert werden.⁴ Das gewählte Imputationsverfahren ist darauf ausgerichtet, eine hohe Qualität der Angaben zu den relevanten rechtlichen Einheiten zu gewährleisten. Die imputierten rechtlichen Einheiten sollten sich somit hinsichtlich ihrer Merkmale strukturell nicht von ähnlichen rechtlichen Einheiten unterscheiden, zu denen Angaben erhoben werden.

³ Der Ende 2018 erreichte Stand der Methodenentwicklung ist in Baumgärtner und andere (2018) detailliert beschrieben.

⁴ Dieses Vorgehen kann als „impute from a sample back to the frame“ bezeichnet werden (Krotki und andere, 2005).

Für die technische Umsetzung der Imputation wird die von Statistics Canada, dem nationalen statistischen Amt Kanadas, entwickelte Software CANCEIS (CANCEIS Development Team, 2015) verwendet.⁵ Es wird ein Hot-Deck-Nearest-Neighbour-Verfahren zur Imputation genutzt, das den vollständigen Merkmalskranz einer Spendereinheit übernimmt. Die Auswahlgrundlagen und die Erhebungsdaten der strukturellen Unternehmensstatistiken sind die Basis für die Imputation.

5 Die Abkürzung CANCEIS steht für „CANadian Census Edit and Imputation System“. Die Software und ihre Dokumentationen können über Statistics Canada erworben werden. CANCEIS wurde in Deutschland unter anderem bereits in den Verdiensterhebungen eingesetzt (Frentzen/Günther, 2017).

Durch die Kombination von Erhebung und Imputation liegen für alle rechtliche Einheiten in komplexen Unternehmen vollständige Merkmalskränze auf der Einzeldatenebene vor. Dies liefert eine vollständige Ausgangsbasis für die Konsolidierung von komplexen Unternehmen.

Hochrechnung der einfachen Unternehmen

In Deutschland werden die Erhebungen der Unternehmensstrukturstatistiken vom Statistischen Bundesamt (zentral) oder von den Statistischen Ämtern der Länder (dezentral) durchgeführt. Sie sind wirtschaftsbereichsspezifisch organisiert, und zwar in den Bereichen Produzierendes Gewerbe, Handel und Gastgewerbe sowie

Übersicht 1

Erhebungen der Unternehmensstrukturstatistiken

Wirtschaftsbereich	Erhebung	Art der Erhebung	Abschneidegrenze	Erhebungsdurchführung ¹
Produzierendes Gewerbe	Kostenstrukturhebung im Bereich Verarbeitendes Gewerbe, Bergbau und Gewinnung von Steinen und Erden	Stichprobe	Unternehmen mit 20 und mehr Tätigen Personen	zentral
	Investitionserhebung im Bereich Verarbeitendes Gewerbe, Bergbau und Gewinnung von Steinen und Erden	Totalerhebung	Unternehmen mit 20 und mehr Tätigen Personen	dezentral
	Strukturhebung für kleine Unternehmen im Bereich Verarbeitendes Gewerbe, Bergbau und Gewinnung von Steinen und Erden	Stichprobe	Unternehmen mit weniger als 20 Tätigen Personen	zentral
	Erhebung der laufenden Aufwendungen für den Umweltschutz im Produzierenden Gewerbe	Stichprobe	Unternehmen mit 50 und mehr Tätigen Personen	zentral
	Erhebung der Investitionen für den Umweltschutz im Produzierenden Gewerbe	Stichprobe	Unternehmen mit 20 und mehr Tätigen Personen	dezentral
	Kostenstrukturhebung im Bereich der Energie- und Wasserversorgung	Totalerhebung	Unternehmen der Wasserversorgung mit einer jährlichen Wasserabgabe von 200 000 m ³ und mehr	zentral
	Investitionserhebung im Bereich der Energie- und Wasserversorgung	Totalerhebung	Unternehmen der Wasserversorgung mit einer jährlichen Wasserabgabe von 200 000 m ³ und mehr	dezentral
	Kostenstrukturhebung im Baugewerbe	Stichprobe	Unternehmen mit 20 und mehr Tätigen Personen	zentral
	Jahreserhebung einschließlich Investitionserhebung im Bauhauptgewerbe	Totalerhebung	Unternehmen mit 20 und mehr Tätigen Personen	dezentral
	Jahreserhebung einschließlich Investitionserhebung im Ausbaugewerbe	Totalerhebung	Unternehmen mit 20 und mehr Tätigen Personen	dezentral
	Strukturhebung für kleine Unternehmen im Baugewerbe	Stichprobe	Unternehmen mit weniger als 20 Tätigen Personen	zentral
Handel und Gastgewerbe	Jahreserhebung im Handel sowie in der Instandhaltung und Reparatur von Kraftfahrzeugen und Gebrauchsgütern	Stichprobe	keine	dezentral (Kfz-Handel, Einzelhandel) zentral (Großhandel, Handelsvermittlung)
	Jahreserhebung im Gastgewerbe	Stichprobe	keine	dezentral
Dienstleistungen	Strukturhebung im Dienstleistungsbereich	Stichprobe	keine	dezentral

Quelle: Baumgärtner und andere (2018, Seite 37); eigene Ergänzungen

1 zentral: Statistisches Bundesamt; dezentral: Statistische Ämter der Länder.

Dienstleistungen. Dies ist historisch gewachsen. Für die einzelnen Branchen wurde eine an die jeweiligen Gegebenheiten und Informationsbedarfe angepasste Datenbereitstellung angestrebt. Daher gibt es in den Einzelstatistiken Unterschiede hinsichtlich der Methodik, des Inhalts, der Art und des Umfangs der Erhebungen. Zusammen ergeben sie jedoch ein Gesamtbild der abgedeckten Branchen der deutschen Wirtschaft, genauer der Wirtschaftsabschnitte B bis N (ausgenommen K) sowie S95. [↗ Übersicht 1](#)

Für eine verlässliche Ergebniserstellung bei der Umsetzung des EU-Unternehmensbegriffs ist eine Harmonisierung der Einzelstatistiken in größerem Maße als bisher erforderlich (Baumgärtner und andere, 2018, hier: Seite 36 f.). Hauptgrund dafür ist, dass komplexe Unternehmen aus rechtlichen Einheiten aus unterschiedlichen Wirtschaftsbereichen bestehen können. Deswegen sind die Auswahlgrundlagen wirtschaftsbereichs- und damit auch statistikübergreifend zu harmonisieren und unterschiedliche Hochrechnungsmethoden anzupassen. Somit ist es nicht mehr möglich, statistikspezifische Besonderheiten bei der Hochrechnung zu berücksichtigen, sodass eine einheitliche freie Hochrechnungsmethodik zur Anwendung kommt. Außerdem können rechtliche Einheiten, die den Wirtschaftsbereich wechseln, im Gesamtbild berücksichtigt werden; das war bisher in den Einzelstatistiken nicht systematisch möglich. Dies zusammen führt zu einer besseren Darstellung des Gesamtbilds der deutschen Wirtschaft. Es führt aber auch dazu, dass Veröffentlichungen von Strukturstatistiken nach rechtlichen Einheiten, die wirtschaftsbereichsbezogener orientiert sind, von den Veröffentlichungen nach statistischen Unternehmen abweichen.

Bei der neuen einheitlichen freien Hochrechnungsmethodik ist zu berücksichtigen, dass sich die Grundgesamtheit der rechtlichen Einheiten in zwei Kategorien aufteilt. Zum einen sind das die erhobenen beziehungsweise imputierten Daten für die komplexen Unternehmen, zum anderen die Erhebungsdaten für einfache Unternehmen. Aufgrund der Imputation liegen Angaben für alle komplexen Unternehmen vor. Sie bilden faktisch eine Totalschicht und tragen mit einem Hochrechnungsfaktor von 1 zu den endgültigen Ergebnissen bei. Einfache Unternehmen werden hingegen nicht imputiert und müssen auch weiterhin hochgerechnet werden. Ihr freier Hochrechnungsfaktor HRF_{eUnt} wird dabei wie folgt

ermittelt: Zuerst wird die Anzahl an einfachen Unternehmen je Schicht N_{rEeUnt} berechnet, indem die Anzahl der rechtlichen Einheiten in komplexen Unternehmen N_{rEkUnt} von der Anzahl rechtlicher Einheiten N_{rE} je Schicht abgezogen wird:

$$(1) \quad N_{rEeUnt} = N_{rE} - N_{rEkUnt}$$

Danach wird die Anzahl an einfachen Unternehmen je Schicht N_{rEeUnt} durch die Anzahl an erhobenen einfachen Unternehmen je Schicht n_{rEeUnt} geteilt und der Hochrechnungsfaktor der einfachen Unternehmen berechnet:

$$(2) \quad HRF_{eUnt} = \frac{N_{rEeUnt}}{n_{rEeUnt}}$$

Dadurch wird eine „Doppelberücksichtigung“ rechtlicher Einheiten, die zu komplexen Unternehmen gehören, vermieden.

Validierung der Imputation rechtlicher Einheiten in komplexen Unternehmen

Die Imputation rechtlicher Einheiten in komplexen Unternehmen „ersetzt“ im Prinzip deren Hochrechnung und damit ein in der amtlichen Statistik gebräuchliches und bewährtes Verfahren. Eine Validierung der Güte der Imputation ist daher essenziell und unerlässlich.

Anhand einer interaktiven Webapplikation können die Strukturen von imputierten und erhobenen Merkmalen verglichen werden (Baumgärtner und andere, 2018, hier: Seite 39 f.). Diese Webapplikation wird zur Validierung der Imputation genutzt. Bei auffälligen abweichenden Mustern in den Strukturen von erhobenen und imputierten Merkmalen können Spezifikationen der Imputation angepasst werden.

Zur weiteren Einschätzung der Güte der Imputation werden die Aggregatsergebnisse nach Imputation mit denen nach Hochrechnung verglichen. Die Spezifikation der Imputation ist darauf ausgerichtet, diese Abweichung möglichst gering zu halten. Jedoch erlaubt dieser Vergleich nur eine Schlussfolgerung darüber, inwieweit die Ergebnisse voneinander abweichen. Eine Aussage darüber, welcher Ansatz (Imputation oder Hochrechnung) für komplexe Unternehmen tatsächlich näher am „wahren“ Ergebnis liegt, kann nicht getroffen werden. Dies liegt daran, dass bei einer Stichprobenerhebung das wahre Ergebnis für die Grundgesamtheit nicht bekannt ist und geschätzt werden muss.

Um hier eine bessere Einschätzung der Güte der Imputation zu geben, wurden Simulationsrechnungen für das Berichtsjahr 2017 durchgeführt. Grundlage waren Substichproben, für die die gesuchten wahren Ergebnisse bekannt sind und die somit zur vergleichenden Bewertung von Imputation und Hochrechnung verwendet werden können.

Der Simulationsansatz wurde beispielhaft für die Aggregate der WZ-Zweisteller des Wirtschaftsabschnittes G - Handel⁶ umgesetzt. Dabei wurden die erhobenen Daten des Wirtschaftsabschnitts G als sogenannte „neue Grundgesamtheit“ definiert. Die Aggregate der Werte für die einzelnen Erhebungsmerkmale in den WZ-Zweistellern sind für diese der Simulation zugrunde liegende Grundgesamtheit bekannt und können berechnet werden. Um die konkrete methodische Herangehensweise bei Imputation und Hochrechnung realitätsnah zu simulieren, wird eine Substichprobe aus der neu definierten Grundgesamtheit gezogen, bei der das Verhältnis Sponder/Population dem realen Verhältnis möglichst nahe kommt. Die Substichprobe wird dann zum einen auf das Aggregat der bekannten Grundgesamtheit frei hochgerechnet. Zum anderen werden die nicht in der Substichprobe enthaltenen Einheiten imputiert, sodass auch hier ein Aggregat für die bekannte Grundgesamtheit gebildet wird. Die Aggregate nach Hochrechnung und Imputation können nun mit den tatsächlichen Aggregaten der neu definierten Grundgesamtheit verglichen werden. Um Zufallseinflüsse zu minimieren, wurde diese Simulation 100 Mal durchgeführt.

Im Ergebnis lassen sich über die 100 Simulationen keine eindeutigen Anzeichen dafür feststellen, dass die Imputation bei ausgewählten bedeutsamen Erhebungsmerkmalen in den betrachteten WZ-Zweistellern zu systematisch schlechteren Ergebnissen führt als die freie Hochrechnung. Dies bestätigt sowohl grundsätzlich die Wahl des Imputationsansatzes als auch dessen konkrete Umsetzung. ➡ Tabelle 1

Tabelle 1

Simulationsergebnisse: Durchschnittliche absolute Abweichung in % zur neu definierten Grundgesamtheit der Simulation über 100 Simulationswiederholungen

	WZ 45: Handel mit Kraftfahrzeugen		WZ 46: Großhandel		WZ 47: Einzelhandel	
	Durchschnittliche absolute Abweichung in % über 100 Simulationen					
	Imputation	Hochrechnung	Imputation	Hochrechnung	Imputation	Hochrechnung
Umsatz	0,80	0,94	1,00	0,69	0,34	0,56
Bruttowertschöpfung	0,95	0,75	1,67	0,60	0,59	0,61
Löhne	0,31	0,60	2,00	0,49	0,45	0,58
Bruttoinvestitionen	3,33	1,78	2,00	1,54	0,88	0,75
Beschäftigte	0,49	0,57	0,89	0,46	0,16	0,58

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

3.2 Profiling von Unternehmen

Die Methode „Profiling“ (beziehungsweise „Profilbildung“) beschreibt das Verfahren zur Identifizierung von Unternehmen gemäß der EU-Einheitenverordnung 696/93, das für Unternehmensgruppen angewendet wird. Es wird im EU-Empfehlungshandbuch zum Unternehmensregister unter dem Abschnitt 19 – Die Behandlung großer und komplexer Unternehmen (Eurostat, 2010) erwähnt:

«Das „Profiling“ ist ein Verfahren zur Analyse der rechtlichen und operationalen Struktur sowie der Rechnungslegungsstruktur einer Unternehmensgruppe auf nationaler und internationaler Ebene, um die statistischen Einheiten innerhalb der Gruppe und die Verbindungen zwischen ihnen sowie die effizientesten Strukturen zur Sammlung statistischer Daten zu ermitteln.»

Profiling wird im engen Sinne als manueller Arbeitsprozess verstanden und teilt sich hier auf die Teilprozesse **Desktop Profiling** und **Intensive Profiling** auf. Diese werden über Fallstudien von den Mitarbeiterinnen und Mitarbeitern der Statistischen Ämter der Länder im Fachbereich Unternehmensregister betreut.

Mit Blick auf den Aufwand in der Fallbearbeitung können nicht für alle Unternehmensgruppen differenzierte Untersuchungen durchgeführt werden, sodass eine Priorisierung auf die bedeutendsten Unternehmensgruppen vorgenommen werden muss. Für die Unternehmensstatistik ist es jedoch unumgänglich, eine vollständige Anpassung auf die neue Darstellungseinheit vorzunehmen, also für alle Unternehmensgruppen in Deutschland

⁶ 45: Handel mit Kraftfahrzeugen; Instandhaltung und Reparatur von Kraftfahrzeugen; 46: Großhandel (ohne Handel mit Kraftfahrzeugen); 47: Einzelhandel (ohne Handel mit Kraftfahrzeugen).

die Abgrenzung einfacher und komplexer Unternehmen zu ermitteln. Dies wird durch den Einsatz des **Automatic Profiling** sichergestellt. Hierbei werden automatisierte Algorithmen zur sinnvollen Zusammenfassung von rechtlichen Einheiten zu Unternehmen aufgebaut. Die Algorithmen untersuchen einerseits die Charakteristik von rechtlichen Einheiten mit Blick auf die Erfüllung bestimmter Funktionen (zum Beispiel Hilfstätigkeit) in der Unternehmensgruppe. Andererseits werden die Beziehungen zwischen bestimmten wirtschaftenden rechtlichen Einheiten (zum Beispiel Produktionseinheit und Handelseinheit, vertikale Integration zwischen Einheiten) analysiert, um gemeinsames wirtschaftliches Handeln abzuleiten. Die Identifikation der Zusammenhänge wird auf Basis von Register- und Erhebungsdaten vorgenommen.

Das Automatic Profiling wird für alle Unternehmensgruppen eingesetzt, für die kein manuelles Profil vorliegt. Diese Entscheidung wird unmittelbar nach Abschluss der „Feldphase“ für manuelles Profiling vorgenommen. Die Ergebnisse des Automatic Profiling werden analog zum manuellen Profiling im Unternehmensregister mit denselben Datenstrukturen festgehalten, sodass ein Wechsel in der Profiling-Methodik in Folgeberichts Jahren prinzipiell jederzeit möglich ist. Von einem Wechsel von der Methodik manuelles Profiling in Richtung Automatic Profiling wird aus fachlichen Gesichtspunkten aber meist abgesehen. Stattdessen werden die Vorjahresprofile manuell auf etwaigen Änderungsbedarf geprüft.

Für das Berichtsjahr 2017 ergibt sich für die Grundgesamtheit der Strukturstatistiken (WZ-Abschnitte B bis N ohne K sowie S95) eine Gesamtzahl von rund 208 000 profiliten Unternehmen, die sich in deutschen (Teilen von) Unternehmensgruppen befinden. Hierunter sind rund 40 000 komplexe Unternehmen (19 %). [↗ Tabelle 2](#)

Im Zusammenspiel mit den Aufgaben Imputation und Konsolidierung werden aus dem Unternehmensregister Datensätze zu profiliten Unternehmen und ihren zugehörigen rechtlichen Einheiten bereitgestellt und mit beschreibenden, profiling-spezifischen Merkmalen ergänzt. In der Gesamtbetrachtung ergibt sich eine Unterscheidung der Datenstrukturen in gruppenrelevante (= profiling-relevante) und nicht gruppenrelevante Unternehmen, zugeschnitten auf die Grundgesamtheit in den Strukturstatistiken. Die in Erhebungen einbezogenen rechtlichen Einheiten verteilen sich ungleichmäßig,

Tabelle 2

Profile Unternehmen in Unternehmensgruppen im Berichtsjahr 2017

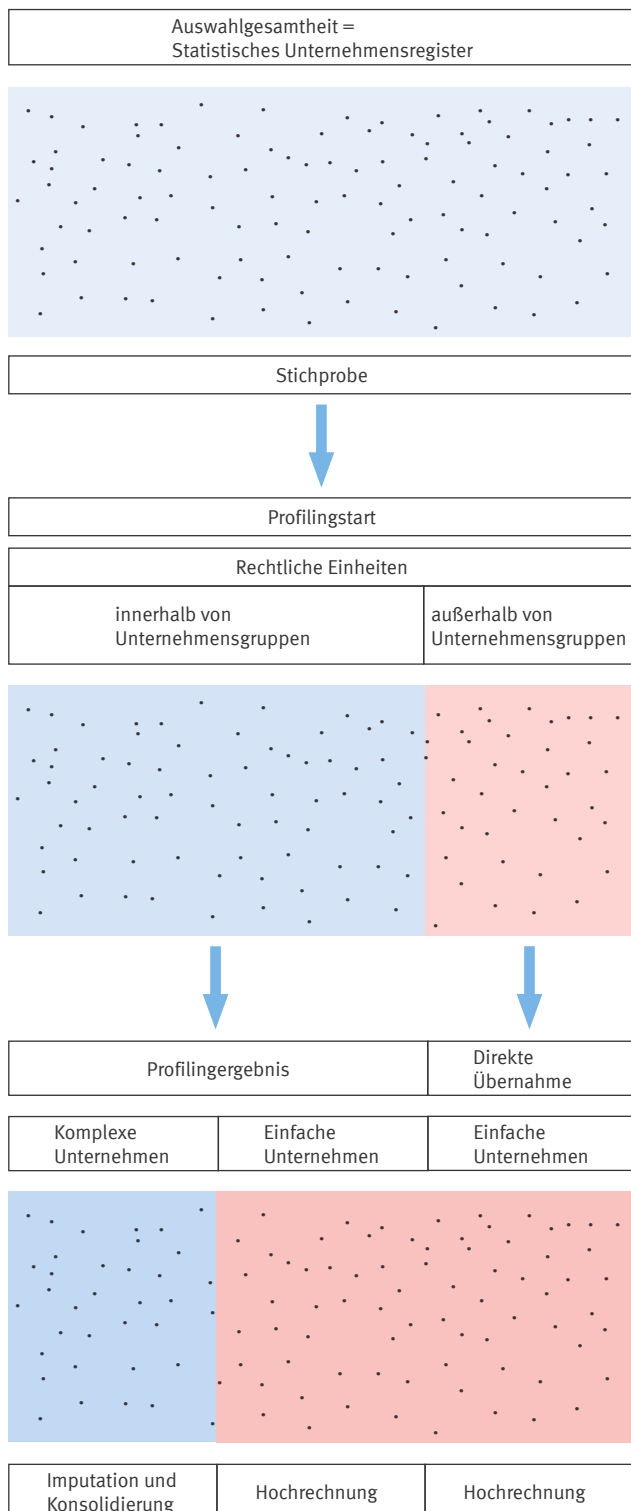
	Anzahl	%
1 rechtliche Einheit im Unternehmen (= einfache Unternehmen)	169 240	81,5
Komplexe Unternehmen:		
2 rechtliche Einheiten im Unternehmen	24 399	11,7
3 rechtliche Einheiten im Unternehmen	6 946	3,3
4 bis 5 rechtliche Einheiten im Unternehmen	4 195	2,0
6 bis 9 rechtliche Einheiten im Unternehmen	1 856	1,0
10 bis 19 rechtliche Einheiten im Unternehmen	816	0,4
20 bis 49 rechtliche Einheiten im Unternehmen	223	0,1
50 oder mehr rechtliche Einheiten im Unternehmen	58	> 0,0
Insgesamt	207 733	100

jedoch nicht aktiv gesteuert auf die beiden Subpopulationen. Die nicht gruppenrelevanten Unternehmen sind stets von einfacher Struktur, weshalb die strukturellen Merkmale entweder erhoben oder hochgerechnet werden. Die gruppenrelevanten Unternehmen sind einfach oder komplex und damit je nach Betrachtung ein Fall für die Imputation und die Konsolidierung. [↗ Grafik 3](#)

Durch das Profiling werden für den Arbeitsschritt Imputation Informationen darüber bereitgestellt, welche rechtliche Einheiten einfache Unternehmen bilden und welche zu komplexen Unternehmen gehören. Für die Konsolidierung steuert das Profiling Angaben über die Branche und den Sitz komplexer Unternehmen bei, außerdem über Rolle und Funktion rechtlicher Einheiten in komplexen Unternehmen, die für die konkreten Konsolidierungsschritte bedeutsam sind.

Grafik 3

Komponenten der Grundgesamtheit zur Aufbereitung von statistischen Unternehmen



2020 - 01 - 0224

3.3 Konsolidierung von komplexen Unternehmen

Grundsätzliche Überlegungen

Bei der Bereitstellung der für die europäischen Strukturstatistiken relevanten Merkmale sollen die Merkmalswerte nur die Marktaktivitäten der Unternehmen abbilden. Die rechtlichen Einheiten, die zu einem komplexen Unternehmen gehören, können jedoch auch untereinander in Austausch treten. Aus diesem Grund kann die einfache Addition der Merkmale der rechtlichen Einheiten, die weiterhin die Erhebungseinheiten darstellen, nicht in jedem Fall erfolgen. Ziel des Arbeitsschrittes Konsolidierung ist es, diese „internen“ Transaktionen bei der Berechnung der Merkmale für das komplexe Unternehmen zu eliminieren. Nach den Arbeitsschritten Profiling und Imputation, die beide unerlässliche Informationen für die Konsolidierung liefern, können die finalen Ergebnisse für Unternehmen durch die Konsolidierung erzeugt werden.

Da weder erhobene Merkmalswerte zum komplexen Unternehmen selbst noch Informationen über interne Transaktionen vorliegen, müssen die konsolidierten Merkmale durch das Treffen von Annahmen und das Durchführen bestimmter Rechenschritte geschätzt werden. Hierfür stehen die erhobenen und imputierten Merkmalswerte aller rechtlichen Einheiten eines komplexen Unternehmens zur Verfügung⁷.

Die entwickelten Rechenschritte zur Konsolidierung der Merkmalswerte für das statistische Unternehmen laufen für alle komplexen Unternehmen automatisiert in einem Programmalgorithmus in SAS⁸ ab. Eine Konsolidierung mit manuellen Schritten oder Zwischenschritten, beispielsweise für sehr große komplexe Unternehmen, findet bisher nicht statt.

⁷ Weitere, gegebenenfalls externe Informationsquellen werden für die Konsolidierung der Merkmale nicht verwendet. Grund hierfür ist, dass die Abgrenzung der Unternehmen in externen Datenquellen, wie beispielsweise Daten über Konzernabschlüsse, mit der statistischen Darstellungseinheit nur in sehr seltenen Fällen übereinstimmt.

⁸ SAS ist eine Software zur Datenauswertung, die Komponenten zum Datenmanagement, zur Durchführung von tabellarischen und grafischen Auswertungen sowie zur komplexen Datenanalyse mithilfe von multivariaten Verfahren umfasst. SAS wird in den Statistischen Ämtern des Bundes und der Länder als Standardsoftware verwendet.

Konsolidierung von zwei rechtlichen Einheiten

Die Merkmale von zwei rechtlichen Einheiten eines komplexen Unternehmens werden in drei Schritten konsolidiert. Das sind 1. die Bestimmung der Transaktionsarten, 2. die Wahl der Konsolidierungsmethode und 3. die Berechnung der Höhe der internen Transaktion. Die Methodenentwicklungen konzentrierten sich zunächst auf die Konsolidierung der Merkmale für zwei verbundene rechtliche Einheiten.⁹ Die folgende Darstellung geht vor allem auf neue Entwicklungen ein und bezieht sich auf die Konsolidierung einer beliebigen Menge an rechtlichen Einheiten in einem komplexen Unternehmen.

(1) Bestimmung der internen Transaktionsarten

Ein möglicher Grund, rechtliche Einheiten einem komplexen Unternehmen zuzuordnen, ist, dass sie bestimmte Funktionen in Bezug auf die wirtschaftliche Haupttätigkeit innerhalb des komplexen Unternehmens ausüben. Eine solche Zuordnung erfolgt also beispielsweise, wenn rechtliche Einheiten Waren oder Dienstleistungen für die sogenannten Haupteinheiten, die den wirtschaftlichen Schwerpunkt bestimmen, intern bereitstellen. Hierbei werden zwei grundlegende Funktionen unterschieden: Einheiten mit Hilfstätigkeiten und vertikal integrierte Einheiten. Zudem kann es Einheiten im komplexen Unternehmen geben, die weder eine Hilfstätigkeit noch eine vertikal integrierte Tätigkeit ausüben. Solche Einheiten werden im Profiling aufgrund anderer Kriterien einem komplexen Unternehmen zugeordnet.

9 Baumgärtner und andere (2018) geben erste Einblicke in die Methoden der Konsolidierung nach dem damaligen Stand der Entwicklung.

Übersicht 2

Funktionen von rechtlichen Einheiten im komplexen Unternehmen

Funktion	Beschreibung
Haupteinheiten	Sie bestimmen den wirtschaftlichen Schwerpunkt des statistischen Unternehmens.
Vertikal integrierte Tätigkeiten	Diese Einheiten können beispielsweise Vorleistungen für die Haupteinheit erbringen (vorgelagerte Tätigkeit) oder die von der Haupteinheit hergestellten Waren am Markt verkaufen (nachgelagerte Tätigkeit).
Hilfseinheiten	Das sind meist Einheiten aus dem Dienstleistungsbereich, die bestimmte Dienstleistungen für die Haupteinheit(en) erbringen.
Sonstige Einheiten	Sie können keiner der oben genannten Tätigkeiten zugeordnet werden, sind aber dennoch Teil des statistischen Unternehmens (z.B. wirtschaftliche Nebentätigkeiten).

Hierzu zählen rechtliche Einheiten, die Nebentätigkeiten ausüben oder der wirtschaftlichen Kontrolle der Haupteinheit unterliegen (Redecker/Sturm, 2017, hier: Seite 16). Es werden demnach vier verschiedene Funktionen von rechtlichen Einheiten im komplexen Unternehmen unterschieden, die für die Annahme von internen Transaktionen bedeutend sind. [↘ Übersicht 2](#)

Informationen über die Funktionen von rechtlichen Einheiten werden beim Profiling gewonnen und in Form von Kennzeichnungen in den Profiling-Daten an den Arbeitsschritt Konsolidierung weitergegeben. Diese Funktionen geben einen Hinweis, ob interne Transaktionen zwischen den Einheiten stattfinden oder nicht. Bei rechtlichen Einheiten mit vertikal integrierten Tätigkeiten ist gekennzeichnet, ob es sich um eine vor- oder nachgelagerte Tätigkeit handelt, da die Richtung der Waren- und Dienstleistungsströme entscheidend für die Konsolidierung ist (Baumgärtner und andere, 2018, hier: Seite 42). Sind rechtliche Einheiten als Hilfstätigkeit oder als vertikal integrierte Tätigkeit gekennzeichnet, so wird angenommen, dass Transaktionen in Form von Waren oder Dienstleistungen mit der Haupteinheit stattfinden. Für alle sonstigen Einheiten, die weder Hilfstätigkeiten ausüben noch vertikal integriert sind, werden keine internen Ströme mit der Haupteinheit vermutet. Hierbei werden alle Merkmale aufsummiert, da keine internen Transaktionen herausgerechnet werden müssen (= additive Konsolidierung).

Um die Art der internen Transaktion zu bestimmen, reicht die Kenntnis über die Funktion der rechtlichen Einheit allein nicht aus. Erst in Kombination mit dem Wirtschaftszweig kann die Art der internen Transaktion mit der Haupteinheit abgeleitet werden. [↘ Übersicht 3](#) zeigt beispielhaft und nicht abschließend mögliche Arten von internen Transaktionen, die anhand der Funktion und des Wirtschaftszweigs einer rechtlichen Einheit (=Typen von rechtlichen Einheiten) bestimmt werden. Die Bestimmung der Transaktionsart stellt eine Annahme dar, die aufgrund der vorhandenen Informationen am wahrscheinlichsten ist. Alle anderen möglichen Transaktionen, die über die direkte Herleitung aus der gekennzeichneten Funktion und dem Wirtschaftszweig hinausgehen, werden nicht betrachtet.

Bei den Hilfstätigkeiten können bezüglich des Wirtschaftszweigs zum Teil sehr feine Unterscheidungen der Transaktionsarten vorgenommen werden (Beispiel:

Übersicht 3

Bestimmung der Transaktionsart nach Typen von rechtlichen Einheiten

Typ der rechtlichen Einheit		Interne Transaktion mit der Haupteinheit	Art der Transaktion mit der Haupteinheit	
Funktion	Wirtschaftszweig (WZ)			
Haupt-einheiten	WZ-unabhängig (= WZ des statistischen Unternehmens)			
Vertikale Integration	vorgelagert	WZ G – Handel	ja	Es handelt sich um Einkaufseinheiten. Es werden Handelswaren am Markt eingekauft und intern an die Haupteinheit weiterverkauft.
		WZ B bis F – Produzierendes Gewerbe	ja	Die Einheit produziert Vorleistungsgüter für eine Haupteinheit. Es werden die eigenen Erzeugnisse intern an die Haupteinheit verkauft.
	nachgelagert	WZ G – Handel	ja	Es handelt sich hierbei um eine Verkaufs- oder Vertriebsseinheit. Diese Einheit kauft intern Waren oder Dienstleistungen (abhängig vom Wirtschaftszweig der Haupteinheit) der Haupteinheit und verkauft diese am Markt.
Hilfstätigkeiten	WZ 68 – Grundstücks- und Wohnungswesen	ja	Es wird die dem Wirtschaftszweig entsprechende Dienstleistung der Haupteinheit zur Verfügung gestellt.	
	WZ 70 – Verwaltung und Führung von Unternehmen und Betrieben	ja	Es wird die dem Wirtschaftszweig entsprechende Dienstleistung der Haupteinheit zur Verfügung gestellt.	
	WZ 78 – Vermittlung und Überlassung von Arbeitskräften	ja	Es wird die dem Wirtschaftszweig entsprechende Dienstleistung der Haupteinheit zur Verfügung gestellt.	
Sonstige Einheiten	WZ-unabhängig	nein		

Überlassung von Arbeitskräften). Dies hängt damit zusammen, dass für diese speziellen Dienstleistungen in den Kostenstrukturerhebungen genaue Kostenpositionen für Dienstleistungstätigkeiten vorliegen, die für die Berechnung des Werts der internen Transaktionen verwendet werden können.

(2) Konsolidierungsmethode

Bei der gewählten Konsolidierungsmethode (Baumgärtner und andere, 2018, hier: Seite 42) geht es um die Einteilung der Merkmale in additiv und nicht additiv. Bei den nicht additiven Merkmalen wird die Summe der Merkmalswerte beider Einheiten um den Wert der internen Transaktion gemindert. Hingegen können bei den additiven Merkmalen die Werte zweier rechtlicher Einheiten aufsummiert werden, da hier keine internen Transaktionen angenommen werden oder diese nicht relevant sind.

(3) Höhe der internen Transaktion

Eine Transaktion von Waren oder Dienstleistungen zwischen zwei rechtlichen Einheiten verursacht die Verbuchung von Erträgen bei der einen und die Verbuchung von Aufwänden bei der anderen rechtlichen Einheit.¹⁰

¹⁰ Wie der Wert der internen Transaktion aus den vorhandenen Ertrags- und Aufwandsvariablen der Strukturstatistiken geschätzt werden kann, zeigt ein Beispiel in Baumgärtner und andere (2018, hier: Seite 42 f.).

Welche konkreten Ertrags- und Aufwandspositionen dabei betrachtet werden, gibt die Annahme über die Art der Transaktion vor. Als zu eliminierender Betrag der internen Transaktion wird das Minimum dieser Aufwands- und Ertragsgrößen angesetzt. Dabei wird unterstellt, dass die Differenz beider Größen am Markt realisiert wird. Sie geht in das konsolidierte Gesamtergebnis des komplexen Unternehmens ein.

Konsolidierung mehrerer rechtlicher Einheiten

Die Mehrheit der komplexen Unternehmen besteht aus zwei rechtlichen Einheiten und kann somit in einem Schritt konsolidiert werden. Jedoch können gerade aus sehr großen Unternehmensgruppen, die manuell profitiert werden, komplexe Unternehmen mit zum Teil 50 oder mehr rechtlichen Einheiten hervorgehen (siehe Tabelle 2). Um eine einfache schrittweise Konsolidierung mit den oben skizzierten Methoden vornehmen zu können, wurde seit 2018 ein automatisiertes Verfahren entwickelt. Im Folgenden wird die schrittweise Konsolidierung mehrerer rechtlicher Einheiten an einem Beispiel erklärt. ➔ Grafik 4

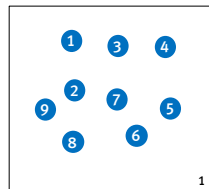
Im ersten Schritt werden rechtliche Einheiten, die dem gleichen Typ (siehe Übersicht 3) entsprechen, zunächst additiv konsolidiert. Hierbei werden die Merkmale aller Einheiten des gleichen Typs addiert (= additive Konsolidierung von Einheiten) und bilden eine neue Teileinheit

Grafik 4

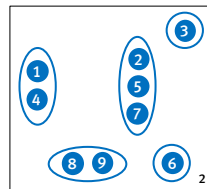
Beispiel zur Konsolidierung eines komplexen Unternehmens mit WZ 21.1

Nr.	Funktion	Wertschöpfungsstufe	Wirtschaftszweig
1	Vorleistung (V)	20.5	Herstellung von sonstigen chemischen Erzeugnissen
2	Haupteinheit (H)	21.1	Herstellung von pharmazeutischen Grundstoffen
3	Sonstige Einheit (S)	20.6	Herstellung von Chemiefasern
4	Vorleistung (V)	20.5	Herstellung von sonstigen chemischen Erzeugnissen
5	Haupteinheit (H)	21.1	Herstellung von pharmazeutischen Grundstoffen
6	Hilfstätigkeit (Hilf)	78.2	Befristete Überlassung von Arbeitskräften
7	Haupteinheit (H)	21.1	Herstellung von pharmazeutischen Grundstoffen
8	Hilfstätigkeit (Hilf)	68.2	Vermietung, Verpachtung von eigenen oder geleasteten Grundstücken, Gebäuden und Wohnungen
9	Hilfstätigkeit (Hilf)	68.2	Vermietung, Verpachtung von eigenen oder geleasteten Grundstücken, Gebäuden und Wohnungen

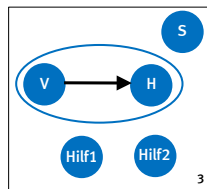
Ausgangssituation:
9 rechtliche Einheiten
in einem komplexen
Unternehmen



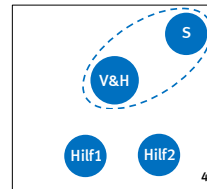
Additive Konsolidierung
der rechtlichen Einheiten
des gleichen Typs



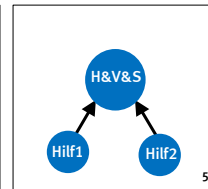
Konsolidierung vertikal
integrierter Einheiten



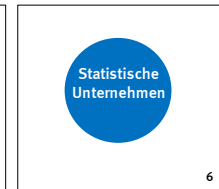
Additive Konsolidierung
der sonstigen Einheiten



Konsolidierung der
Hilfseinheiten



Ergebnis: konsolidierte
Merkmale für das
gesamte statistische
Unternehmen



2020 - 01 - 0225

auf dem Weg zum komplexen Unternehmen. Im Beispiel in der Grafik 4 sind es neun rechtliche Einheiten (1. Bild). Hierbei werden jeweils drei Haupteinheiten (2 + 5 + 7), zwei Vorleistungseinheiten aus dem Verarbeitenden Gewerbe (1 + 4) und zwei Hilfseinheiten mit dem Wirtschaftszweig WZ 68.2 (8 + 9) additiv zusammengefasst. Hintergrund der additiven Zusammenfassung der Einheiten des gleichen Typs ist, dass kein Anlass zur Annahme von Transaktionen zwischen diesen Einheiten besteht. Zwei weitere Typen (Einheiten 3 und 6) kommen nur einmal vor und verbleiben zunächst in ihrem ursprünglichen Zustand. Durch die additive Konsolidierung von Einheiten gleichen Typs wird der Komplexitätsgrad des Unternehmens auf einfache Art und Weise reduziert, sodass das komplexe Unternehmen nur noch auf wenige (Teil-)Einheiten beschränkt wird. Im Beispiel werden neun rechtliche Einheiten auf fünf Teileinheiten reduziert (2. Bild in Grafik 4).

Im nächsten Schritt werden vertikal integrierte Einheiten konsolidiert (3. Bild in Grafik 4). Im Beispiel sind dies Einheiten mit vertikal integrierten Vorleistungen und Haupteinheiten.

Im Anschluss an die Konsolidierung der vertikal integrierten Einheiten werden die „sonstigen Einheiten“, die keinerlei Rückschlüsse auf interne Transaktionen geben, additiv konsolidiert (4. Bild in Grafik 4). Dieser Schritt wird vor der Konsolidierung der Hilfseinheiten vorgenommen. Grund hierfür ist die Annahme, dass Hilfseinheiten, die Dienstleistungen innerhalb des komplexen Unternehmens bereitstellen, ihre Dienste für alle Einheiten der Gruppe anbieten (zum Beispiel auch für wirtschaftliche Nebenaktivitäten), und nicht ausschließlich für die Haupteinheiten.

Die Hilfseinheiten werden am Ende mit dem Ergebnis aus den Schritten der Bilder 2 bis 4 konsolidiert (5. Bild in Grafik 4). Unterschiedliche Typen von Hilfseinheiten können parallel konsolidiert werden, da hierbei unterschiedliche Kostenpositionen der bereits konsolidierten Teileinheit für die Berechnung der internen Transaktionen relevant sind. Aus diesen Konsolidierungsschritten resultiert als Ergebnis das statistische Unternehmen (6. Bild in Grafik 4).

Validierung der Ergebnisse der Konsolidierung

Eine Validierung der Konsolidierungsmethoden ist zur Qualitätssicherung der vorgenommenen Annahmen und Schätzungen sehr wichtig. Sie erfolgt anhand von Prüfroutinen, die normalerweise auf der Ebene von Wirtschaftszweigaggregaten genutzt werden, und die Eurostat zur Validierung der Daten anwendet. Die Regeln umfassen die Einhaltung von Größenverhältnissen bestimmter Variablen, die entweder allgemein oder auch wirtschaftsbereichsspezifisch gültig sind. Beispielsweise wird geprüft, ob der Produktionswert kleiner als der Umsatz ist oder der Anteil der Löhne und Gehälter an den Personalkosten zwischen 70 und 90% liegt. Das entwickelte Verfahren ist insofern strenger, da diese Prüfroutinen auf der Ebene der einzelnen komplexen Unternehmen angewendet werden. Diese Prüfroutinen bestätigen die bisherigen Methoden (Baumgärtner und andere, 2018, hier: Seite 43).

4

Ausblick

Ein weiterer Beitrag, der ebenfalls in Ausgabe 3/2020 dieser Zeitschrift publiziert ist, analysiert die zu erwartenden Auswirkungen auf die Ergebnisse der Strukturstatistiken. Grundlage hierfür sind Testläufe und Simulationen für das Berichtsjahr 2017. [u](#)

LITERATURVERZEICHNIS

Baumgärtner, Luisa/Gräb, Christopher/Leppert, Philipp/Söllner, René/Spies, Lydia/Veith, Stefan/Vorgrimler, Daniel. [Imputation und Konsolidierung: Neue Aufgaben für die Unternehmensstatistik](#). In: WISTA Wirtschaft und Statistik. Ausgabe 6/2018, Seite 33 ff.

Beck, Martin/Baumgärtner, Luisa/Bürk, Katja-Verena/Redecker, Matthias. [Auswirkungen der Einführung des EU-Unternehmensbegriffs](#). In: WISTA Wirtschaft und Statistik. Ausgabe 3/2020, Seite 49 ff.

CANCEIS Development Team. *CANCEIS User's Guide. Version 5.2*. Ottawa 2015.

Frentzen, Kathrin/Günther, Roland. [Korrektur des Antwortausfalls in der Verdiensterhebung 2015](#). In: WISTA Wirtschaft und Statistik. Ausgabe 2/2017, Seite 24 ff.

Eurostat. *Business registers – Recommendations manual: Chapter 19 – The handling of large and complex businesses*. Ausgabe 2010. [Zugriff am 27. Mai 2020]. Verfügbar unter: <https://ec.europa.eu>

Krotki, Karol/Black, Stephen/Creel, Darryl. *Mass Imputation*. ASA Section on Survey Research Methods. 2005.

Opfermann, Rainer/Beck, Martin. [Einführung des EU-Unternehmensbegriffs](#). In: WISTA Wirtschaft und Statistik. Ausgabe 1/2018, Seite 63 ff.

Redecker, Matthias/Sturm, Roland. [Profiling von Unternehmen](#). In: WISTA Wirtschaft und Statistik. Ausgabe 6/2017, Seite 9 ff.

Sturm, Roland/Redecker, Matthias. [Das EU-Konzept des Unternehmens](#). In: WISTA Wirtschaft und Statistik. Ausgabe 3/2016, Seite 57 ff.

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.

Verordnung (EG) Nr. 295/2008 des Europäischen Parlaments und des Rates vom 11. März 2008 über die strukturelle Unternehmensstatistik (Amtsblatt der EU Nr. L 97, Seite 13).

Verordnung (EWG) Nr. 696/93 des Rates vom 15. März 1993 betreffend die statistischen Einheiten für die Beobachtung und Analyse der Wirtschaft in der Gemeinschaft (Amtsblatt der EG Nr. L 76, Seite 1).

AUSWIRKUNGEN DER EINFÜHRUNG DES EU-UNTERNEHMENSBEGRIFFS

Martin Beck, Luisa Baumgärtner, Katja-Verena Bürk,
Matthias Redecker

📖 **Schlüsselwörter:** Unternehmen – Unternehmensstatistik – Strukturstatistik – Imputation – Konsolidierung – Profiling

ZUSAMMENFASSUNG

Im Zuge der Einführung des EU-Unternehmensbegriffs in der Unternehmensstrukturstatistik in Deutschland wurden das Profiling sowie die neuen, zentral durchzuführenden Arbeitsschritte Imputation und Konsolidierung integriert. Testläufe und Simulationen für das Berichtsjahr 2017 sollten sichtbar machen, welche Auswirkungen auf die Ergebnisse der Strukturstatistiken dadurch zu erwarten sind. Der Beitrag analysiert die Ergebnisse des Testlaufs für statistische Unternehmen, beschreibt das geplante Vorgehen bei der Veröffentlichung der Daten ab dem Berichtsjahr 2018 und gibt einen Ausblick auf Folgearbeiten.

📖 **Keywords:** *enterprise – business statistics – structural business statistics – imputation – consolidation – profiling*

ABSTRACT

As part of the introduction of the EU enterprise concept in structural business statistics in Germany, profiling has been incorporated as well as the new work steps of imputation and consolidation, which have to be carried out centrally. Pilot runs and simulations were made for the reference year 2017 to identify the anticipated impacts on the results of structural business statistics. This contribution analyses the results of the pilot run concerning statistical enterprises, describes the plans for publishing the data from the reference year 2018 onwards, and provides an outlook on future work.

Martin Beck

ist Diplom-Ökonom und leitet im Statistischen Bundesamt die Gruppe „Unternehmensregister, Klassifikationen, Verdienste“ sowie die Steuerungsgruppe zur Einführung des EU-Unternehmensbegriffs. Er befasst sich derzeit unter anderem mit der effizienteren Gestaltung der Datengewinnung und -analyse durch die Einführung neuer statistischer Methoden und der Weiterentwicklung der Unternehmensstatistiken.

Luisa Baumgärtner

ist Betriebswirtin und als Referentin im Referat „Konjunktur der Industrie, Produktionsstatistik“ im Statistischen Bundesamt tätig. Im Rahmen des Projekts zur Einführung des europäischen Unternehmensbegriffs ist sie für die Entwicklung der Konsolidierungsmethodik für komplexe Unternehmen verantwortlich.

Katja-Verena Bürk

ist Ökonomin und Referentin im Referat „Handwerk, Gewerbeanzeigen, Insolvenzen“ im Statistischen Bundesamt. Sie beschäftigt sich mit der zentralen Umsetzung, Evaluierung und Automatisierung von Imputations- und Geheimhaltungsverfahren in den Unternehmensstrukturstatistiken.

Matthias Redecker

ist Diplom-Statistiker und als Referent im Referat „Profiling, Unternehmensgruppen, Methodik statistischer Einheiten“ des Statistischen Bundesamtes für das Fachthema Profiling zuständig. Er koordiniert die Einführung von Profiling im Statistischen Verbund, erstellt methodische Konzepte und entwickelt diese weiter.

1

Einleitung

Mit der Einführung der Unternehmensdefinition der Europäischen Union (EU) setzt Deutschland eine der größten methodischen Änderungen in der Unternehmensstatistik der letzten Jahrzehnte ab dem Berichtsjahr 2018 um. Seit 2016 behandelten vier Beiträge in dieser Zeitschrift diese Thematik und dokumentieren die Vorarbeiten der deutschen amtlichen Statistik: die sogenannte Einheitendiskussion auf EU-Ebene und die fachliche und politische Relevanz der EU-Unternehmensdefinition (Sturm/Redecker, 2016), das Profiling als Methode zur Identifizierung der Unternehmen nach dem Konzept der EU (Redecker/Sturm, 2017), den Aktionsplan und das Grobkonzept zur Einführung des EU-Unternehmensbegriffs in den Strukturstatistiken (Opfermann/Beck, 2018) sowie die Konzepte zur Imputation und Konsolidierung als notwendige neue Arbeitsschritte (Baumgärtner und andere, 2018).

Danach hat ein ebenfalls in Ausgabe 3/2020 dieser Zeitschrift publizierter Beitrag zunächst Anlass und Umsetzungskonzept dargestellt und die konkrete Umsetzung, insbesondere des Profiling und der neuen, zentralen Arbeitsschritte Imputation und Konsolidierung beschrieben.

Der folgende Beitrag analysiert zunächst in Kapitel 2 die zu erwartenden Auswirkungen auf die Ergebnisse der Strukturstatistiken durch die neue Unternehmensdefinition auf der Grundlage von Testläufen und Simulationen für das Berichtsjahr 2017. Kapitel 3 beschreibt, wie die Daten zu statistischen Unternehmen ab dem Berichtsjahr 2018 veröffentlicht werden und für welche Zwecke die Nutzung dieser Daten geeignet ist. Der Aufsatz schließt mit einem Fazit und einem Ausblick auf wichtige Folgearbeiten.

2

Ergebnisse nach statistischen Unternehmen

Alle nachfolgenden Ergebnisse beruhen auf einem Testlauf, der auf Grundlage von Informationen aus dem statistischen Unternehmensregister (Profiling) und der Daten der Strukturstatistiken des Berichtsjahrs 2017 durchgeführt wurde. Alle im Testlauf verwendeten Abläufe und Methoden sind weitgehend abgeschlossen, sodass dieselben Methoden auch für das Berichtsjahr der ersten Veröffentlichung (2018) verwendet werden können und keine methodenbedingten Brüche in den Ergebnissen von 2017 und 2018 zu erwarten sind.

2.1 Effekte bei der Erstellung der Ergebnisse für Unternehmen

Bei der Darstellung der Ergebnisse sind nicht vorrangig die absoluten Werte als Gesamtergebnis interessant, sondern vor allem der Vergleich der Ergebnisse nach Unternehmen und nach rechtlichen Einheiten. Dieser Vergleich ist wichtig, um die im Berichtsjahr 2018 parallel veröffentlichten Ergebnisse für die beiden Darstellungseinheiten richtig zu verstehen und zu interpretieren.

Hierbei spielen insgesamt vier Effekte eine Rolle, die letztlich zu unterschiedlichen Ergebnissen für die beiden Darstellungseinheiten führen. Die Effekte entstehen durch die verschiedenen Arbeitsschritte von der rechtlichen Einheit hin zum Ergebnis nach statistischen Unternehmen (Beck und andere, 2020, hier: Kapitel 3):

› Imputationseffekt:

Der Imputationseffekt beschreibt die Abweichung der Merkmale, die sich durch die Imputation von rechtlichen Einheiten in komplexen Unternehmen im Vergleich zur bisherigen Hochrechnung ergeben.

› Hochrechnungseffekt:

Dieser Effekt zeigt die Abweichung der Merkmale, die durch eine gesamtwirtschaftliche Hochrechnung im Vergleich zu einer wirtschaftsbereichsspezifischen Hochrechnung bei den einfachen Unternehmen entsteht.

› Profilingeffekt:

Der Profilingeffekt beschreibt Verschiebungen zwischen den Wirtschaftsbereichen und Regionen (zum Beispiel Bundesländern) aufgrund der Zuordnung von rechtlichen Einheiten (und deren Merkmalswerten) zu komplexen Unternehmen. Der Effekt bewirkt für sich betrachtet eine Verringerung der Anzahl der Unternehmen. Er spielt bei einer gesamtwirtschaftlichen Betrachtung von Merkmalen keine Rolle.

› Konsolidierungseffekt:

Der Konsolidierungseffekt veranschaulicht, wie sich Merkmalswerte aufgrund der Herausrechnung von internen Transaktionen bei nicht additiven Merkmalen (das sind beispielsweise der Gesamtumsatz oder die gesamten Aufwendungen für Waren und Dienstleistungen; siehe Baumgärtner und andere, 2018) reduzieren. Für sich betrachtet bewirkt er eine Verringerung der Werte bei nicht additiven Merkmalen und gleichbleibende Werte bei additiven Merkmalen (wie der Anzahl der Beschäftigten oder den Bruttoinvestitionen in Sachanlagen).

Je nach Typ des Unternehmens (einfach oder komplex) und Art des Merkmals wirken sich die vier Effekte aus oder nicht. [↗ Übersicht 1](#)

Übersicht 1

Übersicht über die Effekte bei der Erstellung der Ergebnisse für Unternehmen

Merkmal	Einfache Unternehmen	Komplexe Unternehmen
Anzahl der Unternehmen	› Hochrechnungseffekt	› Profilingeffekt
Additive Merkmale	› Hochrechnungseffekt	› Profilingeffekt › Imputationseffekt
Nicht additive Merkmale	› Hochrechnungseffekt	› Profilingeffekt › Imputationseffekt › Konsolidierungseffekt

2.2 Ergebnisse des Berichtsjahrs 2017

Der Anteil der komplexen Unternehmen bezogen auf die Gesamtanzahl der Unternehmen mit einem für die Strukturstatistiken relevanten Wirtschaftsabschnitt beträgt 1,5 %. Jedoch generieren diese Unternehmen 60 % des Gesamtumsatzes (konsolidierter Umsatz) und verfügen über fast 40 % aller Beschäftigten. [↗ Tabelle 1](#)

Tabelle 1

Eckwerte für einfache und komplexe Unternehmen 2017

	Insgesamt	Einfache Unternehmen	Komplexe Unternehmen
		%	
Anzahl an Unternehmen	2 650 987	98,5	1,5
Anzahl der Beschäftigten	30 762 492	61,4	38,6
Umsatz (Mill. EUR)	6 564 206	39,6	60,4

Gesamtwirtschaftliche Ergebnisse

Gesamtwirtschaftlich betrachtet spielt vor allem der Konsolidierungseffekt eine Rolle. Bei allen nicht additiven Merkmalen reduziert sich wie zu erwarten das Ergebnis für die Darstellungseinheit Unternehmen, da hier die internen Transaktionen zwischen den rechtlichen Einheiten komplexer Unternehmen herausgerechnet werden. Der Profilingeffekt spielt bei einer gesamtwirtschaftlichen Betrachtung der relevanten Merkmale keine Rolle (Ausnahme: Anzahl der Unternehmen), da er lediglich eine Verschiebung der Merkmalswerte zwischen einzelnen Wirtschaftsbereichen bewirkt. Tabelle 2 zeigt erwartungsgemäß, dass die Anzahl der Unternehmen sinkt, und zwar um 3,1 %. Der Umsatz ist aufgrund der Konsolidierung der Merkmale aller rechtlichen Einheiten in komplexen Unternehmen um 3,2 % niedriger. Die Reduzierung des Umsatzes teilt sich jedoch nicht gleichmäßig auf die einzelnen Umsatzarten auf: Der Umsatz aus industriellen Erzeugnissen reduziert sich nur um 1,1 %, während der Umsatz aus Handels- sowie aus Dienstleistungstätigkeiten um 5,6 beziehungsweise 4,3 % sinkt. Der Produktionswert reduziert sich aufgrund der Konsolidierung um 2,4 %. Bei den additiven Merkmalen gibt es keine Veränderung der Ergebnisse (der Konsolidierungseffekt ist 0).

Jedoch waren für die Realisierung des statistischen Unternehmens Methoden Anpassungen notwendig, die die Erstellung der Ergebnisse für statistische Unternehmen in der amtlichen Unternehmensstatistik in Deutschland überhaupt erst ermöglichen. Hintergründe zu diesen Methoden Anpassungen (Imputation und Anpassung der Hochrechnung) beschreiben Beck und andere (2020, hier: Abschnitt 3.1). Die Spalte „Imputations- und Hochrechnungseffekt“ der Tabelle 2 zeigt die prozentuale Steigerung der Merkmalswerte aufgrund dieser Methoden Anpassungen. Wird dieser Effekt mit einbezogen, erhöht sich die Anzahl der Unternehmen im Vergleich zur Anzahl der rechtlichen Einheiten insgesamt um 5,9 % (siehe Tabelle 2, Spalte „Gesamteffekt“ sowie

Tabelle 2

Gesamtwirtschaftliche Ergebnisse und Effekte nach Arbeitsschritten für ausgewählte Merkmale¹ 2017

	Ergebnis nach rechtlichen Einheiten	Ergebnis nach Imputation und Anpassung der Hochrechnung	Ergebnis nach Unternehmen (nach Konsolidierung)	Imputations- und Hochrechnungseffekt	Konsolidierungseffekt	Gesamteffekt
				%		
Anzahl der Unternehmen	2 504 371	2 736 066	2 650 987 ¹²	+ 9,3	– 3,1 ¹²	+ 5,9
Anzahl der Beschäftigten	29 768 330	30 762 492	30 762 492	+ 3,3	+ 0,0	+ 3,3
Mill. EUR						
Umsatz	6 573 160	6 784 590	6 564 206	+ 3,2	– 3,2	– 0,1
Waren- und Dienstleistungskäufe insgesamt	4 849 240	5 000 858	4 780 474	+ 3,1	– 4,4	– 1,4
Käufe von Waren und Dienstleistungen zum Wiederverkauf in unverarbeitetem Zustand	2 192 402	2 285 475	2 175 927	+ 4,2	– 4,8	– 0,8
Zahlungen für Leasingpersonal (Leiharbeiter)	35 219	35 540	35 004	+ 0,9	– 1,5	– 0,6
Zahlungen für langfristig gemietete und geleaste Produktionsanlagen	11 767	12 058	11 998	+ 2,5	– 0,5	+ 2,0
Produktionswert	4 456 450	4 576 806	4 465 970	+ 2,7	– 2,4	+ 0,2
Bruttogewinnspanne bei Handelsware	544 255	563 097	563 097	+ 3,5	+ 0,0	+ 3,5
Bruttowertschöpfung zu Faktorkosten	1 737 648	1 797 976	1 797 976	+ 3,5	+ 0,0	+ 3,5
Bruttobetriebsüberschuss	636 043	677 671	677 671	+ 6,5	+ 0,0	+ 6,5
Bruttoinvestitionen in Sachanlagen	265 344	270 964	270 964	+ 2,1	+ 0,0	+ 2,1
Bruttoinvestitionen in Grundstücke ¹³	14 900	16 214	16 214	+ 8,8	+ 0,0	+ 8,8

¹ Die Definition aller Merkmale der europäischen Strukturstatistik (Structural Business Statistics) können unter https://ec.europa.eu/eurostat/cache/metadata/en/sbs_esms.htm eingesehen werden (siehe 3.4 Statistical concepts and definitions).

² Hierbei handelt es sich nicht um den Konsolidierungseffekt, sondern um den Profilingeffekt, das Merkmal Anzahl der Unternehmen bildet hier eine Ausnahme.

³ Bruttoinvestitionen in Grundstücke werden im WZ-Abschnitt I (Gastgewerbe) nicht erhoben.

die Erläuterungen in Beck und andere, 2020, Abschnitt 3.1). Der Umsatz mindert sich nur noch um 0,1 %. Der negative Konsolidierungseffekt kehrt sich beim Produktionswert im Gesamteffekt sogar um und beträgt 0,2 %. Durch die Einbeziehung des Imputations- und Hochrechnungseffekts verändern sich auch die additiven Merkmale: Die Anzahl der Beschäftigten steigt um 3,3 % und die Investitionen in Sachanlagen erhöhen sich um 2,1 %. [↗ Tabelle 2](#)

Ergebnisse nach Wirtschaftsbereichen

Im Profiling werden rechtliche Einheiten zu einem komplexen Unternehmen zugeordnet, welches den Wirtschaftszweig der Haupteinheit erhält. Beispielsweise können rechtliche Einheiten und deren Merkmale, die im Handel erhoben wurden, dann einem Unternehmen außerhalb des Handels zugerechnet werden. Hierdurch entstehen Verschiebungen der Merkmale zwischen den Wirtschaftsbereichen im Vergleich zum Ergebnis nach rechtlichen Einheiten.

Die Grafiken 1 bis 4 zeigen die Verschiebungen zwischen den Wirtschaftsbereichen für vier ausgewählte Merkmale. Die Balken geben dabei jeweils die Merkmalswerte in Abhängigkeit von den einzelnen Arbeitsschritten

von der rechtlichen Einheit zum Unternehmen wieder. Ebenso ist die gesamte prozentuale Änderung (Gesamteffekt) der Werte blau dargestellt. In Klammern und grau ist die prozentuale Veränderung durch die Arbeitsschritte Profiling und Konsolidierung angegeben. Diese Änderungsrate drückt die eigentlich zu erwartende Änderung aufgrund der Einführung des neuen Unternehmensbegriffs aus.

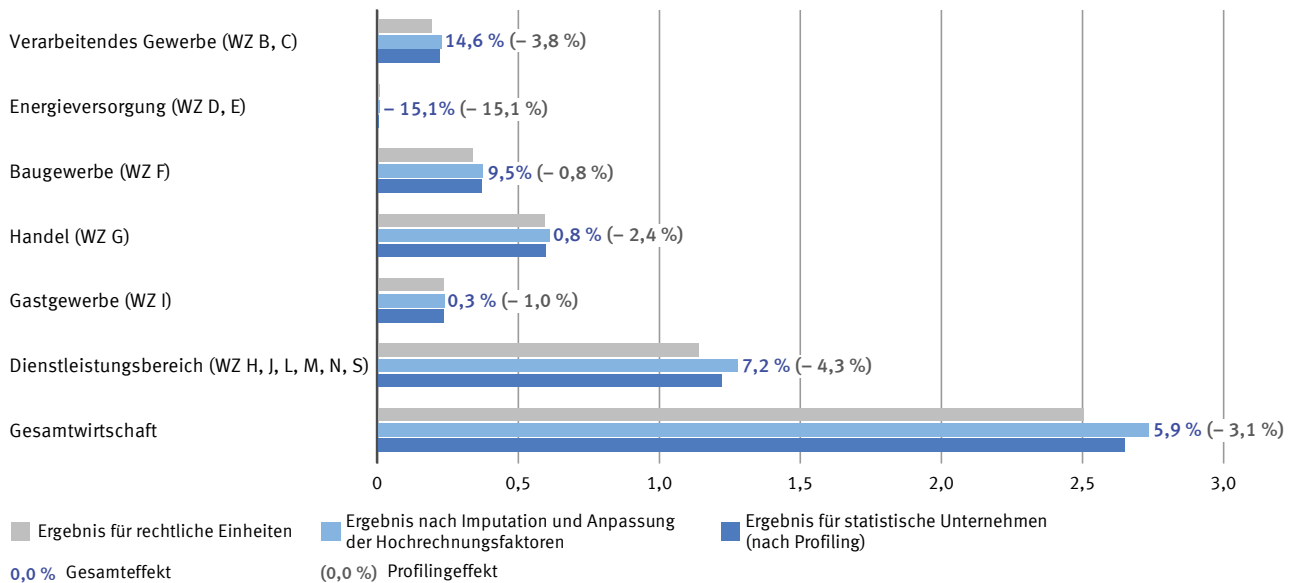
Bei der Anzahl der Unternehmen zeigt [↗ Grafik 1](#), dass mit dem Profiling grundsätzlich eine Reduzierung der Unternehmen in allen Wirtschaftsbereichen einhergeht. Durch den Imputations- und Hochrechnungseffekt wird dieser Effekt jedoch überkompensiert; die Anzahl der Unternehmen steigt. Ausschlaggebend hierfür ist die Harmonisierung der Hochrechnungsmethodik und die vollständigere Abdeckung der Grundgesamtheit. Dies zusammen führt zu einer besseren Darstellung des Gesamtbilds der deutschen Wirtschaft (Beck und andere, 2020, hier: Seite 40).

Eine Ausnahme bildet der Bereich Energieversorgung (WZ-Abschnitte D und E). In diesem Bereich gibt es eine Vollerhebung aller rechtlichen Einheiten. Es müssen weder rechtliche Einheiten vollständig imputiert noch hochgerechnet werden.

Auswirkungen der Einführung des EU-Unternehmensbegriffs

Grafik 1

Ergebnisse der Unternehmensstrukturstatistik nach Wirtschaftsbereichen für Unternehmen im Berichtsjahr 2017
Mill.

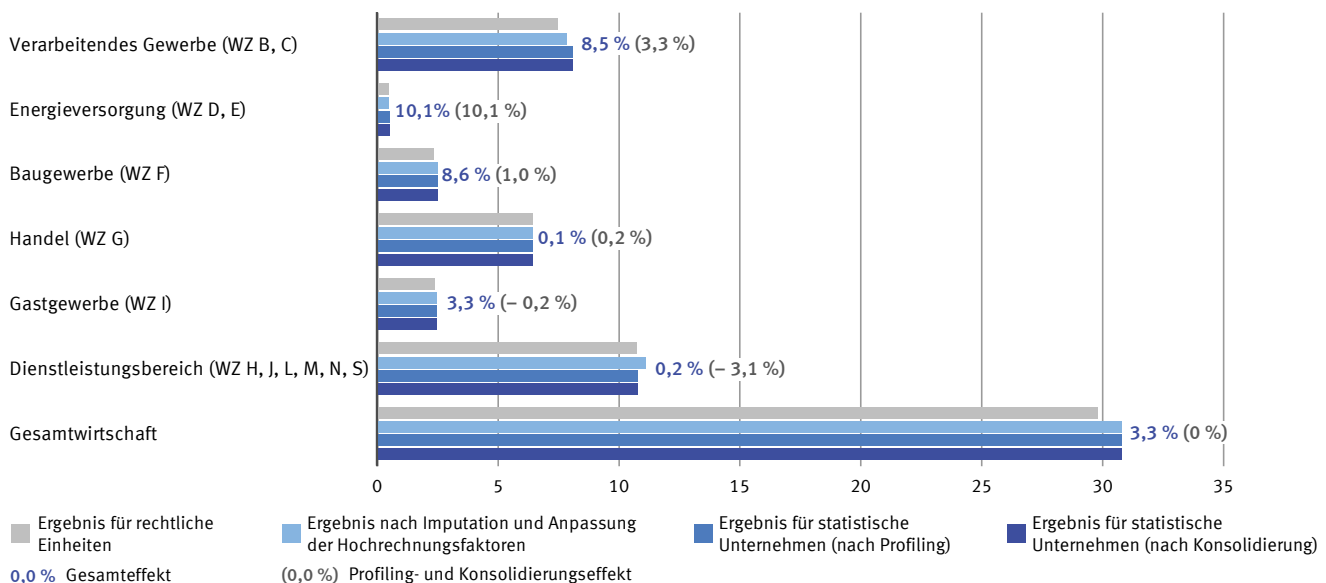


Wirtschaftsabschnitte der Klassifikation der Wirtschaftszweige, Ausgabe 2008 (WZ 2008).

2020 - 01 - 0227

Grafik 2

Ergebnisse der Unternehmensstrukturstatistik nach Wirtschaftsbereichen für Beschäftigte im Berichtsjahr 2017
Mill.

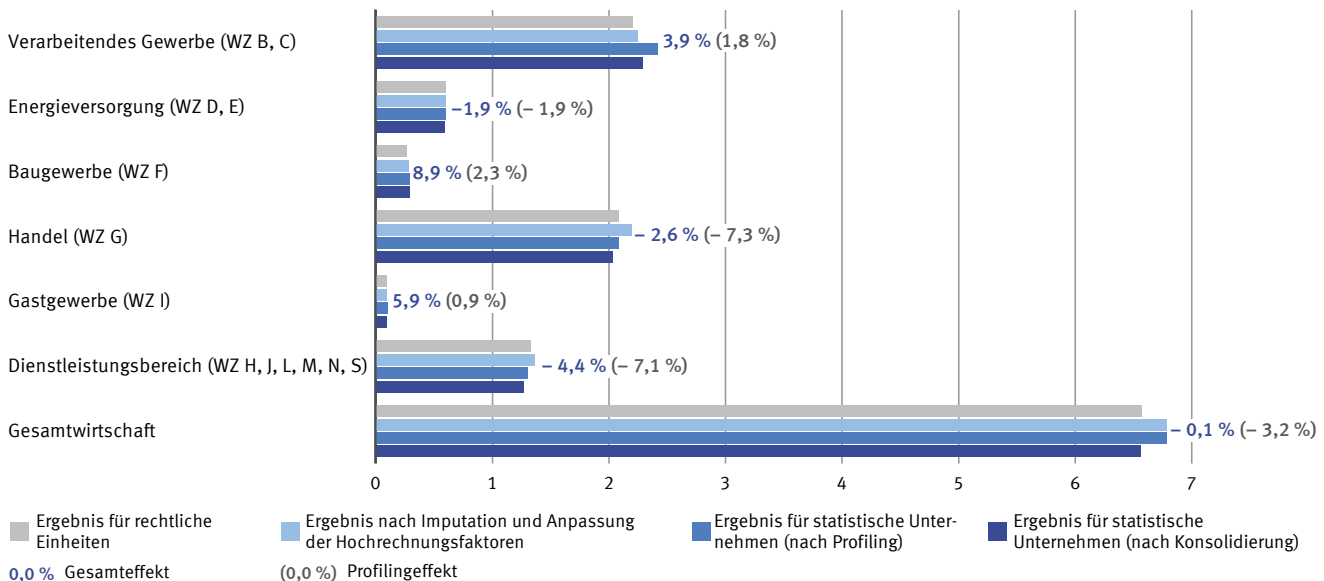


Wirtschaftsabschnitte der Klassifikation der Wirtschaftszweige, Ausgabe 2008 (WZ 2008).

2020 - 01 - 0228

Grafik 3

Ergebnisse der Unternehmensstrukturstatistik nach Wirtschaftsbereichen für den Umsatz im Berichtsjahr 2017
Mrd. EUR



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

2020 - 01 - 0229

Bei den Beschäftigten bewirkt der Profilingeffekt für sich betrachtet eine Reduzierung im Dienstleistungsbereich und im Gastgewerbe, während vor allem das Verarbeitende Gewerbe und der Bereich der Energieversorgung einen prozentualen Zuwachs an Beschäftigten erfahren.

➤ Grafik 2

Eine Erhöhung des Umsatzes ist vor allem im Verarbeitenden Gewerbe und Baugewerbe zu verzeichnen. Dagegen verzeichnen der Handel und der Dienstleistungsbereich größere Rückgänge. ➤ Grafik 3

Beim Merkmal Investitionen in Grundstücke ist der prozentuale Anstieg mit Ausnahme des Dienstleistungsbereichs in allen Bereichen außerordentlich hoch. Teilweise verdoppeln beziehungsweise verdreifachen sich die Merkmalswerte durch die Erstellung der Ergebnisse für statistische Unternehmen. Dies hängt damit zusammen, dass sowohl im automatischen als auch im manuellen Profiling rechtliche Einheiten des Wirtschaftszweigs „Grundstücks- und Wohnungswesen“ (WZ 68) sehr häufig als Hilfeinheiten in einem komplexen Unternehmen tätig sind, dessen Haupteinheit einem anderen Wirtschaftsbereich zugeordnet ist. Diese rechtlichen Einheiten tätigen hohe Investitionen in Grundstücke. Allein im Verarbeitenden Gewerbe verursacht die Zuordnung von

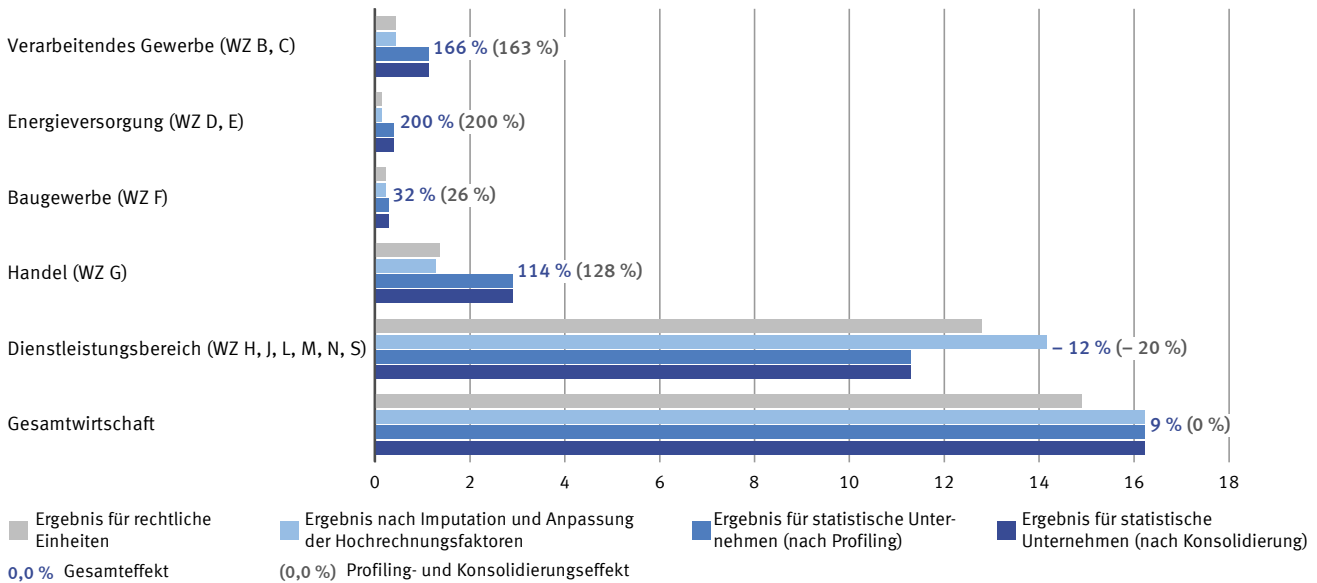
rechtlichen Einheiten aus diesem Wirtschaftszweig etwa 76 % der Erhöhung des Merkmalswerts. (Die übrigen 24 % der Erhöhung resultieren aus der Zuordnung rechtlicher Einheiten anderer Wirtschaftszweige.) ➤ Grafik 4

➤ Grafik 5 zeigt den Anteil der Investitionen in Grundstücke am Produktionswert im Verarbeitenden Gewerbe im Zeitverlauf seit dem Berichtsjahr 1995. Hier lässt sich erkennen, dass dieser Anteil mehr oder weniger kontinuierlich gesunken ist. Ersetzt man für das Berichtsjahr 2017 das Ergebnis durch den Wert für die neue Darstellungseinheit Unternehmen, erreicht der Anteil etwa wieder das Niveau des Berichtsjahres 1996. Die Entwicklung lässt vermuten, dass der Rückgang mit dem zunehmenden „unechten Outsourcing“ von rechtlichen Einheiten im Bereich des Grundstücks- und Wohnungswesens zusammenhängt und nicht beziehungsweise nicht ausschließlich auf strukturelle Änderungen zurückzuführen ist. Grafik 5 zeigt damit auch Auswirkungen, mit denen der Umstieg auf den EU-Unternehmensbegriff begründet wurde.

Auswirkungen der Einführung des EU-Unternehmensbegriffs

Grafik 4

Ergebnisse der Unternehmensstrukturstatistik nach Wirtschaftsbereichen für Investitionen in Grundstücke im Berichtsjahr 2017
Mrd. EUR



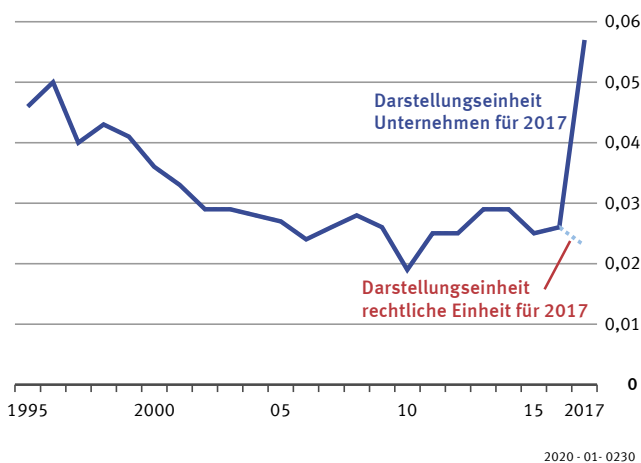
Wirtschaftsabschnitte der Klassifikation der Wirtschaftszweige, Ausgabe 2008 (WZ 2008).

1 Im WZ-Abschnitt I (Gastgewerbe) werden die Bruttoinvestitionen in Grundstücke nicht erhoben. Daher sind Aussagen zum Gesamteffekt nicht möglich.

2020 - 01 - 0229

Grafik 5

Anteil der Investitionen in Grundstücke am Produktionswert für das Verarbeitende Gewerbe nach Berichtsjahren in %



Ergebnisse nach Bundesländern

So wie es zu Verschiebungen zwischen den Wirtschaftsbereichen durch die Einführung der neuen Unternehmensdefinition kommt, ergeben sich auch regionale Verschiebungen der Unternehmen und der dazugehörigen Merkmale aufgrund des Profiling. Der Sitz der Haupteinheit gilt für das gesamte Unternehmen. Daher können rechtliche Einheiten und deren Merkmalswerte aufgrund des Wechsels der Darstellungseinheit einem anderen Bundesland zugeordnet werden.

Nicht für alle zentral durchgeführten Strukturstatistiken werden bislang Ergebnisse für rechtliche Einheiten nach Bundesländern veröffentlicht. [Tabelle 3](#) zeigt daher ausschließlich die prozentuale Änderung durch die beiden Arbeitsschritte Profiling und Konsolidierung für die Merkmale Anzahl der Unternehmen, Umsatz und Anzahl der Beschäftigten.

Die Anzahl der Unternehmen sinkt in allen Bundesländern aufgrund der Bildung von komplexen Unternehmen im Profiling. Eine Analyse der Effekte beim Umsatz und der Anzahl der Beschäftigten macht deutlich, dass aufgrund des Profiling und der Konsolidierung vor allem die neuen Bundesländer überdurchschnittliche negative

Tabelle 3

Effekte nach Bundesländern 2017

	Anzahl der Unternehmen	Umsatz	Anzahl der Beschäftigten
	Profiling- und Konsolidierungseffekt in %		
Schleswig-Holstein	- 2,9	- 5,2	- 2,9
Hamburg	- 5,6	+ 2,3	+ 8,8
Niedersachsen	- 3,4	- 3,3	+ 1,0
Bremen	- 6,2	- 10,0	- 0,6
Nordrhein-Westfalen	- 3,4	- 3,1	+ 0,3
Hessen	- 2,9	- 2,8	+ 3,0
Rheinland-Pfalz	- 2,6	- 3,9	- 2,9
Baden-Württemberg	- 3,0	+ 0,2	+ 2,6
Bayern	- 3,1	- 4,1	- 1,0
Saarland	- 2,7	- 6,2	- 4,4
Berlin	- 3,1	- 5,0	+ 1,4
Brandenburg	- 2,1	- 9,9	- 4,4
Mecklenburg-Vorpommern	- 2,2	- 9,8	- 5,6
Sachsen	- 2,4	- 13,5	- 7,6
Sachsen-Anhalt	- 2,5	- 10,5	- 5,5
Thüringen	- 2,4	- 13,6	- 9,1
Deutschland	- 3,1	- 3,2	0

Änderungsraten aufweisen. Dagegen ist der Rückgang bei der Anzahl der Unternehmen in den neuen Bundesländern unterdurchschnittlich. Hierbei ist jedoch zu beachten, dass diese Ergebnisse für Analysen der regionalen Wirtschaftsstruktur nur eingeschränkt geeignet sind, da Unternehmen – insbesondere nach neuer EU-Definition – häufig bundeslandübergreifend aktiv sind (siehe auch Kapitel 4).

3

Veröffentlichung der Daten zu statistischen Unternehmen

3.1 Datenlieferung an Eurostat

Die Datenlieferung an das Statistische Amt der Europäischen Union (Eurostat) richtet sich nach den europäischen Verordnungen der strukturellen Unternehmensstatistiken (siehe unter „Rechtsgrundlagen“). Die Verordnungen spezifizieren dabei outputorientiert die Wirtschaftszweige, für die Daten bereitgestellt werden müssen, den Merkmalskranz, sowie besondere Gliederungsebenen, beispielsweise Beschäftigtengrößen-

klassen. Diese sogenannten Reihen müssen für das Berichtsjahr 2018 bis zum 30. Juni 2020 erstmals für die Darstellungseinheit „Unternehmen“ im Sinne der EU-Definition zur Verfügung gestellt werden. Für die vorherigen Berichtsjahre bis einschließlich 2017 wurden von Deutschland Daten für die Darstellungseinheit „rechtliche Einheit“ geliefert. Diese Datenlieferung wird somit abgelöst und es kommt aus den in Kapitel 2 dargelegten Gründen für Deutschland zu einem Bruch in den an Eurostat übermittelten Daten.

3.2 Nationale Veröffentlichung

Ab dem Berichtsjahr 2018 steht auch in den nationalen Veröffentlichungen des Statistischen Bundesamtes die Darstellungseinheit „Unternehmen“ als adäquate Darstellungseinheit der strukturellen Unternehmensstatistiken im Fokus. Die gesamtwirtschaftliche Betrachtungsweise der neuen Darstellungseinheit wird bei der nationalen Veröffentlichung berücksichtigt. Das Statistische Bundesamt strebt für den Bund eine Veröffentlichungstabelle an, die alle von den Strukturstatistiken abgedeckten Wirtschaftsabschnitte und deren gemeinsame Merkmale umfasst. Dies erleichtert den Nutzerinnen und Nutzern einen Vergleich zwischen den einzelnen Wirtschaftsabschnitten, strukturelle Veränderungen der Wirtschaftsbereiche können im Zeitvergleich einfacher erkannt werden.

Die Veröffentlichung erfolgt zum einen über die Datenbank GENESIS-Online. Für die Wissenschaft ist zum anderen die Bereitstellung von Einzeldaten im Rahmen der AFiD-Panel¹ der Forschungsdatenzentren der Statistischen Ämter des Bundes und der Länder vorgesehen.

Ergebnisse für die Darstellungseinheit „rechtliche Einheit“ werden in einer Übergangsphase zumindest für die Berichtsjahre 2018 bis 2020 ergänzend zur Verfügung gestellt, sodass die Auswirkungen des Methodenwechsels nachvollzogen werden können. Danach muss evaluiert werden, inwieweit diese Veröffentlichungen weiter fortgeführt werden müssen. Im Gegensatz zur wirtschaftsbereichsübergreifenden Veröffentlichungs-

¹ AFiD ist die Abkürzung für „Amtliche Firmendaten in Deutschland“, einem aus mehreren Modulen bestehenden Angebot von Panel-daten der Forschungsdatenzentren der Statistischen Ämter des Bundes und der Länder. Weitere Informationen finden sich unter www.forschungsdatenzentrum.de/de/alle-daten.

tabelle beim „Unternehmen“ wird die übliche wirtschaftsbereichsspezifische Veröffentlichung bei der Darstellungseinheit „rechtliche Einheit“ beibehalten.

3.3 Geheimhaltung

Die Ergebnisse für Unternehmen setzen sich aus den Daten der komplexen und einfachen Unternehmen zusammen. Die Ergebnisse für die komplexen Unternehmen werden auf Grundlage des Profiling, durch die Imputation fehlender Erhebungsdaten und die anschließende Konsolidierung der rechtlichen Einheiten gewonnen. Sie unterscheiden sich somit von denen für rechtliche Einheiten. Bei den einfachen Unternehmen unterscheiden sich die Ergebnisse für Unternehmen im Vergleich zu rechtlichen Einheiten ebenfalls, da dort eine gesamtwirtschaftlich angepasste Methodik der Hochrechnung (Beck und andere, 2020, hier: Abschnitt 3.1) genutzt wird.

Aus diesem Grund wurde geprüft, ob für die Ergebnisse über Unternehmen eine Tabellengeheimhaltung notwendig ist oder ob die durchgeführten Arbeitsschritte die Ergebnisse so verändern, dass eine Deanonymisierung zugrundeliegender rechtlicher Einheiten ausgeschlossen ist. Theoretische Überlegungen und Auswertungen zeigen, dass die ergriffenen Maßnahmen nicht ausreichen, um eine Geheimhaltung sicherzustellen. Deshalb muss auch bei Tabellen mit Ergebnissen über Unternehmen eine Tabellengeheimhaltung durchgeführt werden.

Für die primäre statistische Tabellengeheimhaltung kommt ein Zellsperrverfahren nach der p-Prozent-Regel zum Einsatz. Die Geheimhaltung für Ergebnisse nach Unternehmen erfolgt dabei zentral und maschinell mithilfe der vom Niederländischen Statistikamt CBS entwickelten Software τ -Argus² (Hundepool und andere, 2011). Dies ermöglicht eine abgestimmte Geheimhaltung zwischen Bundes- und Landesergebnissen.

² Die Software kann über das niederländische Statistikamt bezogen werden.

3.4 Verwendung der Daten

Die Daten für Unternehmen im Sinne der EU-Unternehmensdefinition sind insbesondere für Analysen und Untersuchungen geeignet, die auf Marktakteure und diese kennzeichnende Indikatoren abstellen und von der rechtlichen Struktur der Marktakteure abstrahieren. Beispielhaft seien genannt:

- › Beobachten des Strukturwandels in der Wirtschaft (weil das „unechte“ Outsourcing eliminiert wird),
- › Analyse von kleinen und mittleren Unternehmen sowie evidenzbasierte Mittelstandspolitik (weil kleine und mittlere rechtliche Einheiten, die zu komplexen Unternehmen gehören, ausgeklammert werden),
- › Messen der Unternehmenskonzentration (weil auf Marktakteure abgestellt wird),
- › Messen der Arbeitsproduktivität (weil die Beschäftigten in unternehmensinternen, aber rechtlich ausgegliederten Leiharbeitsfirmen berücksichtigt werden).

Weniger geeignet sind die Daten hingegen für Regionalanalysen, da die betrachteten Wirtschaftsakteure (Unternehmen) bundesländerübergreifend aktiv sind.

4

Fazit und Ausblick

Nach rund fünf Jahren Vorbereitung steht die Einführung des EU-Unternehmensbegriffs in der deutschen Unternehmensstrukturstatistik unmittelbar bevor. Konzeptionell, methodisch und informationstechnisch sind alle Voraussetzungen geschaffen, um zum 30. Juni 2020 erstmals entsprechende Daten an Eurostat zu melden und somit den von der amtlichen Statistik verfolgten Aktionsplan erfolgreich umzusetzen. Über die Ergebnisse für das Berichtsjahr 2018 soll in einer der kommenden Ausgaben dieser Zeitschrift berichtet werden.

Gleichwohl sind in der näheren Zukunft noch eine Reihe von Folgearbeiten zu leisten. Die wichtigsten sind:

- › FRIBS

Mit der neuen Rahmenverordnung für die Unternehmensstatistiken auf europäischer Ebene (FRIBS) sind ab Berichtsjahr 2021 folgende Wirtschaftsbereiche zusätzlich in die Strukturstatistik einzubeziehen:

- › Erbringung von Finanz- und Versicherungsdienstleistungen (K 64.2/64.3/64.9, 66),
- › Erziehung und Unterricht (P),
- › Gesundheits- und Sozialwesen (Q),
- › Kunst, Unterhaltung und Erholung (R),
- › Erbringung von sonstigen, überwiegend persönlichen Dienstleistungen (S 96).

Dadurch werden sich derzeit noch nicht quantifizierbare Verschiebungen im Ergebnismachweis ergeben. Ab 2021 wird es komplexe Unternehmen geben, die ihren Schwerpunkt in den „neuen“ Wirtschaftsbereichen haben, aber auch rechtliche Einheiten aus den „alten“ Wirtschaftsbereichen umfassen, und umgekehrt. Damit werden auch regionale Verschiebungen einhergehen. Die Ergebnisse für die Berichtsjahre 2018 bis 2020 und ab Berichtsjahr 2021 werden sich alleine deshalb für die „alten“ Wirtschaftsbereiche unterscheiden.

Unter FRIBS wird außerdem gefordert, dass die statistischen Ergebnisse alle im Laufe des Berichtsjahrs aktiven Unternehmen berücksichtigen. Das ist in den Strukturstatistiken, die auf den Stichtag 31.12. abstellen, bisher nicht der Fall. Hierzu ist noch ein Konzept auszuarbeiten. Die Zahl der Unternehmen und auch die anderen Variablen werden sich dadurch im Vergleich zum Ist-Zustand erhöhen.

- › Berücksichtigung des Wirtschaftsabschnitts K Erbringung von Finanz- und Versicherungsdienstleistungen


Nach geltendem Recht müssen derzeit nicht für den gesamten Wirtschaftsabschnitt K, sondern lediglich für ausgewählte Bereiche (unter anderem für den Wirtschaftsbereich 64.1 Zentralbanken und Kreditinstitute) Daten an Eurostat geliefert werden, und zwar nach $t + 10$ Monaten (statt nach $t + 18$ Monaten). Dies ist eine Sonderregelung, die unter FRIBS entfallen wird. Erst danach kann die Datenlieferung zum Wirtschaftsabschnitt K auf die Darstellungseinheit statistisches Unternehmen umgestellt werden.

- › Nationale Merkmale

In den einzelnen Strukturstatistiken¹³ werden ergänzend zu den EU-Merkmalen unterschiedliche nationale Merkmale erhoben, um branchenspezifische

Informationsbedarfe abzudecken. Die amtliche Statistik hat sich bewusst dafür entschieden, zunächst Konzepte für die Imputation und Konsolidierung der EU-Merkmale zu entwickeln und die nationalen Merkmale zurückzustellen. Insofern wird sich der in die Veröffentlichungen zu statistischen Unternehmen einzubeziehende Merkmalskranz für die Berichtsjahre 2019 und später noch erweitern.

- › Länderscharfe Ergebnisse

Die Daten zu statistischen Unternehmen, die länderübergreifend wirtschaftlich aktiv sind, sind für Regionalanalysen weniger geeignet. Daher wird zurzeit ein Konzept zur Gewinnung länderscharfer Ergebnisse erarbeitet. 

³ Siehe Beck und andere, 2020, hier: Übersicht 1.

LITERATURVERZEICHNIS

Baumgärtner, Luisa/Gräb, Christopher/Leppert, Philipp/Söllner, René/Spies, Lydia/Veith, Stefan/Vorgrimler, Daniel. [*Imputation und Konsolidierung: Neue Aufgaben für die Unternehmensstatistik*](#). In: WISTA Wirtschaft und Statistik. Ausgabe 6/2018, Seite 33 ff.

Beck, Martin/Baumgärtner, Luisa/Bürk, Katja-Verena/Redecker, Matthias. [*Einführung des EU-Unternehmensbegriffs: Konzept und Umsetzung*](#). In: WISTA Wirtschaft und Statistik. Ausgabe 3/2020, Seite 35 ff.

Hundepool, Anco/van de Wetering, Aad/Ramaswamy, Ramya/de Wolf, Peter-Paul/Giessing, Sarah/Fischetti, Matteo/Salazar, Juan-José/Castro, Jordi/Lowthian, Philip. *τ-ARGUS Version 3.5 User's Manual*. Statistics Netherlands. Den Haag 2011.

Opfermann, Rainer/Beck, Martin. [*Einführung des EU-Unternehmensbegriffs*](#). In: WISTA Wirtschaft und Statistik. Ausgabe 1/2018, Seite 63 ff.

Redecker, Matthias/Sturm, Roland. [*Profiling von Unternehmen*](#). In: WISTA Wirtschaft und Statistik. Ausgabe 6/2017, Seite 9 ff.

Sturm, Roland/Redecker, Matthias. [*Das EU-Konzept des Unternehmens*](#). In: WISTA Wirtschaft und Statistik. Ausgabe 3/2016, Seite 57 ff.

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.

Verordnung (EU) 2015/2112 der Kommission vom 23. November 2015 zur Änderung von Anhang I der Verordnung (EG) Nr. 251/2009 zur Durchführung der Verordnung (EG) Nr. 295/2008 des Europäischen Parlaments und des Rates über die strukturellen Unternehmensstatistiken hinsichtlich der Anpassung der Datenreihen im Anschluss an die Überarbeitung der statistischen Güterklassifikation in Verbindung mit den Wirtschaftszweigen (CPA) (Amtsblatt der EU Nr. L 306, Seite 4).

Verordnung (EU) Nr. 446/2014 der Kommission vom 2. Mai 2014 zur Änderung der Verordnung (EG) Nr. 295/2008 des Europäischen Parlaments und des Rates über die strukturelle Unternehmensstatistik, der Verordnung (EG) Nr. 251/2009 der Kommission und der Verordnung (EU) Nr. 275/2010 der Kommission im Hinblick auf die zu erstellenden Datenreihen und die Kriterien für die Bewertung der Qualität der strukturellen Unternehmensstatistik (Amtsblatt der EU Nr. L 132, Seite 13).

Verordnung (EG) Nr. 251/2009 der Kommission vom 11. März 2009 zur Durchführung und Änderung der Verordnung (EG) Nr. 295/2008 des Europäischen Parlaments und des Rates im Hinblick auf die zu erstellenden Datenreihen für die strukturelle Unternehmensstatistik bzw. die nach der Überarbeitung der statistischen Güterklassifikation in Verbindung mit den Wirtschaftszweigen (CPA) erforderlichen Anpassungen (Amtsblatt der EU Nr. L 86, Seite 170).

Verordnung (EG) Nr. 295/2008 des Europäischen Parlaments und des Rates vom 11. März 2008 über die strukturelle Unternehmensstatistik (Amtsblatt der EU Nr. L 97, Seite 13).

IMMATERIELLES KAPITAL UND PRODUKTIVITÄT IM VERARBEITENDEN GEWERBE

Ergebnisse auf Basis verknüpfter Einzeldaten
(Micro Data Linking)

Dr. Markus Zimmermann

➤ **Schlüsselwörter:** Multifaktorproduktivität – Arbeitsproduktivität – Produktionsfunktionen – immaterielles Kapital – Micro Data Linking

ZUSAMMENFASSUNG

Der Aufsatz untersucht den Einfluss von immateriellem Kapital (Forschung und Entwicklung, Software, Patente) auf die Produktivität am Beispiel des Verarbeitenden Gewerbes in Deutschland. Hierzu werden Einzeldaten der Unternehmensstrukturstatistik verknüpft. Im ersten Schritt wird die Entwicklung der Investitionen in immaterielles Kapital deskriptiv analysiert. Im zweiten Schritt wird die Multifaktorproduktivität mittels Produktionsfunktionen geschätzt und die Ergebnisse verschiedener Modellspezifikationen verglichen. Die Analyse zeigt einerseits, dass immaterielles Kapital positive Effekte auf den Output hat, und andererseits, dass diese Effekte über die Unternehmen hinweg sehr ungleich verteilt sind.

➤ **Keywords:** multifactor productivity – labour productivity – production functions – intangible capital – microdata linking

ABSTRACT

This essay analyses the effects of intangible capital (research & development, software, patents) on productivity, using the example of the German manufacturing sector. For that purpose, microdata from structural business statistics are linked. In the first step, a descriptive analysis is made of the development of intangible investment. In the second step, multifactor productivity is estimated using production functions, and the results of different model specifications are compared. The analysis shows not only that intangible capital has a positive effect on output, but also that these effects are highly unequally distributed across businesses.



Dr. Markus Zimmermann

ist wissenschaftlicher Mitarbeiter im Referat „Analyse der Unternehmensstrukturen“ des Statistischen Bundesamtes und arbeitet zu den Themen Unternehmensstatistiken, Micro Data Linking und Produktivität. Er studierte Volkswirtschaftslehre an den Universitäten Freiburg, Basel und Göteborg. Anschließend promovierte er an der Humboldt-Universität zu Berlin im Bereich der Mikroökometrie.

1

Einleitung

Die Entwicklung der Produktivität ist langfristig die wichtigste Triebkraft des materiellen Wohlstands einer Volkswirtschaft. In den meisten entwickelten Volkswirtschaften hat das Produktivitätswachstum in den vergangenen Jahren abgenommen (Kuntze/Mai, 2020; Sachverständigenrat, 2019) und es gibt Hinweise auf eine zunehmende Ungleichheit zwischen hochproduktiven und weniger produktiven Unternehmen (Andrews und andere, 2019). Gleichzeitig sind jedoch viele methodische Fragen der Produktivitätsmessung sowie viele Determinanten der Produktivität bisher nur unzureichend untersucht (Byrne und andere, 2016; Syverson, 2017; Syverson, 2011).

Um diese Themen besser zu verstehen, finanziert die Europäische Kommission das Forschungsprojekt MICROPROD¹, an dem sich das Statistische Bundesamt gemeinsam mit anderen europäischen Partnern² beteiligt. Der Fokus des Projekts liegt auf der mikrodatenbasierten Messung der Produktivität. Hierzu sollen neue Datenquellen erschlossen werden, auch durch Verknüpfung bereits bestehender Einzeldaten (Micro Data Linking). Außerdem sollen mithilfe der verknüpften Daten innovative Produktivitätsmaße entwickelt werden.

Mit der Beteiligung am Projekt MICROPROD will das Statistische Bundesamt den Dialog zwischen amtlicher Statistik und Wissenschaft stärken. Ziel ist, die Potenziale der Mikrodatenverknüpfung bestmöglich auszuschöpfen und zu erreichen, dass die angewandte Methodik dem Stand der Forschung entspricht. Gleichzeitig ist die Verknüpfung von Einzeldaten in den Unternehmensstatistiken ein wichtiges Werkzeug, um die amtliche Statistik weiterzuentwickeln, ohne die Auskunftspflichten stärker zu belasten. Gesetzliche Grundlage hierfür ist § 13a Bundesstatistikgesetz.³ Das Statistische

Bundesamt greift hierbei auf Erfahrungen aus früheren Forschungsprojekten zu Micro Data Linking zurück (Leppert, 2019; Kaus/Leppert, 2017; Söllner/Jung, 2017; Jung/Käuser, 2016).

Dieser Aufsatz stellt erste Ergebnisse aus dem Projekt vor. Grundlage ist ein Diskussionspapier von Wolfhard Kaus, Viktor Slavtchev und Markus Zimmermann (Kaus und andere, 2020). Der Fokus der Analysen liegt auf dem Einfluss von immateriellem Kapital auf die Produktivität. In modernen Ökonomien sind nicht nur traditionelle Produktionsfaktoren wie Arbeitskräfte, Maschinen oder Bauten relevant, sondern zunehmend auch immaterielle Faktoren, wie Forschung und Entwicklung (FuE), Software oder Patente (Corrado und andere, 2005, 2009). Eine Analyse mithilfe von Einzeldaten soll zeigen, welche Unternehmen in immaterielle Produktionsfaktoren investieren und wie die Erträge dieser Investitionen verteilt sind. Hierzu werden Produktionsfunktionen geschätzt, um den Effekt verschiedener Inputfaktoren auf den Output zu ermitteln. Hieraus werden Schätzungen der Multifaktorproduktivität abgeleitet. Zudem soll untersucht werden, ob die Streuung der Produktivität zwischen Unternehmen durch die unterschiedlich starke Nutzung immaterieller Produktionsfaktoren erklärt werden kann.

Die Analyse beruht auf verknüpften Einzeldaten der Unternehmensstrukturstatistik im Verarbeitenden Gewerbe (Kostenstrukturerhebung und Investitionserhebung) sowie des statistischen Unternehmensregisters. Es sei darauf hingewiesen, dass alle Analysen auf Ebene der rechtlichen Einheiten durchgeführt werden, da Daten der Berichtsjahre bis einschließlich 2017 verwendet werden (siehe den Exkurs am Ende des Kapitels). Der Begriff „Unternehmen“ bezeichnet daher im Folgenden die rechtliche Einheit.

Die weitere Gliederung des Aufsatzes ist wie folgt: Kapitel 2 diskutiert die methodischen Grundlagen der Produktivitätsmessung. Kapitel 3 behandelt die verwendeten Daten und die Operationalisierung der Variablen. Die Darstellung der Ergebnisse folgt in Kapitel 4. Kapitel 5 bietet ein Fazit und einen Ausblick auf den weiteren Verlauf des Projekts.

1 Die Finanzierung des Projekts erfolgte durch das EU-Förderprogramm für Forschung und Innovation Horizon 2020, Grant Agreement No 822390.

2 Dies sind die Universität Aarhus, die Universität Bocconi, der Think Tank Bruegel, das Institut für Wirtschaftsforschung Halle, die Ungarische Akademie der Wissenschaften, die Vrije Universiteit Amsterdam, die Paris School of Economics und das University College London.

3 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.

Unternehmen und rechtliche Einheiten

Bis einschließlich des Berichtsjahrs 2017 entsprach das Unternehmen in der amtlichen Unternehmensstrukturstatistik der kleinsten rechtlichen Einheit, die aus handels- beziehungsweise steuerrechtlichen Gründen Bücher führt. Ab dem Berichtsjahr 2018 setzt die amtliche Statistik jedoch die EU-Einheitenverordnung um (siehe auch die Aufsätze von Beck und anderen, 2020a und 2020b, die ebenfalls in WISTA 3/2020 publiziert sind, sowie Opfermann/Beck, 2018). Die EU-Einheitenverordnung definiert das Unternehmen als „kleinste Kombination rechtlicher Einheiten, die eine organisatorische Einheit zur Erzeugung von Waren und Dienstleistungen bildet und [...] über eine gewisse Entscheidungsfreiheit verfügt“. Somit kann ein Unternehmen nach der neuen Definition auch aus mehreren rechtlichen Einheiten bestehen.

Im vorliegenden Aufsatz werden alle Analysen noch auf Ebene der rechtlichen Einheiten durchgeführt. Grund dafür ist, dass Daten der vergangenen Strukturerhebungen verwendet werden und es zum Beispiel für Panelanalysen wichtig ist, dieselbe Einheit konsistent über mehrere Jahre verfolgen zu können. Wenn im Folgenden bei der Diskussion der Ergebnisse daher von „Unternehmen“ die Rede ist, dann sind damit immer rechtliche Einheiten entsprechend der vor dem Berichtsjahr 2018 angewendeten Definition gemeint.

2

Methodische Grundlagen

2.1 Produktionsfunktionen und immaterielles Kapital

Ausgangspunkt der Analyse ist das Konzept einer Produktionsfunktion, die einen funktionalen Zusammenhang zwischen dem produzierten Output und den verwendeten Inputfaktoren (Produktionsfaktoren) beschreibt. Hierbei wird meist angenommen, dass ein Produktionsfaktor zumindest teilweise durch andere substituiert werden kann. Die traditionellen Produktionsfaktoren im Verarbeitenden Gewerbe sind Vorleistungen, Arbeit sowie physisches Kapital (Maschinen oder Bauten).

Es setzt sich jedoch in der Literatur zunehmend die Erkenntnis durch, dass diese Faktoren in modernen Ökonomien nicht ausreichen, um Produktionsprozesse zu beschreiben. Vielmehr spielen auch immaterielle Produktionsfaktoren eine entscheidende Rolle. Corrado und andere (2005) zählen hierzu die Unterkategorien „computerized information“ (Software und Datenbanken), „innovative property“ (Forschung und Entwicklung, Suchbohrungen, Patente und Lizenzen) sowie „economic competencies“ (Markenwerte, firmenfinanzierte Weiterbildung, Organisationskapital). Ausgaben für diese Zwecke (zum Beispiel FuE-Ausgaben) sind demnach als Investitionen zu zählen, die über die Zeit einen immateriellen Kapitalstock aufbauen, der wiederum in der Produktion eingesetzt wird.

Im Folgenden wird zur formalen Darstellung eine sogenannte Cobb-Douglas-Produktionsfunktion betrachtet. Der von Unternehmen i produzierte Output Y wird als multiplikative Funktion der Vorleistungen M , des Arbeitseinsatzes L , des physischen Kapitals PK , des immateriellen Kapitals IK sowie eines zusätzlichen Faktors A geschrieben:

$$(1) Y_i = M_i^{\beta_m} \cdot L_i^{\beta_l} \cdot PK_i^{\beta_{pk}} \cdot IK_i^{\beta_{ik}} \cdot A_i$$

Die Koeffizienten $\beta_m, \beta_l, \beta_{pk}, \beta_{ik}$ sind hier als Produktionselastizitäten interpretierbar. Beispielsweise bedeutet ein Koeffizient von $\beta_l = 0,3$, dass der Output um 0,3 % steigt, wenn der Arbeitseinsatz um 1 % steigt und alle anderen Produktionsfaktoren konstant gehalten werden.

Die Produktionsfunktion in Formel (1) verwendet den Bruttoproduktionswert als Maß für den Output und berücksichtigt sowohl die primären Inputs Arbeit und Kapital als auch die Vorleistungen („Gross Output“-Spezifikation). Im Folgenden wird diese Spezifikation genutzt, da es plausibel ist, dass im Produktionsprozess die Vorleistungen und die primären Inputs zumindest teilweise substituierbar sind. Manchmal wird aber in der Literatur auch die Bruttowertschöpfung als abhängige Variable verwendet und nur die primären Inputs Arbeit und Kapital berücksichtigt („Value Added“-Spezifikation). Für eine Diskussion dieser alternativen Spezifikation sei auf das bereits zitierte Arbeitspapier (Kaus und andere, 2020) verwiesen.

2.2 Produktivitätsmaße

Allgemein ist Produktivität definiert als Verhältnis des Outputs zu einem oder mehreren Inputs.¹⁴ Ausgehend von der Produktionsfunktion in Formel (1) ist die Arbeitsproduktivität (AP) von Unternehmen i definiert als Output geteilt durch den Arbeitseinsatz:

$$(2) AP_i = \frac{Y_i}{L_i}$$

Die Multifaktorproduktivität (MFP) von Unternehmen i ist definiert als Output geteilt durch das gewichtete Produkt aller Inputs:

$$(3) MFP_i = A_i = \frac{Y_i}{M_i^{\beta_m} \cdot L_i^{\beta_l} \cdot PK_i^{\beta_{pk}} \cdot IK_i^{\beta_{ik}}},$$

wobei jeder Inputfaktor mit der jeweiligen Produktionselastizität gewichtet wird. Während die Arbeitsproduktivität eine einfach interpretierbare Maßeinheit hat (zum Beispiel Umsatz in Euro je Beschäftigten), ist die absolute Größe der Multifaktorproduktivität nicht direkt interpretierbar. Vielmehr interessieren Unterschiede der Multifaktorproduktivität zwischen Unternehmen. Hat zum Beispiel Unternehmen A eine um 50% höhere Multifaktorproduktivität als Unternehmen B, so bedeutet das, dass Unternehmen A mit denselben Inputs einen um 50% höheren Output produzieren kann als Unternehmen B. Unterschiede in der Multifaktorproduktivität spiegeln somit die Unterschiede im Output wider, die nicht durch die beobachtbaren Inputs erklärt werden können. Außerdem lassen sich Veränderungsraten der Multifaktorproduktivität über die Zeit berechnen.

Ein Unternehmen kann im „Ranking“ durchaus unterschiedliche Positionen einnehmen – je nachdem, welches Produktivitätsmaß man betrachtet. So hat ein Unternehmen, das sehr kapitalintensiv produziert, typischerweise eine hohe Arbeitsproduktivität. Bei der Multifaktorproduktivität werden Unterschiede in der Kapitalausstattung jedoch herausgerechnet, sodass dieses Unternehmen im „Ranking“ nach der Multifaktorproduktivität nicht mehr so produktiv dastehen kann.

¹⁴ Davon zu unterscheiden ist das Konzept der Effizienz. Bei diesem wird die Output-Input-Kombination einer Produktionseinheit mit der einer „Best Practice“-Einheit verglichen und gefragt, ob Inputfaktoren optimal eingesetzt werden. Für eine weitere Diskussion siehe zum Beispiel Söllner (2017), der mittels einer Stochastischen Frontieranalyse die Effizienz im Verarbeitenden Gewerbe untersucht.

Zur Berechnung der Multifaktorproduktivität gemäß Formel (3) sind die Produktionselastizitäten zu bestimmen. Diese können zum Beispiel mittels einer Regression geschätzt werden. Hierzu wird Gleichung (1) zuerst logarithmiert:

$$(4) \ln Y_i = \beta_0 + \beta_m \cdot \ln M_i + \beta_l \cdot \ln L_i + \beta_{pk} \cdot \ln PK_i + \beta_{ik} \cdot \ln IK_i + u_i$$

Dabei bezeichnet β_0 die Regressionskonstante und u_i den Fehlerterm der Regression. Die geschätzte logarithmierte Multifaktorproduktivität ist dann das Residuum dieser Regression:

$$(5) \ln \widehat{MFP}_i = \ln Y_i - (\hat{\beta}_m \cdot \ln M_i + \hat{\beta}_l \cdot \ln L_i + \hat{\beta}_{pk} \cdot \ln PK_i + \hat{\beta}_{ik} \cdot \ln IK_i),$$

wobei $\hat{\beta}_m, \hat{\beta}_l, \hat{\beta}_{pk}, \hat{\beta}_{ik}$ die geschätzten Koeffizienten bezeichnen.¹⁵ Die Regression in Formel (4) kann sowohl mittels einer einfachen Kleinst-Quadrate-Schätzung als auch mittels anderer Methoden durchgeführt werden.¹⁶

Die gemessene Multifaktorproduktivität erlaubt von selbst noch keine Aussage über die genauen Ursachen der Produktivitätsunterschiede. Häufig wird sie als Maß für den technischen Fortschritt interpretiert. Allerdings können Unterschiede in der gemessenen Multifaktorproduktivität zwischen Unternehmen auch darauf zurückzuführen sein, dass nicht alle relevanten Inputfaktoren berücksichtigt wurden. Die Aussagekraft des Konzepts hängt daher maßgeblich davon ab, wie gut im jeweils vorliegenden Datensatz die Inputfaktoren gemessen sind. So könnte ein Teil der gemessenen MFP-Unterschiede darauf zurückzuführen sein, dass das immaterielle Kapital, vor allem auf Einzeldatenebene, bisher nicht ausreichend als Inputfaktor berücksichtigt wurde.

¹⁵ Die Konstante β_0 wird meist zum Residuum addiert. Dies hat jedoch auf die Streuung der Multifaktorproduktivität keinen Einfluss.

¹⁶ In der Literatur wird häufig auch die Methode von Wooldridge (2009) verwendet, die auf einer Generalized-Methods-of-Moments (GMM)-Schätzung basiert. Für eine detaillierte Erläuterung sei auf das Arbeitspapier (Kaus und andere, 2020) verwiesen. Allerdings ändert in der vorliegenden Anwendung die Wahl der Schätzmethode die Ergebnisse nicht grundlegend.

3

Daten und Operationalisierung

3.1 Daten

Die Analysen basieren auf Daten der Kostenstruktur-erhebung beziehungsweise der Investitionserhebung im Verarbeitenden Gewerbe.¹⁷ Beide Erhebungen umfassen rechtliche Einheiten mit 20 oder mehr tätigen Personen. Es besteht Auskunftspflicht. Die Investitionserhebung ist eine Vollerhebung mit Abschneidegrenze und umfasst rund 37 000 befragte Einheiten jährlich. Die Kostenstruktur-erhebung ist eine Rotationsstichprobe mit jährlich etwa 15 000 bis 17 000 Einheiten, wobei meist alle vier Jahre neue Stichproben gezogen werden¹⁸. Für deskriptive Analysen oder zur Konstruktion der Kapitalstöcke werden alle verfügbaren Jahre (ab 1995) verwendet. Die Regressionen werden aber für die Jahre 2009 bis 2015 durchgeführt, da in der Investitionserhebung Investitionen in erworbene Software und Patente/Lizenzen erst seit 2009 erhoben werden. Außerdem werden für diese Jahre Daten aus dem statistischen Unternehmensregister verknüpft, um Informationen zur Zugehörigkeit zu Unternehmensgruppen zu erhalten.¹⁹

3.2 Operationalisierung der Variablen

Die für die Schätzung der Produktionsfunktion in Formel (4) nötigen Variablen werden folgendermaßen operationalisiert:

Der **Output** Y wird als Bruttoproduktionswert (ohne Umsatzsteuer) gemessen. Der Bruttoproduktionswert ist die Summe aus Umsatz aus eigenen Erzeugnissen, Umsatz aus Handelsware, Provisionen aus Handelsver-

mittlung, Umsatz aus sonstigen Tätigkeiten, Bestandsveränderung aus eigener Produktion, sowie Wert selbst-erstellter Anlagen.

Die **Vorleistungen** M umfassen Ausgaben für Roh-, Hilfs- und Betriebsstoffe, Vorleistungen aus Handelsware sowie andere Vorleistungen (unter anderem für Leiharbeiterinnen und Leiharbeiter).

Der **Arbeitseinsatz** L ist gemessen als die Zahl der tätigen Inhaberinnen und Inhaber (einschließlich mithelfender Familienangehöriger) zuzüglich der Zahl der Arbeitnehmerinnen und Arbeitnehmer in Vollzeitäquivalenten (einschließlich zum Beispiel Auszubildender und Praktikantinnen/Praktikanten). Aufgrund der in den letzten Jahrzehnten stark steigenden Teilzeitbeschäftigung (Chalupa/Mai, 2018) ist es wichtig, Unterschiede in der geleisteten Arbeitszeit zu berücksichtigen.

Besonders schwierig zu messen sind die in den meisten Unternehmensdatensätzen nicht enthaltenen Kapitalstöcke. Sie sind auf Basis der Investitionen mithilfe der Perpetual-Inventory-Methode (PIM) zu ermitteln (siehe den Exkurs am Ende des Kapitels sowie Gühler/Schmalwasser, 2020). Die Grundidee dieser Methode ist, aus den vergangenen Stromgrößen (Investitionen) die aktuelle Bestandsgröße (Kapitalstock) zu schätzen.

Der **physische Kapitalstock** PK wird aus den vergangenen Investitionen in Maschinen gebildet.

Der **immaterielle Kapitalstock** IK wird gebildet aus den Ausgaben für betriebsinterne Forschung und Entwicklung aus der Kostenstruktur-erhebung sowie den Investitionen in erworbene Patente und erworbene Software aus der Investitionserhebung. Da die FuE-Ausgaben Personal- und Sachkosten ebenso wie Investitionen umfassen, müssen zur Vermeidung von Doppelzählungen die anderen Inputs korrigiert werden. Beispielsweise wird die Zahl der FuE-Beschäftigten von der Gesamtzahl der Beschäftigten abgezogen und die so erhaltene Zahl der „Nicht-FuE-Beschäftigten“ als Maß für den Arbeitseinsatz L verwendet.

Die Struktur-erhebungen im Verarbeitenden Gewerbe enthalten einige Variablen nicht, die in der Literatur ebenfalls zu immateriellem Kapital gezählt werden. Beispielsweise werden weder Informationen zum Wert der selbstgeschaffenen Patente oder Software noch die

7 Unternehmen im Bergbau und der Gewinnung von Steinen und Erden werden nicht einbezogen, da die Produktionsprozesse mit denen im Verarbeitenden Gewerbe schwer vergleichbar sind.

8 Die Verknüpfung von Kostenstruktur-erhebung und Investitionserhebung ist auch als AfID-Panel Industrieunternehmen für Wissenschaftlerinnen und Wissenschaftler in den Forschungsdatenzentren der Statistischen Ämter des Bundes und der Länder nutzbar.

9 Einerseits wird diese Information als Kontrollvariable in den Schätzungen verwendet, andererseits werden als Robustheitsprüfung separate Schätzungen für Unternehmen mit/ohne Gruppenzugehörigkeit durchgeführt.

zugekaufte Forschung und Entwicklung erhoben.¹⁰ Gleiches gilt für Markenwerte, Organisationskapital oder den Wert firmenfinanzierter Weiterbildungsmaßnahmen.

Zusätzliche Maße für immaterielles Kapital ließen sich aus der Erhebung zur Nutzung von Informations- und Kommunikationstechnologien in Unternehmen (IKT-Erhebung) ableiten. Allerdings unterscheiden sich die Berichtskreise und Erhebungsarten der IKT-Erhebung und der Strukturhebungen stark. Daher ist die Schnittmenge der Einheiten, für die aus beiden Datenquellen Angaben vorliegen, zu gering, um für das Verarbeitende Gewerbe aussagekräftige Analysen zu generieren. Weitere denkbare Datenquellen für Deutschland wären außerdem die FuE-Erhebung des Stifterverbands für die deutsche Wissenschaft e.V. oder das Mannheimer Innovationspanel des Leibniz-Zentrums für Europäische Wirtschaftsforschung (ZEW)¹¹. Auf die Einzeldaten dieser Erhebungen hat das Statistische Bundesamt jedoch keinen Zugriff.

Zur Deflationierung der Inputs und Outputs dienen Preisindizes auf WZ-2-Steller-Ebene¹². Sie werden den Veröffentlichungen der Volkswirtschaftlichen Gesamtrechnungen (Fachserie 18, Reihe 1.4) entnommen, und zwar separat für Produktionswerte, Wertschöpfung sowie Investitionsgüter für Ausrüstungen beziehungsweise geistiges Eigentum. Dies ist in der empirischen Literatur üblich, da die eigentlich nötigen unternehmensspezifischen Preisindizes in den meisten Datensätzen nicht zur Verfügung stehen.¹³

➤ Perpetual-Inventory-Methode (PIM)

Die hier verwendete Variante der Perpetual-Inventory-Methode nutzt die folgende Beziehung:

$$(6) K_t = (1 - \delta) \cdot K_{t-1} + I_t,$$

10 In den Strukturhebungen im Dienstleistungsbereich werden seit dem Berichtsjahr 2012 Investitionen in selbstgeschaffene immaterielle Anlagen erfasst. Allerdings werden dort die FuE-Ausgaben nicht erhoben, deshalb ist der Dienstleistungsbereich nicht einbezogen.

11 Das Mannheimer Innovationspanel ist der deutsche Beitrag zum Community Innovation Survey (CIS), der Innovationserhebung der Europäischen Union.

12 Wirtschaftsabteilungen der Klassifikation der Wirtschaftszweige, Ausgabe 2008 (WZ 2008).

13 Für das Verarbeitende Gewerbe wäre es möglich, aus der Produktionserhebung Daten zu Produktionsmengen und Verkaufswerten für einzelne Güter zu nutzen. Hieraus lässt sich jedoch nicht ohne Weiteres ein unternehmensspezifischer Preisindex konstruieren, da nicht klar ist, in welchem Maß Unterschiede in den Verkaufswerten auch Qualitätsunterschiede widerspiegeln.

wobei K_t den Kapitalstock der Periode t , K_{t-1} den Kapitalstock der Vorperiode $t-1$, δ die Abschreibungsrate und I_t die Investitionen bezeichnen.

Bei der Umsetzung der Perpetual-Inventory-Methode ist erstens für die Kapitalgüter eine geeignete Abschreibungsrate zu bestimmen. Beispielsweise ist es plausibel, für immaterielles Kapital eine höhere Abschreibungsrate anzusetzen als für Maschinen, da zum Beispiel Software sehr schnell veraltet. Zweitens ist meist ein „Startwert“ für den Kapitalstock zu bestimmen, da die verfügbaren Beobachtungszeiträume in den meisten Paneldaten nicht ausreichend lang sind, um aus den vergangenen Stromgrößen vollständig die Bestandsgröße zu bilden. Der Einfluss des Startwerts, der stets mit gewissen Messfehlern behaftet ist, sinkt aber mit der Länge der verfügbaren Zeitreihen beziehungsweise der Abschreibungsrate. Der Anhang des diesem Beitrag zugrunde liegenden Papiers diskutiert diese Fragen genauer (Kaus und andere, 2020).

4

Ergebnisse

4.1 Deskriptive Analyse

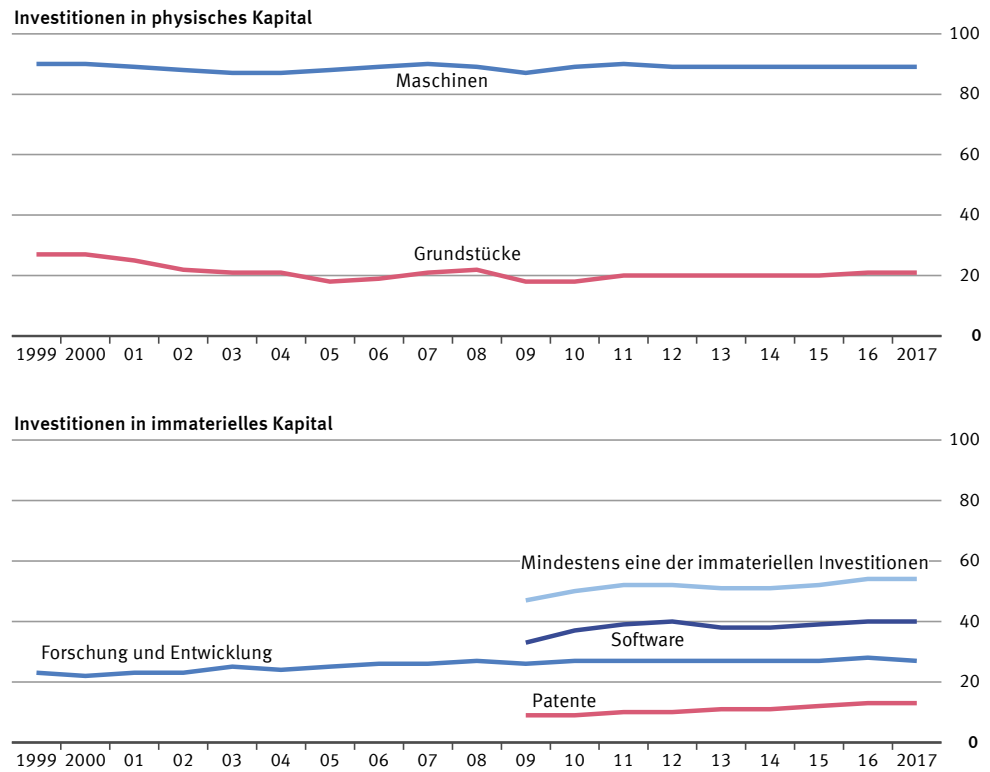
➤ Grafik 1 stellt für jedes Berichtsjahr im Zeitraum von 1999 bis 2017 den Anteil der Unternehmen dar, die angeben, in diesem Jahr bestimmte Investitionen getätigt zu haben. Während im Jahr 2017 die überwiegende Mehrheit (89 %) Investitionen in Maschinen tätigte, investierten nur 54 % in immaterielles Kapital. Investitionen in Software waren dabei mit 40 % am weitesten verbreitet, während Investitionen in Patente (13 %) sowie Forschung und Entwicklung (27 %) seltener vorkamen. Im Zeitablauf haben Investitionen in immaterielles Kapital zugenommen. So stieg der Anteil der in Forschung und Entwicklung aktiven Unternehmen von knapp 23 % im Jahr 1999 auf gut 27 % im Jahr 2017.

➤ Tabelle 1 zeigt die Perzentile der verschiedenen Investitionen. Betrachtet werden die kumulierten Investitionen eines Unternehmens über den Zeitraum 2012 bis 2015.¹⁴ Diese mehrjährige Betrachtung trägt der Tat-

14 Diese Jahre wurden gewählt, da sie den jüngsten vollständigen Rotationszyklus der Kostenstrukturhebung bilden.

Grafik 1

Anteile der Unternehmen mit Investitionen
in %



Gewichtete Ergebnisse der Kostenstrukturerhebung/Investitionserhebung im Verarbeitenden Gewerbe. Rechtliche Einheiten mit 20 und mehr tätigen Personen.

2020 - 01 - 0209

sache Rechnung, dass Investitionen oft nicht in jedem Jahr getätigt werden, sondern sich in einzelnen Jahren häufen. Der Anteil der investierenden Unternehmen in

der vierjährigen Betrachtung (Tabelle 1) ist daher höher als in der einjährigen Betrachtung (Grafik 1). Weiterhin zeigt sich, dass die Verteilung aller Investitionen stark

Tabelle 1

Perzentile der kumulativen Bruttoinvestitionen im Zeitraum 2012 bis 2015

	Anteil der Unternehmen mit >0	Perzentil				
		P10	P25	P50	P75	P90
	%	1 000 EUR				
Investitionen in physisches Kapital	96,7	66	226	775	2 649	8 780
Maschinen	96,7	62	205	669	2 228	7 053
Grundstücke	36,2	0	0	0	99	1 158
Investitionen in immaterielles Kapital	69,8	0	0	28	414	3 202
Forschung und Entwicklung	33,1	0	0	0	207	2 668
Software	61,5	0	0	7	62	265
Patente	25,2	0	0	0	0	52

Gewichtete Ergebnisse der Kostenstrukturerhebung/Investitionserhebung im Verarbeitenden Gewerbe. Rechtliche Einheiten mit 20 und mehr tätigen Personen. Verwendet wird ein balanciertes Panel von N=13 585 Unternehmen, die 2012 bis 2015 durchgängig in den Erhebungen vertreten waren.

rechtsschief ist, das heißt die meisten Unternehmen investieren nichts oder wenig und wenige Unternehmen investieren sehr viel. Diese Konzentration ist bei den Investitionen in immaterielles Kapital noch deutlich stärker als bei den Investitionen in physisches Kapital. So beträgt der Median der immateriellen Investitionen lediglich 28 000 Euro, während das 90. Perzentil bei 3,2 Millionen Euro liegt.

➤ **Tabelle 2** enthält die Perzentile der immateriellen Investitionen nach Wirtschaftsabteilungen (WZ-2-Stellen). Die Wirtschaftsabteilungen mit den höchsten Werten sind C26 Herstellung von Datenverarbeitungsgeräten (Median: 1,2 Millionen Euro), C21 Herstellung von phar-

mazeutischen Erzeugnissen (Median: 875 000 Euro) sowie C20 Herstellung von chemischen Erzeugnissen (Median: 680 000 Euro). Die Tabelle zeigt jedoch auch, dass es innerhalb der Wirtschaftsabteilungen große Unterschiede zwischen Unternehmen gibt.

➤ **Grafik 2** soll weiter verdeutlichen, welche Unternehmen besonders intensiv immaterielles Kapital einsetzen. Hierzu wird als Maß die Intensität von immateriellem Kapital (immaterieller Kapitalstock je Vollzeitäquivalent) betrachtet.¹⁵ Der Datensatz wird in sechs Grup-

15 Der immaterielle Kapitalstock wird, wie in Abschnitt 3.2 dargestellt, gemäß der Perpetual-Inventory-Methode (PIM) berechnet.

Tabelle 2

Perzentile der kumulativen immateriellen Bruttoinvestitionen nach Wirtschaftsabteilungen¹ des Verarbeitenden Gewerbes² im Zeitraum 2012 bis 2015

	Unternehmen	Anteil der Unternehmen mit > 0	Perzentil				
			P10	P25	P50	P75	P90
	Anzahl	%	1 000 EUR				
10 Nahrungs- und Futtermittel	1 708	43,8	0	0	0	18	268
11 Getränke	224	71,5	0	0	34	152	740
13 Textilien	290	71,9	0	0	49	506	1 950
14 Bekleidung	155	72,7	0	0	30	360	1 082
16 Holz-, Flecht-, Korb- und Korkwaren (ohne Möbel)	310	51,0	0	0	0	49	247
17 Papier, Pappe und Waren daraus	382	76,7	0	1	47	319	1 485
18 Druckerzeugnisse; Vervielfältigung von Medien	268	68,8	0	0	19	81	296
20 Chemische Erzeugnisse	755	88,9	0	42	680	3 421	12 619
21 Pharmazeutische Erzeugnisse	167	95,1	7	81	875	4 976	29 367
22 Gummi- und Kunststoffwaren	742	73,7	0	0	37	297	1 676
23 Glas und Glaswaren, Keramik, Steine und Erden	605	71,1	0	0	24	301	1 587
24 Metallherzeugung und -bearbeitung	577	79,2	0	3	71	552	3 020
25 Metallherzeugnisse	1 731	65,2	0	0	14	112	724
26 Datenverarbeitungsgeräte, elektronische und optische Erzeugnisse	620	91,2	1	64	1 198	4 580	20 493
27 Elektrische Ausrüstungen	914	81,2	0	8	241	2 786	10 288
28 Maschinenbau	1 916	82,9	0	11	157	1 544	8 119
29 Kraftwagen und Kraftwagenteile	496	77,5	0	2	108	1 542	12 172
30 Sonstiger Fahrzeugbau	192	84,7	0	19	233	2 941	12 495
31 Möbel	311	64,9	0	0	16	154	1 534
32 Sonstige Waren	515	62,9	0	0	9	381	3 049
33 Reparatur und Installation von Maschinen und Ausrüstungen	579	56,2	0	0	2	30	421

Gewichtete Ergebnisse der Kostenstrukturhebung/Investitionserhebung im Verarbeitenden Gewerbe. Rechtliche Einheiten mit 20 und mehr tätigen Personen. Verwendet wird ein balanciertes Panel von N = 13 585 Unternehmen, die 2012 bis 2015 durchgängig in den Erhebungen vertreten waren.

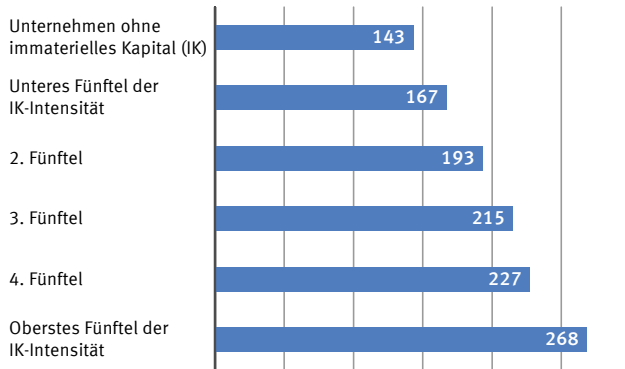
1 2-Steller der Klassifikation der Wirtschaftszweige, Ausgabe 2008 (WZ 2008).

2 Die Wirtschaftsabteilungen C12 (Tabakverarbeitung), C15 (Herstellung von Leder, Lederwaren und Schuhen) und C19 (Kokerei und Mineralölverarbeitung) werden wegen geringer Fallzahlen nicht ausgewiesen.

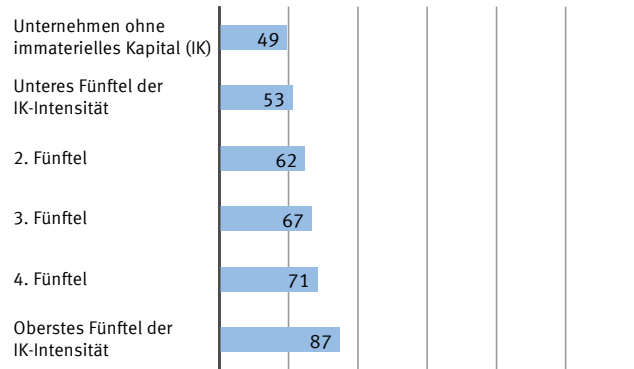
Grafik 2

Eigenschaften der Unternehmen nach der Intensität von immateriellem Kapital (immaterieller Kapitalstock je Vollzeitäquivalent)

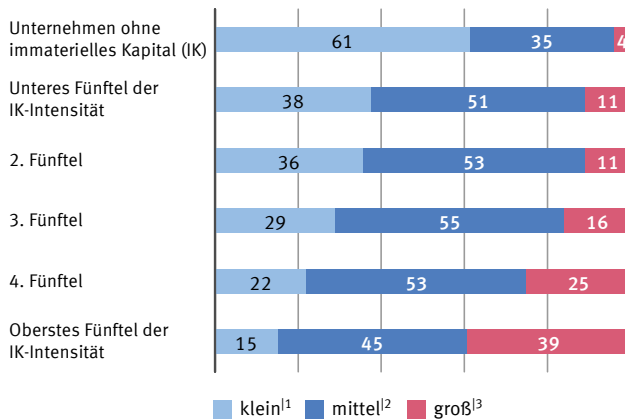
Durchschnittlicher Bruttoproduktionswert je Vollzeitäquivalent in 1 000 EUR



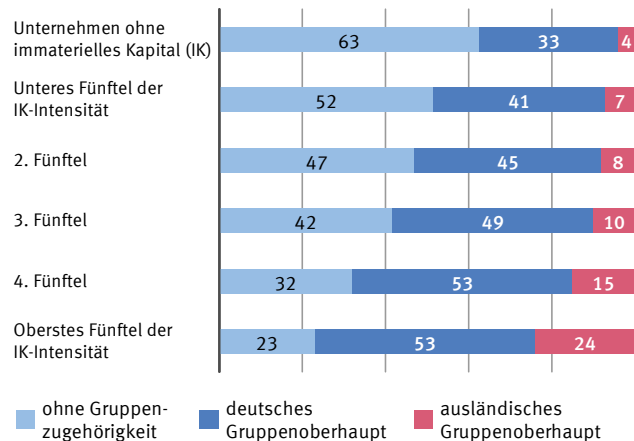
Durchschnittliche Bruttowertschöpfung je Vollzeitäquivalent in 1 000 EUR



Größenklassen der Unternehmen, in %



Zugehörigkeit zu Unternehmensgruppen, in %



Gewichtete Ergebnisse der Kostenstrukturerhebung/Investitionserhebung im Verarbeitenden Gewerbe im Zeitraum 2009 bis 2015.

1 Bis 49 tätige Personen und Jahresumsatz bis 10 Mill. EUR.

2 Bis 249 tätige Personen und Jahresumsatz bis 50 Mill. EUR.

3 250 und mehr tätige Personen oder Jahresumsatz über 50 Mill. EUR.

2020 - 01 - 0210

pen unterteilt. Die erste Gruppe bilden die Beobachtungen, die keinen immateriellen Kapitalstock haben. Die Gruppe der Beobachtungen mit positivem Kapitalstock wird dann gemäß ihrer Intensität in Fünftel (Quintile) unterteilt.

Die Arbeitsproduktivität (gemessen entweder als Bruttoproduktionswert oder als Bruttowertschöpfung je Voll-

zeitäquivalent) steigt mit zunehmender Intensität von immateriellem Kapital an. Das oberste Fünftel der Intensität hebt sich in seiner Arbeitsproduktivität jedoch noch einmal deutlich von den anderen Gruppen ab. Unternehmen mit hoher Intensität von immateriellem Kapital sind häufiger große Unternehmen sowie häufiger Teil einer Unternehmensgruppe.

4.2 Schätzung der Produktionsfunktionen

↗ **Tabelle 3** zeigt die geschätzten Produktionsfunktionen. Hierbei wird ein gemeinsames Modell für alle Unternehmen im Verarbeitenden Gewerbe mittels einer einfachen Kleinste-Quadrate-Regression geschätzt. Spalte 2 weist eine Produktionselastizität für immaterielles Kapital von 0,013 nach, das heißt eine Erhöhung des immateriellen Kapitals um 1 % steigert den Output um 0,013 %, wenn alle anderen Inputfaktoren konstant gehalten werden. In den Spalten 3 bis 6 werden die einzelnen Komponenten des immateriellen Kapitals in die Regression eingefügt. Aufgrund der positiven Korrelation der Variablen untereinander ist der Effekt der Teilkomponenten größer, wenn sie jeweils einzeln in die Regression eingefügt werden (Spalten 3 bis 5) als wenn alle gemeinsam eingefügt werden (Spalte 6). Insgesamt ist der Effekt von Forschung und Entwicklung mit 0,017 höher als der von Software (0,008) und Patenten (0,001). Bezüglich der übrigen Input-Variablen ergeben sich Produktionselastizitäten von etwa 0,72 für Vorleistungen, 0,24 bis 0,25 für Arbeit und 0,033 bis 0,035 für physisches Kapital. Die geschätzten Skalenerträge (das heißt die Summe aller Koeffizienten) sind in allen Regressionen nahe bei eins, was im Einklang mit den Annahmen der ökonomischen Theorie steht.

Im zugrunde liegenden Diskussionspapier (Kaus und andere, 2020) werden neben der hier dargestellten Kleinste-Quadrate-Schätzung auch andere Methoden untersucht, insbesondere eine Fixed-Effects-Schätzung sowie eine GMM-Schätzung nach Wooldridge (2009). Außerdem diskutiert es Ergebnisse der „Value Added“-Spezifikation (mit der Bruttowertschöpfung als abhängiger Variablen, und ohne Kontrolle für die Vorleistungen). In allen Modellspezifikationen bestätigt sich ein positiver Effekt des immateriellen Kapitals, vor allem von Forschung und Entwicklung, allerdings variiert die Größenordnung der Effekte.

↗ **Tabelle 4** zeigt die Schätzungen separat für Beobachtungen mit unterschiedlicher Intensität von immateriellem Kapital (gemessen als Kapitalstock je Vollzeit-äquivalent). Die Effekte des immateriellen Kapitals sind sehr klein für die Unternehmen, die dieses Kapital weniger intensiv nutzen (das unterste, zweite und dritte Fünftel der Intensität). Hingegen besteht ein stark positiver Effekt von 0,073 für das oberste Fünftel, also für die Unternehmen, die immaterielles Kapital besonders intensiv nutzen. Dieses oberste Fünftel sticht auch deswegen hervor, da die Effekte des immateriellen Kapitals größer sind als die des physischen Kapitals, während es bei den anderen Gruppen umgekehrt ist.

Tabelle 3

Produktionsfunktionen (Kleinste-Quadrate-Schätzung) für alle Unternehmen im Verarbeitenden Gewerbe

	Regression 1	Regression 2	Regression 3	Regression 4	Regression 5	Regression 6
Vorleistungen	0,724*** (0,002)	0,718*** (0,002)	0,720*** (0,002)	0,722*** (0,002)	0,723*** (0,002)	0,719*** (0,002)
Arbeit	0,250*** (0,003)	0,239*** (0,003)	0,239*** (0,003)	0,243*** (0,003)	0,248*** (0,003)	0,235*** (0,003)
Physisches Kapital	0,035*** (0,001)	0,033*** (0,001)	0,035*** (0,001)	0,034*** (0,001)	0,035*** (0,001)	0,034*** (0,001)
Immaterielles Kapital		0,013*** (0,001)				
Forschung und Entwicklung			0,018*** (0,001)			0,017*** (0,001)
Software				0,011*** (0,001)		0,008*** (0,001)
Patente					0,006*** (0,001)	0,001 (0,001)
Skalenerträge	1,009	1,004	1,012	1,010	1,012	1,013
R ²	0,985	0,986	0,986	0,986	0,986	0,986
N*T	95 638	95 638	95 638	95 638	95 638	95 638
N	22 439	22 439	22 439	22 439	22 439	22 439

Anmerkungen: Jede Spalte bezeichnet eine separate Regression. Standardfehler in Klammern, geclustert auf Unternehmensebene. * p < 0.1, ** p < 0.05, *** p < 0.01. Weitere Kontrollvariablen: Dummies für WZ 2-Steller, Jahresdummies, ein Dummy für Neue Länder, und Dummies für Zugehörigkeit zu Unternehmensgruppen. Gewichtete Regressionen.

Quelle: Kostenstrukturerhebung/Investitionserhebung im Verarbeitenden Gewerbe 2009 bis 2015. Rechtliche Einheiten mit 20 und mehr tätigen Personen.

Tabelle 4

Produktionsfunktionen (Kleinste-Quadrate-Schätzung), separat nach Quintil des immateriellen Kapitalstocks je Vollzeitäquivalent

	Unternehmen ohne immateriellen Kapitalstock	Fünftel (Quintile) der Intensität des immateriellen Kapitalstocks				
		Unteres Fünftel	2. Fünftel	3. Fünftel	4. Fünftel	Oberes Fünftel
Vorleistungen	0,702*** (0,004)	0,709*** (0,005)	0,713*** (0,006)	0,728*** (0,005)	0,732*** (0,005)	0,715*** (0,006)
Arbeit	0,275*** (0,007)	0,244*** (0,007)	0,229*** (0,009)	0,228*** (0,008)	0,203*** (0,008)	0,197*** (0,007)
Physisches Kapital	0,028*** (0,002)	0,042*** (0,003)	0,051*** (0,003)	0,039*** (0,003)	0,037*** (0,003)	0,015*** (0,003)
Immaterielles Kapital		0,004 (0,003)	0,005 (0,005)	0,004 (0,005)	0,026*** (0,005)	0,073*** (0,005)
Skalenerträge	1,005	0,999	0,998	0,999	0,998	1
R ²	0,972	0,981	0,983	0,984	0,987	0,987
N*T	20 793	12 864	12 999	14 386	16 235	18 361
N	5 871	4 380	5 098	5 073	4 770	4 355

Anmerkungen: Jede Spalte bezeichnet eine separate Regression. Standardfehler in Klammern, geclustert auf Unternehmensebene. * p < 0,1, ** p < 0,05, *** p < 0,01. Weitere Kontrollvariablen: Dummies für WZ 2-Steller, Jahresdummies, ein Dummy für Neue Länder, und Dummies für Zugehörigkeit zu Unternehmensgruppen. Gewichtete Regressionen.

Quelle: Kostenstrukturerhebung/Investitionserhebung im Verarbeitenden Gewerbe 2009 bis 2015. Rechtliche Einheiten mit 20 und mehr tätigen Personen.

Zusätzliche Robustheitsprüfungen (Kaus und andere, 2020) konnten zeigen, dass diese Ungleichheit der Erträge auch innerhalb verschiedener Untergruppen besteht. Beispielsweise gilt dies für verschiedene Zeitperioden (2009 bis 2011 und 2012 bis 2015), für verschiedene Größenklassen sowie für Unternehmen mit und ohne Gruppenzugehörigkeit.

Diese Ergebnisse verdeutlichen, dass die Unternehmen nicht mit einer homogenen Produktionsfunktion operieren und dass diese Heterogenitäten in der Modellspezifikation berücksichtigt werden müssen. Die Ergebnisse lassen sich so interpretieren, dass sich der Einsatz von immateriellem Kapital erst dann lohnt, wenn bereits ein gewisses Maß davon akkumuliert wurde. Die genauen Ursachen für dieses Muster lassen sich mit den vorliegenden Daten aber nicht beantworten. Eine mögliche Erklärung ist, dass Investitionen in immaterielles Kapital mit großen Fixkosten sowie großer Unsicherheit verbunden sind und nur Unternehmen mit entsprechender Erfahrung diese Kosten tragen können. Eine offene Frage ist auch, ob und wie Unternehmen, die bei der Nutzung immateriellen Kapitals zurückliegen, diesen Rückstand aufholen können.

4.3 Streuung der Multifaktorproduktivität vor und nach Berücksichtigung von immateriellem Kapital

Abschließend wird untersucht, in welchem Ausmaß sich die Streuung der Produktivität innerhalb von Wirtschaftsabteilungen durch die unterschiedliche Nutzung immateriellen Kapitals erklären lässt. Hierzu erfolgt – separat auf Ebene der WZ-2-Steller – eine Schätzung der Produktionsfunktionen ohne und mit Einbeziehung von immateriellem Kapital als Inputfaktor. Dann wird für jede der beiden Spezifikationen die geschätzte Multifaktorproduktivität je Unternehmen berechnet und anschließend die Streuung der geschätzten Multifaktorproduktivität über alle Unternehmen in der Stichprobe bestimmt. Als Streuungsmaß wird die prozentuale Differenz zwischen dem 90. und dem 10. Perzentil der Verteilung verwendet („90-10-Differenz“). Gegenüber anderen Verteilungsmaßen wie der Varianz hat dieses Maß den Vorteil, dass es weniger durch Ausreißer beeinflusst wird.

➤ Tabelle 5 auf Seite 72 weist für das Verarbeitende Gewerbe insgesamt eine 90-10-Differenz von 44,5 % nach (ohne Berücksichtigung von immateriellem Kapital). Das bedeutet, dass ein Unternehmen am 90. Perzentil mit denselben Inputs 44,5 % mehr Output produzieren kann als ein Unternehmen am 10. Perzentil. Starke Produktivitätsunterschiede sind ebenfalls innerhalb der Wirtschaftsabteilungen zu verzeichnen (in der

Tabelle 5

Bedeutung des immateriellen Kapitals für die Streuung der Multifaktorproduktivität

	90-10-Differenz		Veränderung
	ohne immaterielles Kapital	mit immateriellem Kapital	%
C Verarbeitendes Gewerbe	44,5	43,2	- 2,9
10 Nahrungs- und Futtermittel	41,9	41,8	- 0,2
11 Getränke	58,3	57,4	- 1,5
13 Textilien	43,4	43,3	- 0,2
14 Bekleidung	47,6	47,1	- 1,1
16 Holz-, Flecht-, Korb- und Korkwaren (ohne Möbel)	31,5	31,2	- 1,0
17 Papier, Pappe und Waren daraus	31,5	31,1	- 1,3
18 Druckerzeugnisse; Vervielfältigung von Medien	52,2	51,3	- 1,7
20 Chemische Erzeugnisse	45,3	43,9	- 3,1
21 Pharmazeutische Erzeugnisse	56,6	54,7	- 3,4
22 Gummi- und Kunststoffwaren	36,0	35,3	- 1,9
23 Glas und Glaswaren, Keramik, Steine und Erden	39,9	39,2	- 1,8
24 Metallherzeugung und -bearbeitung	32,6	31,9	- 2,1
25 Metallherzeugnisse	45,0	44,6	- 0,9
26 Datenverarbeitungsgeräte, elektronische und optische Erzeugnisse	68,4	60,9	- 11,0
27 Elektrische Ausrüstungen	47,5	45,8	- 3,6
28 Maschinenbau	42,4	41,4	- 2,4
29 Kraftwagen und Kraftwagenteile	43,6	42,2	- 3,2
30 Sonstiger Fahrzeugbau	54,6	51,6	- 5,5
31 Möbel	33,3	33,3	0,0
32 Sonstige Waren	61,2	58,4	- 4,6
33 Reparatur und Installation von Maschinen und Ausrüstungen	50,3	50,1	- 0,4

Dargestellt ist die prozentuale Differenz zwischen dem 90. Perzentil und dem 10. Perzentil der geschätzten Multifaktorproduktivität (90-10-Differenz). Wirtschaftsabschnitt und Wirtschaftsabteilungen der Klassifikation der Wirtschaftszweige, Ausgabe 2008 (WZ 2008). Schätzungen basieren auf den Produktionsfaktoren, separat für WZ-2-Steller. Ergebnisse der Kostenstrukturerhebung/Investitionserhebung im Verarbeitenden Gewerbe 2009 bis 2015. Rechtliche Einheiten mit 20 und mehr tätigen Personen.


Größenordnung von etwa 30 bis 60 %). Die Streuung der Multifaktorproduktivität wird kleiner, wenn immaterielles Kapital als Produktionsfaktor berücksichtigt wird. Die Reduktion der Streuung beträgt 2,9 % im Durchschnitt des Verarbeitenden Gewerbes und ist erwartungsgemäß in den Wirtschaftsabteilungen größer, in denen immaterielles Kapital intensiver eingesetzt wird (zum Beispiel in der Wirtschaftsabteilung C26 Herstellung von Datenverarbeitungsgeräten, wo die MFP-Streuung um 11 % sinkt). Insgesamt bleibt jedoch auch nach Berücksichtigung von immateriellem Kapital eine erhebliche Streuung bestehen. Die Produktivitätsunterschiede zwischen Unternehmen lassen sich also nur zu einem kleinen Teil durch diesen Faktor erklären.

5

Fazit und Ausblick

Dieser Aufsatz untersuchte mithilfe verknüpfter Einzeldaten des Verarbeitenden Gewerbes die Bedeutung immateriellen Kapitals für die Produktivität. Investitionen in immaterielles Kapital haben im Zeitablauf zugenommen, sind jedoch immer noch stark auf wenige Unternehmen konzentriert. Mit der Schätzung von Produktionsfunktionen wurde gezeigt, dass vor allem Forschung und Entwicklung – und zu einem geringeren Maß Software und Patente – einen positiven Effekt auf den Output haben. Diese Effekte sind jedoch ebenfalls sehr heterogen und vor allem auf die Unternehmen konzentriert, die sehr intensiv immaterielles Kapital einsetzen. Diese Ergebnisse deuten darauf hin, dass die Adaption neuer Technologien sowie die Nutzung ihrer Potenziale erst mit gewisser Zeitverzögerung möglich sind.

Außerdem wurde deutlich, dass auch nach Berücksichtigung des immateriellen Kapitals als Inputfaktor eine große unerklärte Streuung zwischen sehr produktiven und weniger produktiven Unternehmen verbleibt. Dies kann darauf hindeuten, dass weitere unbeobachtete Faktoren – wie die Fähigkeiten des Managements oder Marktmacht – die Produktivitätsunterschiede erklären.

Insgesamt zeigt sich, dass die Verwendung verknüpfter Einzeldaten ein großes zusätzliches Analysepotenzial birgt. Zudem ist deutlich geworden, dass methodische Fragen, beispielsweise unterschiedliche Modellspezifikationen, durchaus einen Einfluss auf die Ergebnisse haben können. Diese Fragen sollen daher im weiteren Verlauf des Projekts vertieft werden. Außerdem sollen weitere Verknüpfungen von Einzeldaten vorgenommen werden, insbesondere mit Daten der Außenhandelsstatistik. Hierdurch soll ein noch umfassenderes Bild der Determinanten von Produktivität gewonnen werden. 

LITERATURVERZEICHNIS

- Andrews, Dan/Criscuolo, Chiara/Gal, Peter. *The Best versus the Rest: Divergence across Firms During the Global Productivity Slowdown*. CEP Discussion Paper No. 1645. London 2019. [Zugriff am 15. April 2020]. Verfügbar unter: cep.lse.ac.uk
- Beck, Martin/Baumgärtner, Luisa/Bürk, Katja-Verena/Redecker, Matthias. [Einführung des EU-Unternehmensbegriffs: Konzept und Umsetzung](#). In: WISTA Wirtschaft und Statistik. Ausgabe 3/2020, Seite 35 ff. (2020a)
- Beck, Martin/Baumgärtner, Luisa/Bürk, Katja-Verena/Redecker, Matthias. [Auswirkungen der Einführung des EU-Unternehmensbegriffs](#). In: WISTA Wirtschaft und Statistik. Ausgabe 3/2020, Seite 49 ff. (2020b)
- Byrne, David M./Fernald, John G./Reinsdorf, Marshall B. *Does the United States Have a Productivity Slowdown or a Measurement Problem?* In: Brookings Papers on Economic Activity. Ausgabe 1/2016, Seite 109 ff. [Zugriff am 15. April 2020]. Verfügbar unter: brookings.edu
- Chalupa, Johannes/Mai, Christoph-Martin. [Entwicklungen am Arbeitsmarkt in Österreich und Deutschland – zwischen Jobwunder und Produktivitätsparadoxon](#). In: WISTA Wirtschaft und Statistik. Ausgabe 6/2018, Seite 48 ff.
- Corrado, Carol/Hulten, Charles/Sichel, Daniel. *Measuring Capital and Technology: An Expanded Framework*. In: Corrado, Carol/Haltiwanger, John/Sichel, Daniel (Herausgeber). *Measuring Capital in the New Economy*. Chicago 2005, Seite 11 ff. Verfügbar unter: www.nber.org/chapters/c0202.pdf
- Corrado, Carol/Hulten, Charles/Sichel, Daniel. *Intangible Capital and U.S. Economic Growth*. In: *Review of Income and Wealth*. Jahrgang 55. Ausgabe 3/2009, Seite 661 ff. Verfügbar unter: onlinelibrary.wiley.com
- Gühler, Nadine/Schmalwasser, Oda. [Anlagevermögen, Abschreibungen und Abgänge in den Volkswirtschaftlichen Gesamtrechnungen](#). In: WISTA Wirtschaft und Statistik. Ausgabe 3/2020, Seite 76 ff.
- Jung, Sandra/Käuser, Stefanie. [Herausforderungen und Potenziale der Einzeldatenverknüpfung in der Unternehmensstatistik](#). In: WISTA Wirtschaft und Statistik. Ausgabe 2/2016, Seite 95 ff.
- Kaus, Wolfhard/Slavtchev, Viktor/Zimmermann, Markus. *Intangible Capital and Productivity. Firm-level Evidence from German Manufacturing*. IWH Discussion Paper No. 1/2020. [Zugriff am 2. Juni 2020]. Verfügbar unter: www.iwh-halle.de
- Kaus, Wolfhard/Leppert, Philipp. [Außenhandelsaktive Unternehmen in Deutschland: neue Perspektiven durch Micro Data Linking](#). In: WISTA Wirtschaft und Statistik. Ausgabe 3/2017, Seite 22 ff.
- Kuntze, Peter/Mai, Christoph-Martin. [Arbeitsproduktivität – nachlassende Dynamik in Deutschland und Europa](#). In: WISTA Wirtschaft und Statistik. Ausgabe 2/2020, Seite 11 ff.

Leppert, Philipp. [*The Micro Data Linking-Panel: A Combined Firm Dataset*](#). In: Journal of Economics and Statistics (Jahrbücher für Nationalökonomie und Statistik). Jahrgang 240. Ausgabe 4/2019, Seite 525 ff. [Zugriff am 16. April 2020]. Verfügbar unter: degruyter.com

Opfermann, Rainer/Beck, Martin. [*Einführung des EU-Unternehmensbegriffs*](#). In: WISTA Wirtschaft und Statistik. Ausgabe 1/2018, Seite 63 ff.

Sachverständigenrat zur Begutachtung der gesamtwirtschaftlichen Entwicklung. [*Jahresgutachten 2019/20: Den Strukturwandel meistern*](#). [Zugriff am 16. April 2020]. Verfügbar unter: www.sachverstaendigenrat-wirtschaft.de/

Söllner, René/Jung, Sandra. *The impact of foreign trade and ownership on enterprise exits: new insights through micro data linking*. In: AStA Wirtschafts- und Sozialstatistisches Archiv. Ausgabe 1/2017, Seite 8 ff. [Zugriff am 2. Juni 2020]. Verfügbar unter: www.springerprofessional.de

Söllner, René. [*Heterogenität und strukturelle Dynamik im Verarbeitenden Gewerbe*](#). In: WISTA Wirtschaft und Statistik. Ausgabe 4/2017, Seite 72.

Syverson, Chad. [*Challenges to Mismeasurement Explanations for the US Productivity Slowdown*](#). In: Journal of Economic Perspectives. Jahrgang 31. Ausgabe 2/2017, Seite 165 ff. [Zugriff am 16. April 2020]. Verfügbar unter: www.aeaweb.org

Syverson, Chad. [*What Determines Productivity?*](#) In: Journal of Economic Literature. Jahrgang 49. Ausgabe 2/2011, Seite 326 ff. [Zugriff am 16. April 2020]. Verfügbar unter: www.aeaweb.org

Wooldridge, Jeffrey M. [*On estimating firm-level production functions using proxy variables to control for unobservables*](#). In: Economic Letters. Jahrgang 104. Ausgabe 3/2009, Seite 112 ff. [Zugriff am 16. April 2020]. Verfügbar unter: sciencedirect.com

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.



Nadine Gühler

ist Diplom-Kauffrau und arbeitet seit 2009 im Statistischen Bundesamt. Sie ist Referentin im Referat „Vermögensrechnung“ und beschäftigt sich schwerpunktmäßig mit der Weiterentwicklung methodischer Lösungskonzepte und Rechenabläufe auf dem Gebiet der volkswirtschaftlichen Abschreibungen und des Anlagevermögens.



Dr. Oda Schmalwasser

ist Diplomwirtschaftlerin und nach einigen Jahren an einem Wirtschaftsforschungsinstitut seit 1992 im Statistischen Bundesamt tätig. Sie hat in verschiedenen Bereichen der Volkswirtschaftlichen Gesamtrechnungen gearbeitet und leitet das Referat „Vermögensrechnung“. Ihre Schwerpunkte liegen in der Weiterentwicklung und praktischen Umsetzung der Vermögensrechnung auf nationaler und internationaler Ebene.

ANLAGEVERMÖGEN, ABSCHREIBUNGEN UND ABGÄNGE IN DEN VOLKSWIRTSCHAFTLICHEN GESAMTRECHNUNGEN

Nadine Gühler, Dr. Oda Schmalwasser

➤ **Schlüsselwörter:** Volkswirtschaftliche Gesamtrechnungen – Vermögensrechnung – nichtfinanzielle Vermögensgüter – Anlagevermögen – Abschreibungen

ZUSAMMENFASSUNG

Die Vermögensrechnung ist ein Teilsystem der Volkswirtschaftlichen Gesamtrechnungen, insbesondere das Anlagevermögen und die Abschreibungen sind von großer Bedeutung. Der Beitrag erläutert die zugrunde liegende Klassifikation der nichtfinanziellen Vermögensgüter, die auch die Abgrenzung der in das Anlagevermögen einzubeziehenden Anlegegüter vorgibt. Das Anlagevermögen kann nach dem Brutto- und nach dem Nettokonzept dargestellt werden, deren Berechnungswege werden ausführlich beschrieben. Die Bedeutung der dabei ermittelten Abschreibungen und Abgänge für die Diskussion über notwendige Ersatzinvestitionen in die öffentliche Infrastruktur wird erörtert.

➤ **Keywords:** *national accounts – national wealth accounts – non-financial assets – stock of fixed assets – consumption of fixed capital*

ABSTRACT

National wealth accounts are a national accounts subsystem. Especially the stock of fixed assets and consumption of fixed capital are of major importance. The article explains the underlying classification of non-financial assets which includes the boundary of fixed assets. The stock of fixed assets can be shown according to the net and gross concepts; the article describes in detail the relevant calculation methods. It also addresses the current discussion about necessary replacement investments in public infrastructure and the importance of consumption of fixed capital and retirements in this context.

1

Einleitung

Die Volkswirtschaftlichen Gesamtrechnungen (VGR) geben ein quantitatives Gesamtbild über das wirtschaftliche Geschehen in einer Volkswirtschaft. Neben der Inlandsproduktberechnung als Kernbereich ist die Vermögensrechnung ein weiterer Bestandteil dieses umfassenden Gesamtrechnungssystems. Der Beitrag dokumentiert den aktuellen Stand der Berechnung von Anlagevermögen und Abschreibungen im Rahmen der Vermögensrechnung. Darüber hinaus stellt er weitere Größen der Anlagevermögensrechnung, wie die Abgänge, vor und erläutert ihre wirtschaftspolitische Aussagefähigkeit.

Die Bedeutung der Vermögensrechnung – insbesondere der Abschreibungen – für die Berechnung des Bruttoinlandsprodukts wird in Kapitel 2 herausgearbeitet. Das dritte Kapitel geht auf die Klassifikation der nichtfinanziellen Vermögensgüter ein. Diese gibt unter anderem vor, welche Anlagegüter aktuell in die Berechnungen einbezogen werden müssen. Kapitel 4 erklärt den Übergang vom Jahresanfangs- zum Jahresendbestand einschließlich der Ermittlung von Umbewertungsgewinnen/-verlusten für das Brutto- und das Nettoanlagevermögen und zeigt die zahlenmäßige Entwicklung seit 1991. Schließlich greift Kapitel 5 die jüngere Diskussion über notwendige (Ersatz-)Investitionen und die Rolle der zuvor erläuterten Brutto- und Nettokonzepte in diesem Zusammenhang auf.

2

Vermögensrechnung als Teilsystem der Volkswirtschaftlichen Gesamtrechnungen

Im Mittelpunkt der Volkswirtschaftlichen Gesamtrechnungen steht die Inlandsproduktberechnung mit der Ermittlung des Bruttoinlandsprodukts als zentrale Größe für laufende Konjunkturanalysen und die Wirtschaftspolitik. Das Bruttoinlandsprodukt kann über die Entstehungs-, die Verwendungs- und die Verteilungsseite berechnet werden. Bei der Entstehungsrechnung,

die die Produktion abbildet, ergibt die Bruttowertschöpfung aller Wirtschaftsbereiche – nach Hinzufügen beziehungsweise Absetzen der Gütersteuern/-subventionen – das Bruttoinlandsprodukt. Verwendungsseitig wird das Bruttoinlandsprodukt als Summe aus privaten und staatlichen Konsumausgaben, Investitionen und dem Außenbeitrag (Exporte abzüglich Importe) ermittelt. Die dritte Möglichkeit der Berechnung – die Verteilungsrechnung – erfasst die entstandenen Einkommen. Sie kann in Deutschland nur zur Plausibilisierung des Bruttoinlandsprodukts herangezogen werden, da die Basisdaten über die Unternehmens- und Vermögenseinkommen lückenhaft sind. Eine komplette Berechnung über die verschiedenen Einkommensarten ist somit nicht möglich.

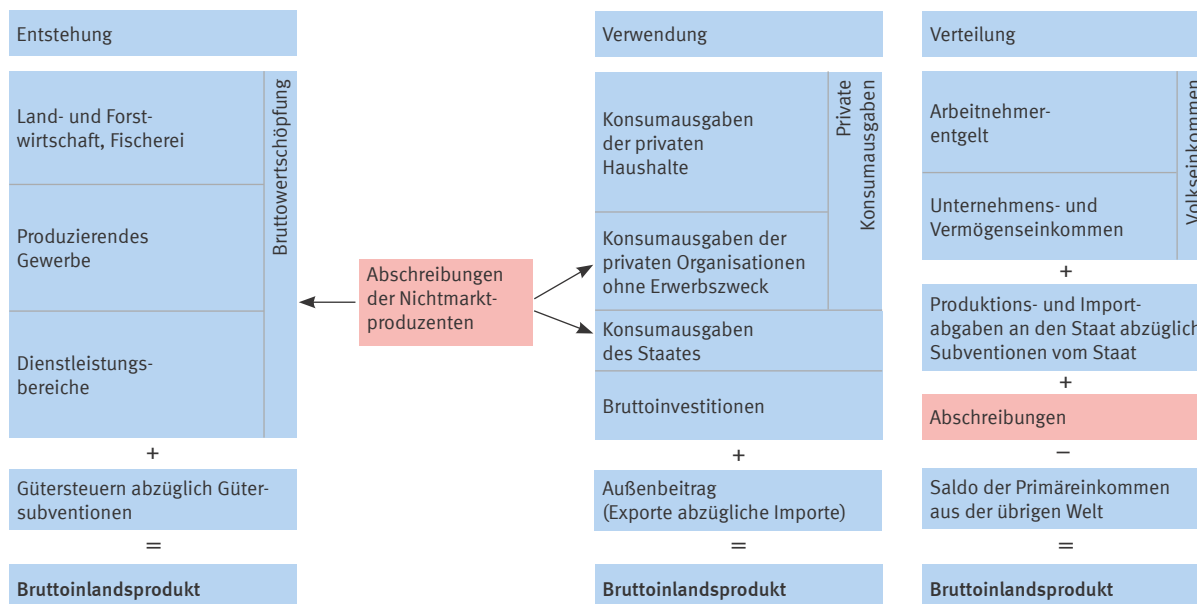
Die Aufgabe der Vermögensrechnung – als weiterer Bestandteil der Volkswirtschaftlichen Gesamtrechnungen – besteht darin, sowohl die Höhe als auch die Zusammensetzung und die Veränderung von Vermögensbeständen zu ermitteln. In Deutschland errechnet das Statistische Bundesamt das nichtfinanzielle Vermögen (oftmals auch als Sachvermögen bezeichnet), während die Deutsche Bundesbank das Geldvermögen ermittelt. Die Ergebnisse werden in integrierten Vermögensbilanzen zusammengetragen und gemeinsam veröffentlicht (Statistisches Bundesamt, 2019).

Die Anlagevermögensrechnung ist der wichtigste Teil der Vermögensrechnung des Statistischen Bundesamtes. Durch Bruttoanlageinvestitionen – einer Verwendungsposition des Bruttoinlandsprodukts – wird ein Kapitalstock in Form des Bruttoanlagevermögens aufgebaut. Die Anlagegüter werden mehrere Jahre zu Produktionszwecken genutzt und über ihre Nutzungsdauer abgeschrieben. Die Anlagevermögensrechnung bildet das Anlagevermögen sowie Abschreibungen und Abgänge ab. Neben dem eigenständigen Zweck der (Anlage-)Vermögensrechnung werden Teilergebnisse daraus auch für die Entstehungsrechnung benötigt. Die Abschreibungen der Nichtmarktproduzenten sind ein notwendiger Baustein für die Ermittlung der Wertschöpfung der Nichtmarktproduktion und somit auch für die Ermittlung des Bruttoinlandsprodukts. Verwendungsseitig schlagen sie sich in den Konsumausgaben des Staates sowie der privaten Organisationen ohne Erwerbszweck nieder. Die gesamtwirtschaftlichen Abschreibungen sind ein Teilaggregat der Verteilung des Bruttoinlandsprodukts.

➤ Grafik 1

Grafik 1

Bedeutung der Abschreibungen bei der Berechnung des Bruttoinlandsprodukts



2020-01-0214

Bruttoanlageinvestitionen und Abschreibungen sind wie die meisten VGR-Aggregate Stromgrößen. Sie beziehen sich, wie das Bruttoinlandsprodukt, auf einen Zeitraum, beispielsweise ein Jahr oder ein Vierteljahr. Demgegenüber ist das Anlagevermögen eine der wenigen zeitpunktbezogenen Bestandsgrößen, die es in den Volkswirtschaftlichen Gesamtrechnungen gibt. Die beiden Produktionsfaktoren Arbeit und (Sach-)Kapital entsprechen in den Volkswirtschaftlichen Gesamtrechnungen den Erwerbstätigen und dem Anlagevermögen. Beim Anlagevermögen wird zwischen Brutto- und Nettobeständen unterschieden (siehe Kapitel 4).

Die Vermögensrechnung basiert, wie die Inlandsproduktberechnung und alle weiteren Rechnungen innerhalb der Volkswirtschaftlichen Gesamtrechnungen, auf den verbindlichen Konzepten des Europäischen Systems Volkswirtschaftlicher Gesamtrechnungen (ESVG) 2010 (Eurostat, 2014). Das ESGV 2010 gibt sowohl die im folgenden Kapitel erläuterte Klassifikation der Vermögensgüter als auch die Berechnungsvorschriften für Vermögensbestände und Abschreibungen sowie deren Bewertung vor. Das Lieferprogramm zum ESGV 2010 regelt darüber hinaus, wann welche Ergebnisse an das Statistische Amt der Europäischen Union (Eurostat) zu

übermitteln sind. Die Vermögensrechnung liefert die Ergebnisse für die Tabellen 20 „Kreuztabelle des Anlagevermögens nach Wirtschaftsbereichen und Anlagearten“ und 26 „Nichtfinanzielle Vermögensbilanzen“ sowie die Abschreibungen für eine Reihe weiterer Tabellen des ESGV-Lieferprogramms.

3

Klassifikation der nichtfinanziellen Vermögensgüter

Die Klassifikation der nichtfinanziellen Vermögensgüter gibt vor, welche Güter in den Volkswirtschaftlichen Gesamtrechnungen zum nichtfinanziellen Vermögen gehören. Sie gilt für Strom- und Bestandsgrößen wie die Bruttoinvestitionen und das Anlagevermögen. Die Abgrenzung der nichtfinanziellen Vermögensgüter hat sich im Zeitverlauf geändert. So wurden mit dem Übergang auf das ESGV 1995 (Europäische Union, 1996) unter anderem die immateriellen Vermögensgüter eingeführt. Mit dem derzeit gültigen ESGV 2010 wurde die Vermögensgrenze erneut ausgedehnt, sodass unter an-

derem Forschung und Entwicklung sowie militärische Waffensysteme seit der Generalrevision der Volkswirtschaftlichen Gesamtrechnungen im Jahr 2014 zu den Anlagegütern zählen.¹

In [Grafik 2](#) sind die Gliederungsebenen der nichtfinanziellen Vermögensgüter schematisch dargestellt. Die erste Gliederungsebene unterscheidet produzierte von nichtproduzierten Vermögensgütern: Durch wirtschaftliches Geschehen können nur produzierte Vermögensgüter gemehrt werden, während nichtproduzierte Vermögensgüter, insbesondere die natürlichen Ressourcen, lediglich genutzt (zum Beispiel Grund und Boden) beziehungsweise verbraucht (zum Beispiel Bodenschätze) werden. Produzierte Vermögensgüter sind Ergebnisse von Produktionsprozessen und ihre Produktion wird im Kontensystem der Volkswirtschaftlichen Gesamtrechnungen erfasst. Demgegenüber sind

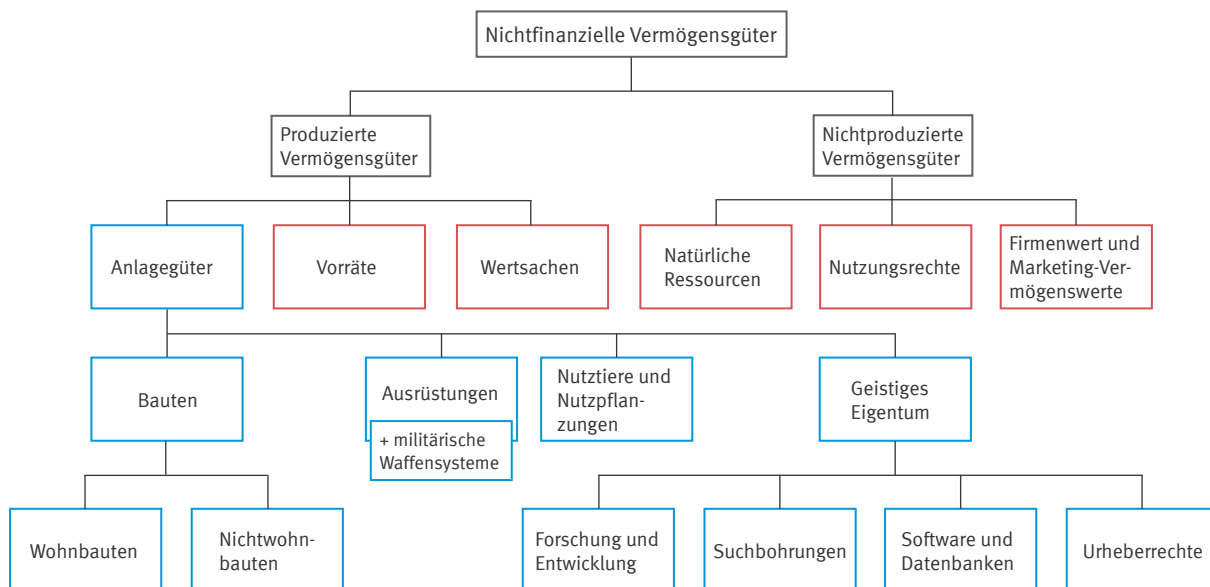
nichtproduzierte Vermögensgüter nicht durch einen Produktionsprozess entstanden, sondern werden nach der Art ihrer Entstehung gegliedert. Sie kommen entweder in der Natur vor (natürliche Ressourcen) oder sie sind durch rechtliche oder buchhalterische Regelungen entstandene, von der Gesellschaft entwickelte Konstrukte (Nutzungsrechte sowie Firmenwert und einzeln veräußerbare Marketing-Vermögenswerte).

«Die produzierten Vermögensgüter (...) werden nach ihrer Rolle im Produktionsprozess gegliedert. Dazu zählen Anlagegüter, die länger als ein Jahr wiederholt oder dauerhaft im Produktionsprozess eingesetzt werden, Vorräte, die als Vorleistungen im Produktionsprozess verbraucht, verkauft oder anderweitig verwendet werden, sowie Wertsachen, die nicht in erster Linie für Zwecke der Produktion oder des Konsums verwendet, sondern primär als Wertaufbewahrungsmittel erworben werden.» (Eurostat, 2014, Ziffer 7.23) Die Käufe und Verkäufe sowie die Eigenproduktion von produzierten Vermögensgütern werden in den Bruttoinvestitionen nachgewiesen.

Die in Grafik 2 blau gekennzeichneten Anlagegüter sind Bestandteil der Bruttoanlageinvestitionen; für sie wer-

1 Eine historische Darstellung zur Klassifikation der nichtfinanziellen Vermögensgüter ist in Gühler/Schmalwasser (2019), Seite 117 ff., nachzulesen. Die Einführung des ESGV 2010 mit der Generalrevision 2014 beschreiben Räh/Braakmann (2014). Eine ausführliche Darstellung der Buchung von Forschung und Entwicklung in den Volkswirtschaftlichen Gesamtrechnungen findet sich in Adler und andere (2014), die Berechnung von Vermögensbeständen und Abschreibungen auf Seite 712 f.

Grafik 2
Gliederung der nichtfinanziellen Vermögensgüter



 = Bestandteile des Anlagevermögens und der Bruttoanlageinvestitionen
 = Übrige nichtfinanzielle Vermögensgüter

2020-01-0215

den innerhalb der Anlagevermögensrechnung Bestände und Abschreibungen sowie weitere Stromgrößen berechnet. Die Anlagegüter umfassen Bauten, Ausrüstungen einschließlich militärischer Waffensysteme, Nutztiere und Nutzpflanzungen sowie geistiges Eigentum. Geistiges Eigentum wurde früher als immaterielle Anlagen bezeichnet; es umfasst Forschung und Entwicklung, Urheberrechte, Software und Datenbanken sowie Suchbohrungen. Bei den Vorräten werden Vorleistungsgüter, unfertige Erzeugnisse, Fertigerzeugnisse und Handelsware unterschieden. Zu den Wertsachen gehören Edelmetalle und Edelsteine, Antiquitäten und Kunstgegenstände sowie sonstige Wertsachen (beispielsweise besonders wertvoller Schmuck).

Über das Anlagevermögen hinaus wird in der Vermögensrechnung der Wert für den Grund und Boden als bedeutendster Bestandteil der natürlichen Ressourcen ermittelt (Schmalwasser/Brede, 2015) und in den gemeinsam mit der Deutschen Bundesbank herausgegebenen sektoralen und gesamtwirtschaftlichen Vermögensbilanzen veröffentlicht (Statistisches Bundesamt, 2019). Vorratsbestände werden als Teil des ESVG-Lieferprogramms berechnet und für interne Zwecke an Eurostat übermittelt, jedoch nicht veröffentlicht, da die Datenqualität bisher nicht den Anforderungen der Volkswirtschaftlichen Gesamtrechnungen entspricht.

4

Die Anlagevermögensrechnung – Brutto- und Nettokonzept

Das Anlagevermögen und die zugehörigen Stromgrößen werden in Deutschland anhand der Kumulationsmethode – Perpetual Inventory Method (PIM) – ermittelt.¹² Diese Methode geht von der Grundüberlegung aus, dass der heute vorhandene Vermögensbestand aus den kumulierten Bruttoanlageinvestitionen der Vergangenheit erstellt wurde. Mittels spezifischer Nutzungsdauern der Anlagegüter lässt sich für die Zugänge/die Investitionen der zurückliegenden Jahre berechnen, welcher Anteil sich am Ende des Berichtsjahrs noch im Bestand

befindet und wann diese Güter rechnerisch endgültig aus dem Bestand ausscheiden. Zusammen mit der Abschreibungsmethode können daraus die Abschreibungen jeder Berichtsperiode ermittelt werden. Voraussetzung für die Anwendung der Kumulationsmethode sind weit in die Vergangenheit zurückreichende Investitionsreihen sowie differenzierte Nutzungsdaueransätze für die einzelnen Anlagegüterarten.¹³

↳ VGR-Revision 2019 – Aufteilung der Nichtwohnbauten in Hoch- und Tiefbauten

Die Generalrevision der Volkswirtschaftlichen Gesamtrechnungen im Jahr 2019 brachte viele Änderungen, die auch Einfluss auf die Vermögensrechnung in Deutschland hatten. Die meisten Revisionsänderungen mit Bezug zur Vermögensrechnung wurden bereits beschrieben (Hauf/Schäfer, 2019). Insbesondere die dort dargestellten Änderungen bei den Bruttoanlageinvestitionen wurden in vollem Umfang auch für das Anlagevermögen umgesetzt. Eine weitere wichtige Neuerung für die Anlagevermögensrechnung ist, dass in den nichtfinanziellen Vermögensbilanzen nun die getrennte Darstellung der Nichtwohnbauten in Nichtwohngebäude (Hochbauten) und sonstige Bauten (Tiefbauten) möglich ist.

Bisher war die getrennte Darstellung nur für den Sektor Staat möglich: Seit der erstmaligen Berechnung der Abschreibungen auf öffentliche Tiefbauten zur VGR-Revision 1999 ist die komplette Trennung von staatlichen Hoch- und Tiefbauten rechentechnisch umgesetzt. In den nichtstaatlichen Sektoren¹⁴ konnten Hoch- und Tiefbauten bislang nicht voneinander getrennt werden, weil die langen Investitionsreihen nur für die gewerblichen Bauten ohne weitere Untergliederungen vorlagen. Deutschland nutzte hierzu eine Ausnahmeregelung zum ESVG-Lieferprogramm; nach der Revision 2019 ist dies nicht mehr notwendig.

Für die Vermögensrechnung müssen die Investitionen in Nichtwohnbauten aus der Inlandsproduktberechnung sektoral nach Hoch- und Tiefbauten getrennt werden. Zur Wahrung der ganzheitlichen Konsistenz der Ergebnisse wurde die Hochbau-Tiefbau-Trennung sowohl für die nichtstaatlichen Sektoren als auch für die Wirtschafts-

2 Eine Ausnahme bilden Nutztiere und Nutzpflanzungen, für die direkte Bestandsangaben zu Anbauflächen sowie Tierbeständen aus der Landwirtschaftsstatistik und entsprechende Preise genutzt werden, um die notwendigen VGR-Angaben zu generieren.

3 Ausführliche Erläuterungen zum mathematischen Modell enthalten die VGR-Methodenbeschreibung (Statistisches Bundesamt, 2016, hier: Kapitel 4.12) sowie ein Aufsatz zur Kapitalstockrechnung (Schmalwasser/Schidlowski, 2006). Das aktuell verwendete Berechnungsverfahren wird in Schmalwasser/Weber (2012) näher erläutert (Seite 935 f.).

4 Das sind nichtfinanzielle und finanzielle Kapitalgesellschaften, private Haushalte und private Organisationen ohne Erwerbszweck.

zweige in der Rechartiefe der Vermögensrechnung umgesetzt. Hoch- und Tiefbauanteile für die verschiedenen Wirtschaftszweige, die im Rahmen der Investorenrechnung ermittelt werden, dienen als Anhaltspunkt für die Aufteilung ab dem Jahr 1991. Für die Jahre vor 1991 fand ein vereinfachtes Verfahren Anwendung, indem die Investitionen mit den Mittelwerten der Hochbau- und Tiefbauanteile der Jahre 1991 bis 2017 aufgeteilt wurden.

Für jeden Wirtschaftszweig gab es bisher einen Nutzungsdaueransatz, welcher – teilweise nur implizit – die Nutzungsdauern der verschiedenen Hoch- und Tiefbauten enthielt. Diese Nutzungsdaueransätze wurden nach Hoch- und Tiefbauten getrennt, dabei fließen die Nutzungsdauern für Modernisierungsmaßnahmen an bestehenden Gebäuden in die Nutzungsdaueransätze für die Hochbauten ein. Gleichzeitig wurden die Nutzungsdauern überarbeitet.

Durch diese Arbeiten zur Trennung von Hoch- und Tiefbauten ergaben sich größere Änderungen beim Anlagevermögen an und den Abschreibungen auf Bauten sowohl der Wirtschaftszweige als auch der Sektoren.

Die Vermögensbestände sind gemäß ESVG-Lieferprogramm nach dem Brutto- und dem Nettokonzepth jeweils zu Wiederbeschaffungspreisen⁵ des Berichtsjahres und zu Wiederbeschaffungspreisen des Vorjahres nachzuweisen. Das Bruttoanlagevermögen bildet die in der Produktion eingesetzten Anlagegüter ab. Es ist ein Maß für die Produktionskapazität. Das Nettoanlagevermögen ergibt sich aus dem Bruttoanlagevermögen, wenn alle kumulierten Abschreibungen der aktiven Anlagegüter vom Bruttoanlagevermögen abgezogen werden. Es stellt damit den (Zeit-)Wert dar. Nachfolgend werden beide Konzepte näher erläutert und die komplette Überleitung vom Jahresanfangsbestand zum Jahresendbestand für die Volkswirtschaft insgesamt, den Staat und die nicht-staatlichen Sektoren vorgestellt.⁶

5 Unter Wiederbeschaffungspreisen wird der Betrag verstanden, der hätte gezahlt werden müssen, wenn die Anlagegüter zum Berichtszeitpunkt angeschafft worden wären. Die Anlagegüter werden mit den jeweiligen Preisen der Berichtsperiode bewertet – als wären sie in der Berichtsperiode neu angeschafft worden (beim Bruttokonzepth beziehungsweise unter Abzug der aufgelaufenen Abschreibungen (beim Nettokonzepth)).

6 In der Fachserie 18 Reihe 1.4 werden das Netto- und das Bruttoanlagevermögen zu Wiederbeschaffungspreisen und preisbereinigt (dargestellt durch den Kettenindex) als Bestand am Jahresende veröffentlicht (Tabelle 3.1.3). Der Nachweis der Abschreibungen (Tabelle 3.1.4) und der Abgänge (Tabelle 3.1.5) erfolgt in jeweiligen Preisen und preisbereinigt. Außerdem zeigen die Tabellen 3.2.23 bis 3.2.28 Vermögensbestände und Abschreibungen für 64 Wirtschaftsbereiche (Statistisches Bundesamt, 2020).

4.1 Bruttoanlagevermögen und Abgänge

Beim Bruttokonzepth bleiben die Anlagegüter bis zum endgültigen Ausscheiden aus dem Produktionsprozess mit ihrem vollen Wert (Neuwert), ohne Berücksichtigung der Wertminderung, im Vermögensbestand. Dahinter steckt die Annahme, dass im Produktionsprozess jeweils das ganze Anlagegut eingesetzt wird und während der Nutzungsdauer in etwa den gleichen jährlichen Produktionsoutput ermöglicht – unabhängig von seinem Alter. Regelmäßige Wartung und Reparaturen werden vorausgesetzt. Das Bruttoanlagevermögen steht in den Volkswirtschaftlichen Gesamtrechnungen für die Produktionskapazität. Der Übergang vom Jahresanfangs- zum Jahresendbestand sieht beim Bruttoanlagevermögen zu Wiederbeschaffungspreisen wie folgt aus:

$$\begin{aligned} & \text{Bruttobestand am Jahresanfang (entspricht dem} \\ & \text{Bruttobestand am Jahresende des Vorjahres)} \\ & + \text{ Zugänge zum Anlagevermögen} \\ & - \text{ Abgänge vom Anlagevermögen (einschließlich} \\ & \text{Sonderabgängen)} \\ & + \text{ Umbewertungsgewinne/-verluste – brutto} \\ & = \text{ Bruttobestand am Jahresende} \end{aligned}$$

Die jährlichen Zugänge zum Anlagevermögen sind die Bruttoanlageinvestitionen aus der Inlandsproduktberechnung. Sie enthalten zum einen die Käufe neuer Anlagegüter und zum anderen die Käufe abzüglich der Verkäufe gebrauchter Anlagegüter, ebenso die durch die Investoren selbst erstellten Anlagen. Wie bereits im dritten Kapitel definiert, zählen zu den Anlagegütern produzierte Güter, die länger als ein Jahr in der Produktion eingesetzt werden. Ausgenommen sind geringwertige nicht aktivierte Güter wie Kleinwerkzeuge und Büromittel. Größere Reparaturen, die zu einer wesentlichen Wertsteigerung des Anlageguts führen und/oder dessen Nutzungsdauer verlängern, sind dagegen als Bruttoanlageinvestitionen abzubilden.

Die Abgänge umfassen Anlagevermögensgüter, die in einer Periode endgültig durch Verschrottung und Abbruch aus dem Bruttobestand ausscheiden, weil sie das Ende ihrer Nutzungsdauer erreicht haben. Da keine statistischen Angaben über Abgänge vorliegen, müssen sie mithilfe der Kumulationsmethode geschätzt werden. Hierzu sind Annahmen über die durchschnittlichen Nut-

zungsdauern und die Abgangsfunktion der Anlagegüter erforderlich. Es wird angenommen, dass die Abgänge etwa glockenförmig um die durchschnittliche Nutzungsdauer verteilt sind. Neben den planmäßigen, durch die Kumulationsmethode geschätzten Abgängen werden – soweit Informationen vorliegen – außerdem Abgänge berücksichtigt, wenn Anlagegüter vorzeitig aus unerwarteten Gründen aus dem Bestand ausscheiden. Als Sonderabgänge gebucht wurden beispielsweise die Verluste an Anlagevermögen durch die Hochwasser im Sommer 2002 und Frühjahr 2013 sowie die vorzeitige Abschaltung von acht Kernkraftwerken im Jahr 2011.

Das Bruttoanlagevermögen zwischen zwei Zeitpunkten wird bei der Bewertung zu Wiederbeschaffungspreisen nicht nur durch die Investitionen erhöht und durch die Abgänge vermindert. Zusätzlich spielen auch Wertänderungen durch Umbewertungen¹⁷ der Vermögensbestände eine erhebliche Rolle. Diese Wertänderungen resultieren aus Preisänderungen für neue Anlagegüter. Vor allem in Zeiten starker Preissteigerungen führen

Umbewertungen zu höheren Zuwächsen des Vermögens. Sie ergeben sich aus folgender Differenz:¹⁸

Bruttobestand am Jahresende (zu Wiederbeschaffungspreisen des Berichtsjahres)

- Bruttobestand am Jahresanfang (entspricht dem Bruttobestand am Jahresende des Vorjahres, das heißt zu Wiederbeschaffungspreisen des Vorjahres)
- Zugänge zum Anlagevermögen (Investitionen in jeweiligen Preisen)
- + Abgänge vom Anlagevermögen (in jeweiligen Preisen)
- = Umbewertungsgewinne/-verluste – brutto (im Berichtsjahr)

7 Nominale Umbewertungsgewinne/-verluste (K.7) werden im ESVG 2010 von Ziffer 6.26 bis Ziffer 6.45 behandelt (Eurostat, 2014).

8 Die Umbewertungsgewinne/-verluste lassen sich wie hier dargestellt als Restwert ermitteln. Eine andere Form der Berechnung ist die Multiplikation der Preisänderung der jeweiligen Güterart im Berichtsjahr mit dem Jahresanfangsbestand zu Wiederbeschaffungspreisen des Vorjahres dieser Güterart. Das erläutern Schmalwasser/Brede (2015, hier: Seite 47) am Beispiel Grund und Boden.

Tabelle 1

Bruttoanlagevermögen zu Wiederbeschaffungspreisen – Überleitung vom Anfangsbestand zum Endbestand

	1991	1995	2000	2005	2010	2015	2018
Mrd. EUR							
Volkswirtschaft insgesamt							
Bruttobestand am Jahresanfang	7 317,9	9 514,3	10 828,4	12 027,9	14 414,0	16 985,0	18 753,9
+ Zugänge zum Anlagevermögen	398,7	450,3	492,6	441,6	508,4	613,2	711,3
– Abgänge vom Anlagevermögen (einschließlich Sonderabgängen)	151,8	189,1	233,5	281,9	340,4	406,4	458,9
+ Umbewertungsgewinne/-verluste – brutto	317,4	155,0	69,8	97,5	195,2	289,6	751,6
Bruttobestand am Jahresende	7 882,3	9 930,5	11 157,3	12 285,1	14 777,2	17 481,5	19 757,9
Staat							
Bruttobestand am Jahresanfang	1 319,0	1 635,6	1 710,4	1 812,6	2 126,0	2 434,8	2 636,2
+ Zugänge zum Anlagevermögen	49,7	49,8	49,1	44,0	60,4	64,8	74,5
– Abgänge vom Anlagevermögen (einschließlich Sonderabgängen)	20,7	26,2	29,9	34,0	41,7	52,1	61,2
+ Umbewertungsgewinne/-verluste – brutto	59,9	19,2	12,2	13,0	30,3	35,9	126,0
Bruttobestand am Jahresende	1 407,8	1 678,5	1 741,7	1 835,7	2 175,1	2 483,3	2 775,5
Nichtstaatliche Sektoren							
Bruttobestand am Jahresanfang	5 998,9	7 878,8	9 118,0	10 215,3	12 288,0	14 550,2	16 117,7
+ Zugänge zum Anlagevermögen	349,1	400,5	443,6	397,5	447,9	548,4	636,8
– Abgänge vom Anlagevermögen (einschließlich Sonderabgängen)	131,0	162,9	203,7	247,9	298,8	354,2	397,7
+ Umbewertungsgewinne/-verluste – brutto	257,5	135,7	57,6	84,5	164,9	253,8	625,6
Bruttobestand am Jahresende	6 474,5	8 252,1	9 415,6	10 449,5	12 602,0	14 998,2	16 982,4

Die vollständige Überleitung von den Jahresanfangsbeständen zu den Jahresendbeständen des Bruttoanlagevermögens einschließlich der Umbewertungsgewinne/-verluste zeigt beispielhaft [Tabelle 1](#).

4.2 Nettoanlagevermögen und Abschreibungen

In der Vermögensbilanz wären gemäß ESVG 2010 die Anlagegüter mit ihren Marktpreisen (Zeitwert) am Bilanzstichtag zu bewerten. Für die meisten Anlagegüter – mit Ausnahme von Kraftfahrzeugen – gibt es in der Regel keine Marktpreise für gebrauchte Anlagen verschiedenen Alters. Daher werden beim Nettokonzept die seit dem Investitionszeitpunkt aufgelaufenen Abschreibungen abgezogen. Aus dieser Berechnung ergibt sich das Nettoanlagevermögen zu Wiederbeschaffungspreisen, das damit der aktuellen Vermögensposition im Sinne des Zeitwerts des Anlagevermögens entspricht. Es gilt folgende Staffelrechnung:

Nettobestand am Jahresanfang (entspricht dem
Nettobestand am Jahresende des Vorjahres)

- + Zugänge zum Anlagevermögen
- Abschreibungen auf das Anlagevermögen
- + Sonstige reale Änderungen des Nettovermögens
- + Umbewertungsgewinne/-verluste – netto
- = Nettobestand am Jahresende

Die Abschreibungen messen laut ESVG 2010 Ziffer 3.139 „die Wertminderung von Anlagegütern durch normalen Verschleiß und wirtschaftliches Veralten.“ (Eurostat, 2014). In den Volkswirtschaftlichen Gesamtrechnungen stellen die Abschreibungen somit den Aufwand an Anlagevermögen beziehungsweise den Wertverlust des Anlagevermögens, der im Produktionsprozess verbraucht wird, dar. Die volkswirtschaftlichen Abschreibungen sind ebenso wie die steuer- und handelsrechtlichen Abschreibungen ein kalkulatorischer Posten, da sie nicht direkt messbar sind. Sie unterscheiden sich jedoch hinsichtlich ihrer Bewertung, der Abgrenzung der abzuschreibenden Güter und der verwendeten Nut-

zungsdauern von den nach steuer- und handelsrechtlichen Grundsätzen ermittelten Abschreibungen.¹⁹

Abgeschrieben wird in den Volkswirtschaftlichen Gesamtrechnungen das gesamte Anlagevermögen mit Ausnahme der Nutztiere. Dabei werden die Abschreibungen, dem ESVG 2010 Ziffer 3.143 folgend, anhand der linearen Abschreibungsmethode berechnet. Der Wert der Anlageinvestitionen wird mit jährlich konstanten Beträgen entsprechend der erwarteten wirtschaftlichen Nutzungsdauer der einzelnen Güter auf die Gesamtzeit ihrer Nutzung verteilt.¹¹⁰ Somit ist der Abschreibungssatz eines Investitionsguts der Kehrwert der Nutzungsdauer. Ausnahmen bilden das Investitions- und das Abgangsjahr. In diesen Jahren ist der Abschreibungssatz lediglich halb so hoch, weil angenommen wird, dass sich die Anlagegüter im Durchschnitt nur jeweils ein halbes Jahr im Bestand befinden. Die Nutzungsdauer in den Volkswirtschaftlichen Gesamtrechnungen umfasst den Zeitraum, der zwischen dem Zugang der Anlagen zum Bestand und dem endgültigen Ausscheiden der Anlagen aus dem Produktionsprozess liegt. Sie ist eine in die Zukunft gerichtete – und somit unsichere – zu schätzende Größe.

Wenn Anlagegüter aus unerwarteten Gründen vorzeitig aus dem Vermögensbestand ausscheiden, wird ihr Restwert nicht mehr abgeschrieben, sondern als sonstige reale Vermögensänderung gebucht.¹¹¹ Sonderabschreibungen sind im Gegensatz zum Steuer- und Handelsrecht in den Volkswirtschaftlichen Gesamtrechnungen nicht zulässig. Die bereits genannten Hochwasserschäden der Jahre 2002 und 2013, die vorzeitige Abschaltung von acht Kernkraftwerken im Jahr 2011 sowie das frühere Ausscheiden von Wohnbauten der ehemaligen DDR (überwiegend Plattenbauten) im Rahmen des Programms Stadtumbau Ost gehen momentan als sonstige reale Vermögensänderungen in die Anlagevermögensrechnung ein.

9 Eine umfassende Gegenüberstellung der Grundsätze der Berechnung und Bewertung der Abschreibungen in den Volkswirtschaftlichen Gesamtrechnungen und im einzelwirtschaftlichen Rechnungswesen enthält der Aufsatz von Schäfer und Schmidt (Schäfer/Schmidt, 1983).

10 Diese Rechnung erfolgt auf preisbereinigter Basis, das heißt auf der Grundlage tief gegliederter Investitionsreihen in Form verketteter Volumenangaben (Statistisches Bundesamt, 2016, hier: Seite 310).

11 Das ESVG 2010 geht in den Ziffern 6.03 bis 6.25 ausführlich auf die sonstigen realen Vermögensänderungen (K.1 bis K.6) ein (Eurostat, 2014).

Tabelle 2

Nettoanlagevermögen zu Wiederbeschaffungspreisen – Überleitung vom Anfangsbestand zum Endbestand

	1991	1995	2000	2005	2010	2015	2018
Mrd. EUR							
Volkswirtschaft insgesamt							
Nettobestand am Jahresanfang	4 643,0	5 989,5	6 665,4	7 165,6	8 333,2	9 550,4	10 399,6
+ Zugänge zum Anlagevermögen	398,7	450,3	492,6	441,6	508,4	613,2	711,3
– Abschreibungen auf das Anlagevermögen	245,9	309,3	356,2	393,1	462,0	541,2	608,7
+ Sonstige reale Änderungen des Nettovermögens	0,0	-0,1	-0,1	-0,7	-0,7	0,0	0,0
+ Umbewertungsgewinne/-verluste – netto	203,7	97,8	36,5	54,0	105,5	156,1	410,3
Nettobestand am Jahresende	4 999,5	6 228,3	6 838,2	7 267,4	8 484,3	9 778,6	10 912,4
Staat							
Nettobestand am Jahresanfang	829,9	1 002,3	1 001,6	1 016,2	1 146,5	1 270,0	1 350,5
+ Zugänge zum Anlagevermögen	49,7	49,8	49,1	44,0	60,4	64,8	74,5
– Abschreibungen auf das Anlagevermögen	37,8	44,8	46,3	48,7	57,0	66,9	75,5
+ Sonstige reale Änderungen des Nettovermögens	0,0	0,0	0,0	0,0	0,0	0,0	0,0
+ Umbewertungsgewinne/-verluste – netto	38,3	12,2	7,5	7,3	16,3	18,5	64,6
Nettobestand am Jahresende	880,1	1 019,4	1 011,8	1 018,9	1 166,3	1 286,3	1 414,1
Nichtstaatliche Sektoren							
Nettobestand am Jahresanfang	3 813,1	4 987,2	5 663,7	6 149,3	7 186,6	8 280,4	9 049,1
+ Zugänge zum Anlagevermögen	349,1	400,5	443,6	397,5	447,9	548,4	636,8
– Abschreibungen auf das Anlagevermögen	208,1	264,4	309,9	344,5	405,0	474,2	533,3
+ Sonstige reale Änderungen des Nettovermögens	0,0	-0,1	-0,1	-0,7	-0,7	0,0	0,0
+ Umbewertungsgewinne/-verluste – netto	165,4	85,7	29,0	46,7	89,2	137,7	345,7
Nettobestand am Jahresende	4 119,4	5 208,8	5 826,4	6 248,4	7 318,1	8 492,3	9 498,3

Auch für die Nettobestände werden die Wertänderungen auf Grundlage von Umbewertungen als Differenz ermittelt. Es gilt die nachfolgende Berechnungsvorschrift:

Nettobestand am Jahresende (zu Wiederbeschaffungspreisen des Berichtsjahres)

- Nettobestand am Jahresanfang (entspricht dem Nettobestand am Jahresende des Vorjahres, das heißt zu Wiederbeschaffungspreisen des Vorjahres)
- Zugänge zum Anlagevermögen (Investitionen in jeweiligen Preisen)
- + Abschreibungen auf das Anlagevermögen (in jeweiligen Preisen)
- Sonstige reale Änderungen des Nettovermögens (in jeweiligen Preisen)
- = Umbewertungsgewinne/-verluste – netto (im Berichtsjahr)

Mit den ermittelten Umbewertungsgewinnen beziehungsweise -verlusten lässt sich für die Nettorechnung

der Übergang vom Jahresanfangsbestand zum Jahresendbestand darstellen. ➔ Tabelle 2

5

Ersatzinvestitionen in Höhe der Abschreibungen?

In der öffentlichen wissenschaftlichen Diskussion wird im Allgemeinen gefordert, dass Investitionen mindestens in Höhe der Abschreibungen getätigt werden müssen, um den Kapitalstock und damit die Produktionskapazität zu erhalten. Nach dieser Auffassung bildet das Netto- und nicht das Bruttoanlagevermögen den Kapitalstock, weil das Nettoanlagevermögen in der Theorie gleich den abdiskontierten künftigen Erträgen ist, und somit der Produktionskapazität entspräche. Das ist aus volkswirtschaftlicher Sicht der Güterproduktion zu hinterfragen.

Während der Nutzungsdauer wird das ganze Anlagegut im Produktionsprozess eingesetzt und ermöglicht eine mehr oder weniger gleichbleibende Nutzung über die gesamte Lebensdauer, also solange es nicht demontiert oder verschrottet wird. Die jährliche Produktionskapazität gemessen in physischen Einheiten, wie Megawattstunden bei Kraftwerken oder Tonnenkilometer bei Lastkraftwagen, bleibt über die gesamte Nutzungsdauer relativ konstant – vorausgesetzt, das Anlagegut wird regelmäßig gewartet und instand gehalten. Die Produktionskapazität stellt den (potenziellen) Beitrag zum realen jährlichen Bruttoinlandsprodukt dar. Sie verringert sich nur marginal im Laufe der Nutzungsdauer, weil die Wartungsintervalle möglicherweise verkürzt werden müssen oder doch häufiger Reparaturen anfallen.

Diese reale jährliche Produktionskapazität ist zu unterscheiden von der wertmäßigen Restproduktionskapazität eines Anlageguts über die Restnutzungsdauer, wie sie im Leitbild der abdiskontierten künftigen Erträge zum Ausdruck kommt. Die reale jährliche Produktionskapazität wird vom Brutto- und nicht vom Nettoanlagevermögen abgebildet. Aus volkswirtschaftlicher Sicht ist sie aber ein wichtiger analytischer Aspekt, der durch die international weit verbreitete Beschränkung auf die Nettobetrachtung und das Nettoanlagevermögen nicht abgedeckt wird.

Es ist ein Unterschied, ob durch Investitionen der Wertverlust des Anlagevermögens in Höhe der Abschreibungen kompensiert werden soll oder ob es die tatsächlichen physischen Abgänge aus dem Anlagevermögen zu ersetzen gilt. Es ist zu unterscheiden zwischen Wertverlust beziehungsweise Wertersatz auf der einen und physischem Abgang beziehungsweise physischem Ersatz auf der anderen Seite. Im ersten Fall geht es um die Werterhaltung des Anlagevermögens, ausgedrückt im Nettoanlagevermögen, im zweiten Fall um den Erhalt der jährlichen Produktionskapazität, die im Bruttoanlagevermögen zum Ausdruck kommt.

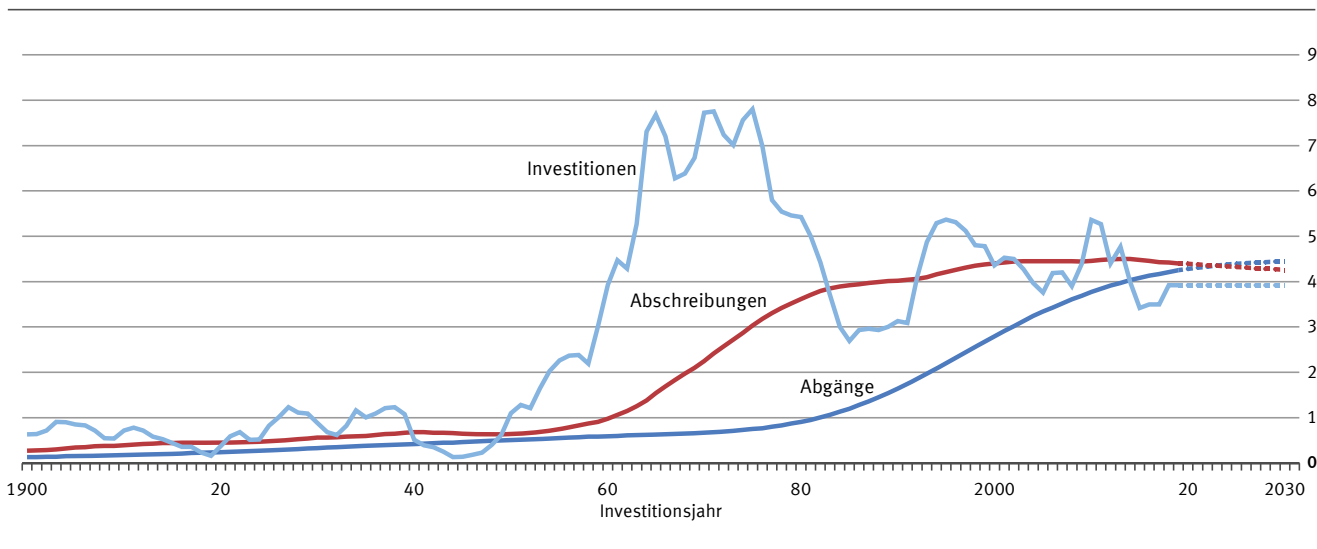
Je nach Erkenntnisziel hat auch das Bruttokonzept neben dem weit verbreiteten und auch der Vermögensbilanz zugrunde liegenden Nettokonzept seine Berechtigung. Die Entwicklung des Bruttoanlagevermögens unterscheidet sich insbesondere dann von der des Nettoanlagevermögens, wenn die Entwicklung der Investitionen keine konstanten Wachstumsraten aufweist und wenn es sich um Anlagegüter mit einer langen

Nutzungsdauer handelt. Beides trifft beispielsweise auf die Infrastrukturbauten des Staates zu. Hinzu kommt, dass nach der Aufbauphase infolge des zweiten Weltkriegs auch Sättigungseffekte auftraten, die in einigen Bereichen, wie Schulen oder Straßen, sehr gut zu beobachten sind. Durch die deutsche Vereinigung kam es in verschiedenen Bereichen, insbesondere bei Straßen und bei Entsorgungsanlagen, erneut zu stärkeren Investitionen. In der Folge ist gerade bei öffentlichen Bauten eine eher wellenförmige Investitionsentwicklung zu erkennen. Dadurch fallen Abschreibungen und Abgänge stärker auseinander, als es zum Beispiel bei den Ausrüstungen oder beim geistigen Eigentum der Fall ist, die eine viel kürzere Nutzungsdauer haben. Wegen der Sättigungseffekte sind Investitionen in Höhe der tatsächlichen Abgänge, aber nicht in Höhe der Abschreibungen notwendig, um die Kapazitäten der öffentlichen Infrastruktur zu erhalten.

Die Bedeutung der langfristigen Wellen bei den Bauinvestitionen ist ein weiterer Aspekt, der in der neueren Diskussion zur notwendigen Höhe öffentlicher Infrastrukturinvestitionen (Expertenkommission „Stärkung von Investitionen in Deutschland“, 2015; Grömling und andere, 2019; Feld und andere, 2019; Bardt und andere, 2019) berücksichtigt werden sollte. Wie sich wellenförmige Investitionsentwicklungen auf die Abgänge und Abschreibungen auswirken, zeigt [Grafik 3](#) auf Seite 86. Zugrunde gelegt sind eine reale Investitionsreihe für Bauten im Bereich Erziehung und Unterricht und eine durchschnittliche Nutzungsdauer von 50 Jahren. Während sich Abschreibungen und Abgänge zunächst nahezu parallel entwickeln, erhöhen sich die Abschreibungen bei starkem Investitionsanstieg sehr viel schneller als die Abgänge. Erst wenn die Anlagegüter der „Investitions-Hochphasen“ endgültig aus dem Anlagenbestand ausscheiden, sind die Abgänge höher als die Abschreibungen, falls die Investitionen nicht rechtzeitig wieder steigen.

Grafik 3

Abhängigkeit der Abschreibungen und Abgänge vom Investitionsverlauf¹ bei einer Nutzungsdauer von 50 Jahren
Mrd. EUR



1 Reale Investitionen für Bauten im Bereich Erziehung und Unterricht.

2020 - 01-0216


6

Fazit

Die Vermögensrechnung hat die Aufgabe, sowohl die Höhe als auch die Zusammensetzung und die Veränderung von Vermögensbeständen zu ermitteln. Sie stellt wichtige Bestandsgrößen für die gesamtwirtschaftlichen und sektoralen Vermögensbilanzen und nach Wirtschaftszweigen bereit. Darüber hinaus sind die mit der Anlagevermögensrechnung ermittelten Abschreibungen als Stromgröße für die Berechnung des Bruttoinlandsprodukts von Bedeutung. Die Abschreibungen der Nichtmarktproduzenten des Sektors Staat und der privaten Organisationen ohne Erwerbszweck beeinflussen direkt die Höhe des Bruttoinlandsprodukts. Die Abgrenzung der einzubeziehenden nichtfinanziellen Vermögensgüter sowie die Methodik sind vom ESVG 2010 vorgegeben. Das ESVG 2010 sieht sowohl die Berechnung und Lieferung des Anlagevermögens nach dem Brutto- und dem Nettokzept als auch der Abschreibungen, nicht aber der Abgänge und weiterer Stromgrößen vor.

Nettoanlagevermögen und Abschreibungen standen bisher stärker im Interesse der wissenschaftlichen Diskussion zu Produktionspotenzial und zur notwendigen Höhe von Investitionen vor allem im öffentlichen

Bereich. Zur Unterstützung der neueren Diskussion, welche die Bruttobetrachtung (wieder) in die Analyse einbezieht (ausgelöst durch Grömling und andere, 2019), wurden die unterschiedlichen analytischen Aussagen von Netto- und Bruttokzept der Anlagevermögensrechnung näher erläutert. Zusätzlich hat das Statistische Bundesamt die Abgänge (wieder) in das Veröffentlichungsprogramm aufgenommen (Statistisches Bundesamt, 2020, Tabelle 3.1.5).

Ergänzend wurde in diesem Aufsatz auch der Übergang vom Jahresanfangsbestand zum Jahresendbestand sowohl für das Brutto- als auch für das Nettoanlagevermögen vollständig mit allen Stromgrößen nachgewiesen. Damit wird die Aufgabe der Vermögensrechnung, auch die Veränderung von Vermögensbeständen aufzuzeigen, für das Anlagevermögen vollständig erfüllt. Mit diesem Beitrag wurden die bisherigen Lücken in den Veröffentlichungen in Form der Abgänge sowie der sonstigen realen Vermögensänderungen und der Umbewertungsgewinne/-verluste geschlossen. Die Zusammensetzung der Vermögensbestände und der zugehörigen Stromgrößen wurde anhand der Klassifikation der Vermögensgüter gezeigt. Sie ist bereits in den Standardveröffentlichungen der Volkswirtschaftlichen Gesamtrechnungen zur (Anlage-)Vermögensrechnung enthalten. 

LITERATURVERZEICHNIS

Adler, Walther/Gühler, Nadine/Oltmanns, Erich/Schmidt, Daniel/Schmidt, Pascal/Schulz, Ingeborg. [*Forschung und Entwicklung in den Volkswirtschaftlichen Gesamtrechnungen*](#). In: Wirtschaft und Statistik. Ausgabe 12/2014, Seite 703 ff.

Bardt, Hubertus/Dullien, Sebastian/Hüther, Michael/Rietzler, Katja. *Für eine solide Finanzpolitik: Investitionen ermöglichen!* Institut der Deutschen Wirtschaft. IW-Policy Paper Nr. 10/2019 – November 2019.

Europäische Union. *Europäisches System Volkswirtschaftlicher Gesamtrechnungen 1995*. Verordnung (EG) Nr. 2223/96 vom 25. Juni 1996 (Amtsblatt der EG Nr. L 310).

Eurostat. *Europäisches System Volkswirtschaftlicher Gesamtrechnungen ESVG 2010*. Luxemburg 2014.

Expertenkommission „Stärkung von Investitionen in Deutschland“. *Stärkung von Investitionen in Deutschland*. Bericht der Expertenkommission im Auftrag des Bundesministers für Wirtschaft und Energie, Sigmar Gabriel. Berlin 2015.

Feld, Lars P./Christofzik, Désirée I./Yeter, Mustafa. *Öffentliche Investitionen: Wie viel ist zu wenig?* In: Schweizer Monat – Die Autorenzeitschrift für Politik, Wirtschaft und Kultur. Ausgabe 1064 – März 2019, Seite 60 ff.

Grömling, Michael/Hüther, Michael/Jung, Markos. *Verzehrt Deutschland seinen staatlichen Kapitalstock?* In: Wirtschaftsdienst. 99. Jahrgang. Heft 1/2019, Seite 25 ff.

Gühler, Nadine/Schmalwasser, Oda. *Der Staat in der Vermögensrechnung*. In: Mink, Reimund/Voy, Klaus (Herausgeber). *Der Staat in den Volkswirtschaftlichen Gesamtrechnungen und in der Finanzstatistik*. Berliner Beiträge zu den Volkswirtschaftlichen Gesamtrechnungen – Band 4. Marburg 2019. Seite 113 ff.

Hauf, Stefan/Schäfer, Dieter. [*Revision der Volkswirtschaftlichen Gesamtrechnungen 1991 bis 2018*](#). In: WISTA Wirtschaft und Statistik. Ausgabe 5/2019, Seite 61 ff.

Räth, Norbert/Braakmann, Albert. [*Generalrevision der Volkswirtschaftlichen Gesamtrechnungen 2014 für den Zeitraum 1991 bis 2014*](#). In: Wirtschaft und Statistik. Ausgabe 9/2014, Seite 502 ff.

Schäfer, Dieter/Schmidt, Liane. [*Abschreibungen nach verschiedenen Bewertungs- und Berechnungsmethoden*](#). In: Wirtschaft und Statistik. Ausgabe 12/1983, Seite 919 ff.

Schmalwasser, Oda/Brede, Sascha. [*Grund und Boden als Bestandteil der volkswirtschaftlichen Vermögensbilanzen*](#). In: WISTA Wirtschaft und Statistik. Ausgabe 6/2015, Seite 43 ff.

Schmalwasser, Oda/Schidlowski, Michael. [*Kapitalstockrechnung in Deutschland*](#). In: Wirtschaft und Statistik. Ausgabe 11/2006, Seite 1107 ff.

LITERATURVERZEICHNIS

Schmalwasser, Oda/Weber, Nadine. [*Revision der Anlagevermögensrechnung für den Zeitraum 1991 bis 2011*](#). In: Wirtschaft und Statistik. Ausgabe 11/2012, Seite 933 ff.

Statistisches Bundesamt. *Volkswirtschaftliche Gesamtrechnungen: Inlandsprodukt und Nationaleinkommen nach ESG 2010 - Methoden und Grundlagen*. Fachserie 18 Reihe S. 30. 2016.

Statistisches Bundesamt. *Sektorale und gesamtwirtschaftliche Vermögensbilanzen 1999 - 2018*. 2019.

Statistisches Bundesamt. *Volkswirtschaftliche Gesamtrechnungen: Inlandsproduktberechnung – Detaillierte Jahresergebnisse 2019*. Fachserie 18 Reihe 1.4. 2020.

HAUSHALTE IN DER BERICHTERSTATTUNG DES MIKROZENSUS AB 2020

Tim Hochgürtel, Julia Weinmann

📌 **Schlüsselwörter:** Mikrozensus 2020 – Haushalt – Haupteinkommensperson – Bevölkerung in Privathaushalten – Hauptwohnsitz

ZUSAMMENFASSUNG

Bislang wurden die Haushaltsbefragungen Mikrozensus (in den die EU-Arbeitskräfteerhebung bereits integriert ist), die Gemeinschaftsstatistik über Einkommen und Lebensbedingungen und die Erhebung über die private Nutzung von Informations- und Kommunikationstechnologien separat durchgeführt. Die Zusammenführung zu einer gemeinsamen Erhebung ab dem Berichtsjahr 2020 erfordert Harmonisierungen, um eine kohärente Berichterstattung zu gewährleisten. Der Aufsatz beschreibt, welche Veränderungen sich für die Definitionen und Abgrenzungen zur Darstellung von Haushalten ergeben.

📌 **Keywords:** *microcensus 2020 – household – main income earner – population in households – main residence*

ABSTRACT

The microcensus (into which the EU labour force survey was integrated), the Community statistics on income and living conditions, and the survey on the use of information and communication technologies in households and by individuals have all been carried out separately. Merging these three household surveys into one joint survey from reporting year 2020 requires harmonization to ensure coherent reporting. This paper describes the changes to be made to the definitions relating to households.



Tim Hochgürtel

studierte Soziologie in Mainz und ist seit 2008 im Statistischen Bundesamt tätig, seit 2016 als Referent im Referat „Bevölkerungsstatistische Auswertungen und Analysen aus dem Mikrozensus“. Schwerpunkte seiner Arbeit sind Analysen sowie Publikationen zu privaten Haushalten und Familien.

Julia Weinmann

ist Soziologin und seit über zehn Jahren im Statistischen Bundesamt im Mikrozensus tätig, derzeit als Referentin im Referat „Auswertung und Analyse der Haushaltserhebungen, Wohnen“. Davor war sie über viele Jahre mit der Veröffentlichung und den Analysen von Mikrozensusdaten zum Thema Haushalte und Familien betraut.

1

Einleitung

Seit 1957 wird im früheren Bundesgebiet der Mikrozensus als Mehrthemenbefragung in privaten Haushalten durchgeführt. In den neuen Ländern wird der Mikrozensus seit 1991 erhoben. Bei der jährlich durchgeführten Erhebung wird 1 % der Bevölkerung zu Themen wie Arbeitsmarktpartizipation, Bildungsbeteiligung oder Haushaltszusammensetzung befragt. Für den Mikrozensus gilt Auskunftspflicht (Emmerling/Riede, 1997).

Für die Stichprobe der Mikrozensususerhebung eines Jahres wird 1 % der Bevölkerung Deutschlands ausgewählt. Hierzu werden zufällig Auswahlbezirke gezogen, in denen die Befragung durchgeführt wird. Bei einem Auswahlbezirk handelt es sich um eine kleine Anzahl benachbarter Wohnungen oder um Teile einer Gemeinschaftsunterkunft. Die benachbarten Wohnungen eines Auswahlbezirks können im gleichen Haus liegen oder sich auf mehrere Häuser verteilen (Herberger, 1957; Bihler/Zimmermann, 2016).

In einer bewohnten Wohnung sind entweder ein Haushalt oder eventuell auch mehrere Haushalte vorhanden. Die grundlegende Definition eines privaten Haushalts hat sich seit Beginn des Mikrozensus nicht verändert. So galt bereits zur Einführung des Mikrozensus (Schubnell, 1959a) und gilt laut Mikrozensusgesetz noch immer: „Einen Haushalt bilden alle Personen, die gemeinsam wohnen und wirtschaften. Wer allein wohnt oder allein wirtschaftet, bildet einen eigenen Haushalt.“ (§3 Absatz 2 Mikrozensusgesetz) Dazu zählen Einpersonenhaushalte ebenso wie Mehrpersonenhaushalte. Gemeinschaftsunterkünfte, beispielsweise Alten- und Pflegeheime oder Klöster, gehören nicht zu den Privathaushalten.

In der Regel bilden die Bewohnerinnen und Bewohner einer Wohnung einen Haushalt. Lebt aber beispielsweise eine Wohngemeinschaft in einer Wohnung, so kann es sich dabei um mehrere Haushalte handeln, sofern das Kriterium des „gemeinsamen Wirtschaftens“ nicht erfüllt ist. Welche Personen im Einzelnen dem Haushalt angehören und anschließend auch auskunftspflichtig zu diesem Haushalt befragt werden, legt der Haushalt selbst fest. Auch Personen, die nur vorübergehend nicht im Haushalt leben, weil sie zum Beispiel

zum Berichtszeitpunkt im Krankenhaus waren, werden zu den Mitgliedern eines Haushalts gezählt.

Bei der Mikrozensusbefragung wird unter anderem erfasst, ob die Personen mit Haupt- oder mit Nebenwohnsitz in der befragten Wohnung leben. Dabei ist die Zugehörigkeit einer Person nicht an den melderechtlichen Status gebunden, sondern soll im Einklang mit dem intuitiven, subjektiven Verständnis stehen. Der Hauptwohnsitz bestimmt sich dadurch, wo eine Person ihren Lebensmittelpunkt hat. Die Wahl des Melderechtsstatus kann auch andere, zum Beispiel finanzielle Gründe haben.

Die Methodik des Mikrozensus wurde in der Vergangenheit immer wieder neuen Anforderungen angepasst (Emmerling/Riede, 1997; Lotze/Breiholz, 2002). So wurde bereits 1968 die Arbeitskräfteerhebung der Europäischen Union (Labour Force Survey, LFS) in den Mikrozensus integriert. Die europaweit harmonisierte Arbeitskräfteerhebung liefert Daten zur Erwerbssituation, die über alle Länder der Europäischen Union (EU) vergleichbar sind.

Mit dem Erhebungsjahr 2020 erfolgte erneut eine solche Anpassung (Hundenborn/Enderer, 2020; Hochgürtel, 2017). Hierbei bleiben grundsätzliche Strukturelemente des Mikrozensus erhalten. Beispielsweise werden nach wie vor für den Mikrozensus 1 % der Bevölkerung aus zufällig gezogenen Auswahlbezirken mit Auskunftspflicht befragt. Auch das grundlegende Verständnis eines Haushalts bleibt unverändert.

Die Anforderungen an haushaltsstatistische Befragungen sind gestiegen. Daher werden die Erhebungen „LEBEN IN EUROPA“¹ (SILC) und ab 2021 die Erhebung über die private Nutzung von Informations- und Kommunikationstechnologien (IKT) in den Mikrozensus integriert. Hieraus resultiert ein Harmonisierungsbedarf der Berichterstattung. Der vorliegende Beitrag stellt die zentralen Konzepte der Berichterstattung zu Haushalten ab 2020 dar.

Das folgende Kapitel 2 beschreibt, wie der Mikrozensus durch die Integration der bisher separat durchgeführten Erhebungen SILC und IKT diesen neuen Anforderungen begegnet. Diese Integration macht eine Harmonisierung der Berichterstattung über die bisher getrennten Erhe-

1 Das ist die deutsche Bezeichnung der Gemeinschaftsstatistik über Einkommen und Lebensbedingungen (Statistics on Income and Living Conditions, SILC).

bungen hinweg erforderlich. Die Folgekapitel erläutern die grundlegenden Konzepte einer solchen harmonisierten Berichterstattung: die Haushalte am Hauptwohnsitz, das Konzept der Haupteinkommensperson sowie die Abgrenzung der Bevölkerung in Haushalten am Hauptwohnsitz. Der Beitrag schließt in Kapitel 6 mit einem Ausblick.

2

Der integrierte Mikrozensus ab 2020

Die Veröffentlichung der europäischen Rahmenvereinbarung IESS² im Oktober 2019 hat auch auf europäischer Ebene ein neues rechtliches Fundament geschaffen. Sie enthält unter anderem die Rahmenbedingungen der europäisch harmonisierten Haushaltserhebungen.

Mit dieser Rahmenvereinbarung steigen die Anforderungen an die amtliche Statistik hinsichtlich der Erhebungen zu privaten Haushalten. So erhöhen sich etwa die Genauigkeitsanforderungen an SILC. Darüber hinaus wird durch die Rahmenvereinbarung eine unterjährige Wiederholungsbefragung für die Arbeitskräfteerhebung etabliert.


Um diesen neuen Anforderungen zu begegnen, wurde mit dem Inkrafttreten des neuen Mikrozensusgesetzes das System der Haushaltsstatistiken in Deutschland neu konzipiert. Die bisher separat durchgeführte Erhebung über Einkommen und Lebensbedingungen (SILC) wurde ebenso wie die Befragung zu Informations- und Kommunikationstechnologien in den privaten Haushalten (IKT) in den Mikrozensus integriert. Somit sind nun neben der Arbeitskräfteerhebung zwei weitere europäische Statistiken Teil des Mikrozensus (Hundenborn/Enderer, 2019).

Im Mikrozensus gab es bereits vor 2020 Parallelen zu den beiden Haushaltsstatistiken SILC und IKT. Beide werden europaweit harmonisiert erhoben und sind mit Datenlieferungspflichten an die EU verbunden. Dies gilt für weite Teile der Erhebungsmerkmale dieser Haushaltsstatistiken, aber auch für die Befragungsinstrumente, das Erhebungsmanagement und die Auswertungsverfahren. Ziel der Integration von SILC (seit 2020) und IKT

(ab 2021) in den Mikrozensus ist, durch diese Gemeinsamkeiten Synergieeffekte zu nutzen und Inkohärenzen sowie Redundanzen zwischen den Befragungen zu vermeiden.

Der Mikrozensus fragt ab dem Erhebungsjahr 2020 weite Teile des Merkmalskranzes, insbesondere soziodemografische Merkmale sowie die Haushaltszusammensetzung, gemeinsam in einem Kernprogramm ab. Alle Auskunftgebenden aus privaten Haushalten der 1%-Stichprobe des Mikrozensus beantworten die Fragen des Kernprogramms.

In Gemeinschaftsunterkünften (zum Beispiel Alten- oder Pflegeheimen, Klöstern) lebende Personen werden bereits seit dem Mikrozensus 2017 standardmäßig nur noch zu einer kleinen Teilmenge des Merkmalskranzes des Mikrozensus befragt. Somit liegen für sie nur ausgewählte Informationen aus der Befragung vor. Sie erhalten beispielsweise keine Fragen mehr nach ihrem höchsten Bildungsstand, ihrer Erwerbsbeteiligung oder ihrem Migrationshintergrund. Diese Maßnahme soll einerseits die Belastung durch die Befragung für die Respondenten reduzieren und andererseits die Qualität der Daten zu Personen in Gemeinschaftsunterkünften verbessern.

Den Mitgliedern von privaten Haushalten werden im Rahmen des Kernprogramms unter anderem Fragen zur Anzahl der in ihrem Haushalt lebenden Personen gestellt. Hierzu zeigt  Grafik 1 die entsprechende Frage aus dem Selbstausfüllerfragebogen ab dem Mikrozensus 2020. Zeitlich gesehen bezieht sich diese auf den Donnerstag in der Berichtswoche. In den bis einschließlich 2019 getrennt durchgeführten Erhebungen Mikrozensus, SILC und IKT waren die Referenzzeitpunkte für diese Frage unterschiedlich: Im Mikrozensus bezog sie sich auf den Mittwoch der Berichtswoche, bei SILC lautete die Frage zur Anzahl der Personen im Haushalt bis 2019: „Wie viele Personen gehören derzeit zum Haushalt?“. In der IKT-Erhebung lautete die Frage „Wie viele Personen gehörten am 31.12. (des Vorjahres) zu Ihrem Haushalt?“ Die Nennung eines konkreten Stichtags legte einen anderen Referenzzeitpunkt fest als bei den beiden anderen Erhebungen. Diese bisherigen Unterschiede schafft eine harmonisierte Abfrage in einem gemeinsamen Kernprogramm ab.

² Integrated European Social Statistics (IESS): Rahmenverordnung für die Erstellung europäischer Statistiken für Personen und Haushalte.

Grafik 1

Frage nach der Anzahl der Personen eines Haushalts

2 Wie viele Personen haben am Donnerstag in der Berichtswoche insgesamt in Ihrem Haushalt gelebt?

Zeitweise abwesende Personen
gehören zum Haushalt, wenn sie z. B. aus beruflichen oder gesundheitlichen Gründen abwesend sind, aber normalerweise hier wohnen.

Keine Haushaltsmitglieder
sind Untermieter/-innen, Personen, die zu Besuch anwesend sind, und Hausangestellte.

Anzahl der Personen in Ihrem Haushalt
(Sie selbst mit einbezogen)

Hinweis!
Die Berichtswoche finden Sie auf dem Deckblatt.

Auszug aus dem Selbstausfüllerbogen ab dem Mikrozensus 2020

2020 - 01 - 0231

Die Angabe, welche Personen in einer für die Stichprobe ausgewählten Wohnung einen gemeinsamen Haushalt bilden, kommt direkt von den Befragten. Hierbei entscheiden die Befragten etwa, welche zum Zeitpunkt der Befragung abwesenden Personen dem Haushalt noch zugerechnet werden. Welche Personen als zeitweise abwesend galten, unterschied sich bis einschließlich 2019 in den einzelnen Erhebungen teilweise durch den Detaillierungsgrad der Erläuterung, teilweise gab es auch inhaltliche Abweichungen. Ab dem Mikrozensus 2020 erhalten die Befragten über alle Erhebungsteile hinweg die gleiche Frage und dazu die gleiche Erläuterung (siehe die mit einem „i“ gekennzeichnete Erläuterung in Grafik 1). Dieses Vorgehen minimiert auch diesbezüglich bislang existierende Unterschiede.

Mit dem Erhebungsjahr 2020 wird im Mikrozensus auch das Konzept der festen Berichtswoche eingeführt. Sofern nicht anders angegeben, beziehen sich die Fragen auf eine feste Kalenderwoche des Jahres: die Berichtswoche. Jedem Auswahlbezirk ist eine solche Berichtswoche zugeordnet. Das Konzept der festen Berichtswoche löst das bis einschließlich 2019 im Mikrozensus praktizierte Konzept der gleitenden Berichtswoche ab, wonach Haushalten Fragen zu jener Woche gestellt wurden, welche der Befragung vorausging.

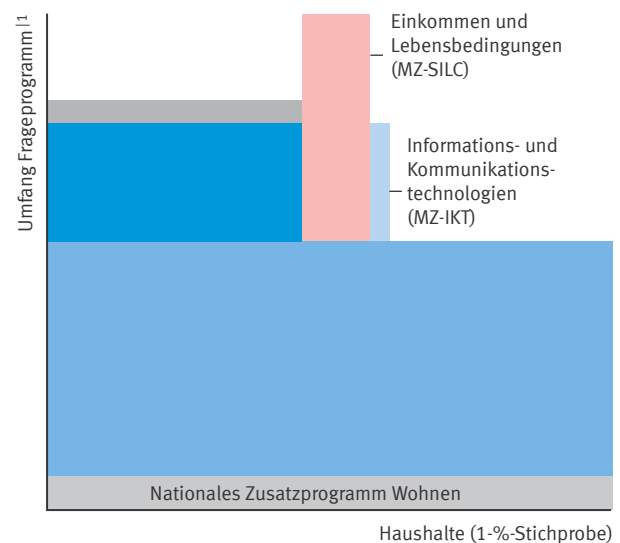
➤ Grafik 2 zeigt, wie und in welchem Umfang weitere Erhebungsteile das Kernprogramm ergänzen. Einige Befragte der 1%-Stichprobe tragen neben dem Kernprogramm zu einem weiteren Erhebungsteil bei. Diese Personen beantworten noch weitere Fragen zu maximal einem der weiteren Erhebungsteile.

Die größte Unterstichprobe bildet der Erhebungsteil zum Arbeitsmarkt (MZ-LFS) mit einem Unterauswahlsatz von bis zu 45 % des Gesamtauswahlsatzes. Dieser Erhebungsteil vertieft Angaben zur Arbeitsmarktbeteiligung aus dem Kernprogramm um weitere Aspekte. Befragte im MZ-LFS sind alle Personen ab 15 Jahren aus privaten Haushalten in den für eine Befragung zu MZ-LFS ausgewählten Auswahlbezirken.

Der Erhebungsteil zu Einkommen und Lebensbedingungen (MZ-SILC) umfasst bis zu 12 % der in der 1%-Stichprobe ausgewählten Haushalte. Sie werden sehr detailliert zu den Themen Einkommen, materielle Deprivation,

Grafik 2

Aufbau des integrierten Mikrozensus ab 2020



1 Umfang angenähert auf Basis des Erhebungsjahrs 2020.

Quelle: Hundenborn/Enderer (2019), Seite 11

2020 - 01 - 0232

Wohnkosten, Kinderbetreuung, Gesundheit und Lebensbedingungen befragt. Für diesen Erhebungsteil werden ausschließlich Personen ab 16 Jahren mit Hauptwohnsitz in der für MZ-SILC ausgewählten Wohnung herangezogen.

Die Erhebung zu Informations- und Kommunikationstechnologien in privaten Haushalten (MZ-IKT) bildet mit bis zu 3,5 % aller Befragten/Haushalte die kleinste Unterstichprobe. Sie umfasst – auf freiwilliger Basis – Personen zwischen 16 und 74 Jahren am Hauptwohnsitz aus für MZ-IKT ausgewählten Auswahlbezirken. Die Schwerpunkte liegen auf Fragen zu Art, Häufigkeit und ausgewählten Zwecken der Internetnutzung³ sowie zu weiteren, aktuellen IKT-Themenbereichen⁴. Zusätzlich werden Informationen darüber erhoben, welche Bedenken und Hindernisse Menschen von der Nutzung von Informations- und Kommunikationstechnologien abhalten.

Die nationalen Zusatzprogramme zu Themen wie Wohnen, Gesundheit, Behinderung, Pendlerverhalten oder Krankenversicherungsschutz erweitern das Merkmalspektrum und werden teilweise jährlich, teilweise im Abstand von vier Jahren rotierend abgefragt.

Auch die Berichterstattung auf Basis des Mikrozensus wird ab dem Jahr 2020 harmonisiert. Um eine kohärente Berichterstattung über die sozialen und wirtschaftlichen Verhältnisse in Deutschland auf Basis des Mikrozensus zu realisieren, werden die bisher verwendeten Konzepte der Berichterstattung vereinheitlicht. Hiervon ist auch die Berichterstattung zu Haushalten auf Basis des Mikrozensus betroffen.

3

Haushalte am Hauptwohnsitz

Die skizzierte abfrageseitige Harmonisierung der Haushaltszusammensetzung bildet ab dem Mikrozensus 2020 die Grundlage, auch auswertungsseitig die Haushalte über alle integrierten Statistiken hinweg entsprechend einheitlich darzustellen. Aufgrund von Datenlieferungspflichten an die EU orientiert sich die neue

Definition des Haushalts ab dem Mikrozensus 2020 an der bislang für SILC europaweit gültigen Haushaltsdefinition des Statistischen Amtes der Europäischen Union (Eurostat).

Die grundlegende Haushaltsdefinition bleibt auch in der Berichterstattung auf Basis des harmonisierten Mikrozensus unverändert: Einen Haushalt bilden Personen, die gemeinsam wohnen und wirtschaften. Jedoch verschiebt die Berichterstattung zu Haushalten ihren Fokus. Bis zum Berichtsjahr 2019 ist sie am Informationsbedarf zur Infrastruktur ausgerichtet, weil beispielsweise für städtebauliche Maßnahmen Informationen zu regionalen Besonderheiten von Bevölkerung und Haushaltezusammensetzung benötigt werden. Entsprechend werden für die Berichterstattung sowohl Haushalte am Hauptwohnsitz als auch am Nebenwohnsitz berücksichtigt. Da Haushalte unabhängig vom Status des Haupt- oder Nebenwohnsitzes die lokale Infrastruktur in Anspruch nehmen, ist die bloße Existenz des Haushalts Anlass für die Berücksichtigung in der Zählung.

Mit dem Berichtsjahr 2020 verschiebt sich der vorrangige Blick hin zu einer ökonomisch motivierten Perspektive. Die Berichterstattung stützt sich hierbei auf Haushalte am Hauptwohnsitz. Nebenwohnsitzhaushalte bleiben bei dieser Betrachtung unberücksichtigt, da deren Bewohnerinnen und Bewohner nicht unabhängig von ihrer Zugehörigkeit zum Hauptwohnsitzhaushalt wirtschaften. Dieses Vorgehen entspricht der gegenwärtigen Praxis der Berichterstattung zur Einkommensverteilung von Haushalten, die Haushalte mit ausschließlich Personen am Nebenwohnsitz nicht in die Analysen aufnimmt.

Die Neuerungen in der Berichterstattung zu Haushalten des Mikrozensus ab 2020 basieren damit auf der Unterscheidung von sogenannten Hauptwohnsitzhaushalten und Nebenwohnsitzhaushalten. Diese Unterscheidung war auf Grundlage des Mikrozensus bereits bisher möglich. Neu ist, dass ab dem Mikrozensus 2020 der Fokus bei der Veröffentlichung der Ergebnisse auf den Hauptwohnsitzhaushalten liegt. Aus europäischer Sicht ist ein Haushalt primär eine sozioökonomische Einheit, bei welcher Doppelzählungen zum Beispiel des Einkommens einer Person unerwünscht sind.

Nebenwohnsitzhaushalte werden nicht grundsätzlich außer Acht gelassen, sie werden ab dem Mikrozensus 2020 jedoch nicht mehr standardmäßig in der Gesamtzahl der Haushalte in Deutschland ausgewiesen.

3 Zum Beispiel Nutzung sozialer Netzwerke, Blogs und Foren, Suche nach Informationen, E-Commerce, E-Government, E-Learning.

4 Wie etwa E-Skills, Internetsicherheit, Schutz der Privatsphäre im Internet oder Smart Home.

Ziel ist, eine kohärente Berichterstattung über alle Erhebungsteile des Mikrozensus ab 2020 zu gewährleisten. Die Befragungen für MZ-SILC und MZ-IKT erfolgen nur in solchen Haushalten, in denen mindestens eine Person ab 16 Jahre ihren Hauptwohnsitz hat. Entsprechend sind Haushalte ausgeschlossen, in denen ausschließlich Personen mit Nebenwohnsitz leben. Für eine einheitliche Berichterstattung über alle Erhebungsteile hinweg bildet eine identische Abgrenzung der Haushalte die Grundlage der Haushalteberichterstattung in den jeweiligen Erhebungsteilen. Hierzu werden die Haushalte am Hauptwohnsitz verwendet.

Hauptwohnsitzhaushalte werden hierbei wie folgt operationalisiert: Sie sind die Teilmenge der Haushalte, in welchen mindestens eine Person dieses Haushalts dort mit Hauptwohnsitz lebt und 16 Jahre oder älter ist. Ist die einzige Person mit Hauptwohnsitz in einem Haushalt jünger als 16 Jahre, dann wird dieser Haushalt zu den Nebenwohnsitzhaushalten gezählt. Solche Haushalte sind jedoch so selten, dass diese Population selbst mit dem Mikrozensus nicht nachweisbar ist.

➤ Grafik 3 zeigt, wie sich die Menge der Hauptwohnsitzhaushalte untergliedert: in Haushalte, in denen alle Haushaltsmitglieder am Hauptwohnsitz leben, und in Haushalte, in denen Personen sowohl am Haupt- als auch am Nebenwohnsitz wohnen. In Nebenwohnsitzhaushalten haben in der Regel alle Haushaltsmitglieder ihren Nebenwohnsitz. Einzige Ausnahme bildet die oben skizzierte Altersregelung. Sie wird aus Konsistenzgründen zu der Unterstichprobe MZ-SILC und der Datenlieferungspflicht aus dieser an Eurostat beibehalten. Danach

werden die personenbezogenen Fragen zu Einkommen und Lebensbedingungen ausschließlich den Personen am Hauptwohnsitz gestellt, die 16 Jahre und älter sind.

4

Haupteinkommensperson

Traditionell erfolgt in der Berichterstattung zu Haushalten eine Darstellung unmittelbar nach Eigenschaften des Haushalts (zum Beispiel Haushaltsgröße, Haushaltsnettoeinkommen, Anzahl der Erwerbstätigen im Haushalt). Daneben werden auch Individualeigenschaften auf Ebene der Haushalte ausgewertet, beispielsweise der höchste schulische oder berufliche Bildungsabschluss (Schwenk, 1959, hier: Seite 244).

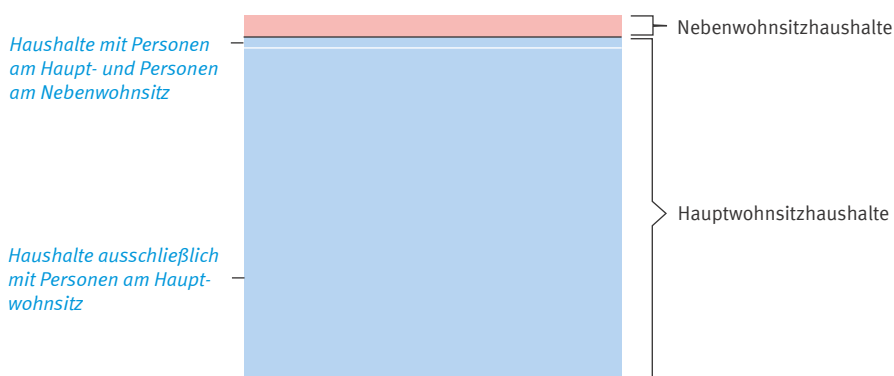
Bis 1984 erfolgte eine Darstellung der Haushalte des Mikrozensus nach dem Haushaltsvorstand⁵ zur Haushaltsrepräsentanz. Danach wurde der Haushaltsvorstand durch die Haushaltsbezugsperson⁶ zur Haushaltsrepräsentanz abgelöst. Seit 2005 erfolgt eine Darstellung von Individualeigenschaften einer Haushaltsrepräsentanz einer Haupteinkommensbezieherin beziehungsweise eines Haupteinkommensbeziehers.

Bei der Haupteinkommensbezieherin beziehungsweise dem Haupteinkommensbezieher handelt es sich im

- 5 In der Befragung wurde eine Person des Haushaltes von den Haushaltsmitgliedern selbst als Haushaltsvorstand bestimmt (Schubnell, 1959b, hier: Seite 64).
- 6 Die Bezugsperson ist die Person, welche als erste Person im Fragebogen eingetragen ist (Neubauer, 1993, hier: Seite 222).

Grafik 3

Haupt- und Nebenwohnsitzhaushalte



2020 - 01 - 0233

Mikrozensus 2005 bis 2019 um die Person ab 15 Jahren mit dem höchsten monatlichen Nettoeinkommen im Haushalt. Sofern mehrere Haushaltsmitglieder über das gleiche persönliche monatliche Nettoeinkommen verfügen, entscheidet die Reihenfolge, in der die Personen im Fragebogen eingetragen sind. Haupteinkommensbezieherin beziehungsweise Haupteinkommensbezieher ist dann – aus dem Kreis aller Personen mit höchster persönlicher Nettoeinkommensklasse im Haushalt – das Haushaltsmitglied mit der niedrigsten Personennummer. Hat im Ausnahmefall kein Haushaltsmitglied Angaben zum persönlichen monatlichen Nettoeinkommen gemacht, ist die Haushaltsbezugsperson gleichzeitig Haupteinkommensbezieherin beziehungsweise Haupteinkommensbezieher des Haushalts.

Ab dem Mikrozensus 2020 erhält die Haupteinkommensbezieherin beziehungsweise der Haupteinkommensbezieher über alle Erhebungsteile hinweg einheitlich die geschlechtsneutrale Bezeichnung „Haupteinkommensperson“. Eine Veränderung der Operationalisierung der Haushaltsrepräsentantin beziehungsweise des Haushaltsrepräsentanten findet sich hinsichtlich des Mindestalters. Die Altersgrenze, ab wann eine Person Haupteinkommensperson eines Haushalts sein kann, wurde von 15 auf 18 Jahre angehoben, da mit Vollendung des 18. Lebensjahres unbeschränkte Geschäftsfähigkeit⁷ erreicht wird. Haupteinkommenspersonen sind somit ausschließlich volljährige Personen, es sei denn, alle im Haushalt lebenden Personen sind minderjährig. In diesen sehr seltenen Fällen kann auch eine minderjährige Person Haupteinkommensperson eines Haushalts sein.

5

Bevölkerung in Hauptwohnsitzhaushalten

Neben der Darstellung von Haushalten als Einheit erfolgt die Berichterstattung auch über die Bevölkerung in privaten Haushalten. Dabei korrespondiert die dargestellte Bevölkerung mit den Haushalten, über welche berichtet wird.

⁷ Im Bürgerlichen Gesetzbuch (BGB) wird nicht konkret der Eintritt der vollen Geschäftsfähigkeit geregelt. Es stuft grundsätzlich alle Menschen als voll geschäftsfähig ein, allerdings mit Ausnahmen: So regelt § 106 BGB die beschränkte Geschäftsfähigkeit Minderjähriger.

Bis 2019 umfasst die Berichterstattung die Haushalte am Haupt- und Nebenwohnsitz sowie mit der Bevölkerung in Privathaushalten am Haupt- und Nebenwohnsitz jene Personen, die in diesen Haushalten leben. Ausgenommen ist der ausschließlich in Gemeinschaftsunterkünften lebende Teil der Bevölkerung. Personen mit mehreren Wohnsitzen in privaten Haushalten können hingegen mehrfach zur Bevölkerung in privaten Haushalten am Haupt- und Nebenwohnsitz beitragen.

Analog zur Berichterstattung zu Haushalten am Hauptwohnsitz richtet sich bei Analysen und Veröffentlichungen auf Personenebene der Fokus ab dem Mikrozensus 2020 standardmäßig auf die Bevölkerung, die in Hauptwohnsitzhaushalten wohnt. Sie erhält die Bezeichnung „Bevölkerung in Hauptwohnsitzhaushalten“. Ihr Kennzeichen ist, dass der Großteil der in den Hauptwohnsitzhaushalten lebenden Personen dort seinen Hauptwohnsitz hat. In geringem Umfang leben in diesen Haushalten auch Personen mit Nebenwohnsitz. Da eine Person in mehreren Privathaushalten wohnberechtigt sein kann, sind somit Mehrfachzählungen möglich. Die „Bevölkerung in Hauptwohnsitzhaushalten“ umfasst wiederum nicht den Teil der Bevölkerung, der ausschließlich in Gemeinschaftsunterkünften lebt.

➤ Grafik 4 fasst zusammen, wie die Bevölkerung in Deutschland anhand der Kriterien Privathaushalt/Gemeinschaftsunterkunft und Haupt-/Nebenwohnsitz unterteilt werden kann.

Grafik 4


Bevölkerung in Deutschland nach Wohnsitz und Wohnort

Bevölkerung in Hauptwohnsitzhaushalten	Bevölkerung in Nebenwohnsitzhaushalten
Bevölkerung am Hauptwohnsitz in Gemeinschaftsunterkünften	Bevölkerung am Nebenwohnsitz in Gemeinschaftsunterkünften

2020 - 01 - 0234

Allerdings ist auf Folgendes hinzuweisen: Die Summe der Personen, die in einem Hauptwohnsitzhaushalt leben, und der in den Standardveröffentlichungen unberücksichtigten Bevölkerung am Hauptwohnsitz in Gemeinschaftsunterkünften wird nicht die aus der laufenden Bevölkerungsfortschreibung publizierte Bevöl-

kerungszahl ergeben. Dies ist unter anderem darauf zurückzuführen, dass zur „Bevölkerung in Hauptwohnsitzhaushalten“ Menschen mit mehreren Wohnsitzen mehrfach beitragen. Die Bevölkerungsfortschreibung zählt Menschen mit mehreren Wohnsitzen nur einmal, allerdings unabhängig davon, ob sie in einem Privathaushalt oder in einer Gemeinschaftsunterkunft leben.

gungsergebnisse, die zum Zeitpunkt der Aufbereitung der Erstveröffentlichung noch nicht vorliegen, werden für die spätere Veröffentlichung von Endergebnissen berücksichtigt (Hundenborn/Enderer, 2019). 

6

Fazit

Mit der europäischen Rahmenvereinbarung IESS sind neue Anforderungen für die amtlichen Haushaltsstatistiken verbunden. Um diesen gerecht zu werden, hat das Mikrozensusgesetz 2017 die Grundlage für eine Integration von SILC und IKT in den Mikrozensus geschaffen. Mit der Integration verbunden ist die Notwendigkeit einer harmonisierten Berichterstattung. Für die Berichterstattung zu Haushalten des Mikrozensus bedeutet dies, dass sich der primäre Fokus der Veröffentlichungstätigkeit von einem infrastrukturelleiten Informationsbedarf hin zu einer ökonomischen Betrachtungsweise verschiebt.

Die harmonisierten Konzepte „Haushalte am Hauptwohnsitz“, „Haupteinkommensperson“ sowie „Bevölkerung in Hauptwohnsitzhaushalten“ ermöglichen eine vergleichende und kohärente Betrachtung über alle Erhebungsteile sowie über den gesamten Mikrozensus hinweg.

Eine Darstellung des Mikrozensus nach den bisherigen nationalen Abgrenzungen⁸ ist auch auf Basis des Mikrozensus ab 2020 weiterhin möglich. Diese Abgrenzungen stellen aber nicht mehr die zentrale Perspektive der Veröffentlichung dar.

Eine weitere Neuerung in Bezug auf die Veröffentlichungstätigkeit zu Haushalten auf Basis des Mikrozensus ergibt sich durch die Publikation von Erst- und Endergebnissen. Künftig werden bereits drei bis vier Monate nach Ende des Berichtsjahres erste Ergebnisse bereitgestellt, rund vier Monate früher als derzeit. Befra-

⁸ Das sind beispielsweise die Haushalte am Haupt- und Nebenwohnsitz, die Haupteinkommensbezieherin beziehungsweise der Haupteinkommensbezieher sowie die Bevölkerung am Haupt- und Nebenwohnsitz.

LITERATURVERZEICHNIS

Bihler, Wolf/Zimmermann, Daniel. [*Die neue Mikrozensusstichprobe ab 2016*](#). In: WISTA Wirtschaft und Statistik. Ausgabe 6/2016, Seite 20 ff.

Lotze, Sabine/Breiholz, Holger. [*Zum neuen Erhebungsdesign des Mikrozensus. Teil 1*](#). In: Wirtschaft und Statistik. Ausgabe 5/2002, Seite 359 ff.

Emmerling, Dieter/Riede, Thomas. [*40 Jahre Mikrozensus*](#). In: Wirtschaft und Statistik. Ausgabe 3/1997, Seite 160 ff.

Herberger, Lothar. [*Der Mikrozensus als neues Instrument zur Erfassung sozial-ökonomischer Tatbestände*](#). In: Wirtschaft und Statistik. Ausgabe 4/1957, Seite 209 ff.

Hochgürtel, Tim. [*Das künftige System der amtlichen Haushaltsstatistiken*](#). In: Wirtschaft und Statistik. Ausgabe 7/2013, Seite 457 ff.

Hundenborn, Janina/Enderer, Jörg. [*Die Neuregelung des Mikrozensus ab 2020*](#). In: WISTA Wirtschaft und Statistik. Ausgabe 6/2019, Seite 9 ff.

Neubauer, Werner. *Idealtypus und Adäquation des privaten Haushalts in der deutschen amtlichen Statistik: Wohngemeinschaft? Budgetgemeinschaft? Leistungsgemeinschaft? Produktionsgemeinschaft?* In: Gräbe, Sylvia (Herausgeberin). *Der Private Haushalt im wissenschaftlichen Diskurs*. Frankfurt am Main 1993, Seite 193 ff.

Schubnell, Hermann. *Haushalt und Familie II. Das neue Konzept der amtlichen Statistik zur Ermittlung und Analyse der Struktur von Haushalt und Familie*. In: Allgemeines Statistisches Archiv. Jahrgang 43. 1959a, Seite 221 ff.

Schubnell, Hermann. *Das Haushalts- und Familienkonzept der kommenden Volks- und Berufszählung*. In: Harmsen, Hans (Herausgeber). *Haushalt und Familie. Ermittlung und Analyse ihrer Struktur*. 1959b, Seite 54 ff.

Schwenk, Heinz. *Die haushalts- und familienstatistischen Konzepte der amtlichen Statistik im Ausland*. In: Allgemeines Statistisches Archiv. Jahrgang 43. 1959, Seite 238 ff.

RECHTSGRUNDLAGEN

Gesetz zur Durchführung einer Repräsentativstatistik über die Bevölkerung und die Arbeitsmarktbeteiligung sowie die Wohnsituation der Haushalte (Mikrozensusgesetz – MZG) vom 7. Dezember 2016 (BGBl. I Seite 2826).

Verordnung (EU) 2019/17001700 des Europäischen Parlaments und des Rates vom 10. Oktober 2019 zur Schaffung eines gemeinsamen Rahmens für europäische Statistiken über Personen und Haushalte auf der Grundlage von Einzeldaten aus Stichprobenerhebungen, zur Änderung der Verordnungen (EG) Nr. 808/2004, (EG) Nr. 452/2008 und (EG) Nr. 1338/2008 des Europäischen Parlaments und des Rates sowie zur Aufhebung der Verordnung (EG) Nr. 1177/2003 des Europäischen Parlaments und des Rates und der Verordnung (EG) Nr. 577/98 des Rates (Amtsblatt der EU Nr. L 261, Seite I/1).

Jörg Feuerhake

ist Diplom-Volkswirt und Referent im Referat „Maschinelles Lernen und Imputationsverfahren“ des Statistischen Bundesamtes. Er befasst sich unter anderem mit der methodischen Weiterentwicklung und dem Einsatz maschineller Lernverfahren.

Kerstin Lange

hat Statistik studiert und ist seit 2017 im Statistischen Bundesamt im Bereich mathematisch-statistische Methoden tätig. Im Referat „Maschinelles Lernen und Imputationsverfahren“ ist sie schwerpunktmäßig für Imputationsverfahren zuständig.

Annelen Siegismund

hat Volkswirtschaftslehre in Mainz studiert und ist Referentin im Referat „Räumliche Bevölkerungsbewegungen, Gebietsgliederungen“ des Statistischen Bundesamtes. Sie beschäftigt sich insbesondere mit Datenqualitätsaspekten der Wanderungsstatistik.

Elsa Vigneau

hat Politikwissenschaft studiert und ist seit 2018 Wissenschaftliche Mitarbeiterin im Referat „Räumliche Bevölkerungsbewegungen, Gebietsgliederungen“ des Statistischen Bundesamtes. Zu ihren Aufgabenschwerpunkten gehören methodische Fragen der Wanderungsstatistik und europäische Datenlieferungen.

KODIERUNG DES GEBURTSSTAATS IN DER WANDERUNGSSTATISTIK

Ein Vergleich regelbasierter Signierung mit Verfahren
des maschinellen Lernens

Jörg Feuerhake, Kerstin Lange, Annelen Siegismund, Elsa Vigneau

➤ **Schlüsselwörter:** Geburtsstaat – Signierung – Wanderungsstatistik – Random Forest – Maschinelles Lernen

ZUSAMMENFASSUNG

Seit dem Berichtsjahr 2008 enthalten die Datenlieferungen zur deutschen amtlichen Wanderungsstatistik Angaben zum Geburtsstaat von zu-, fort- und umziehenden Personen. Wegen unzureichender Qualität wurden diese Daten bislang nur für Schätzungen nach Geburtsstaatsgruppen im Rahmen europäischer Lieferverpflichtungen genutzt. Künftig sollen auch Aussagen über die einzelnen Geburtsstaaten der Wandernden möglich sein. Daher wurden im Jahr 2019 verschiedene Methoden untersucht, um ein Verfahren zur automatisierten Plausibilisierung und Signierung des Merkmals zu entwickeln. Der Beitrag stellt zwei Ansätze vor und vergleicht sie miteinander: eine regelbasierte Geburtsstaatssignierung basierend auf Leitdateien und den Einsatz von maschinellen Lernverfahren.

➤ **Keywords:** country of birth – classification – migration statistics – random forest – machine learning

ABSTRACT

Since reference year 2008, the data provided for German official migration statistics have included information on the country of birth of immigrants, emigrants, or persons moving their main residence inside Germany. Due to insufficient quality, these data have only been used for estimations in a breakdown by country-of-birth groups in order to meet European data transmission obligations. To enable future analysis of the migrants' country of birth, various methodological investigations were carried out in 2019 to develop a method for automatically checking the consistency of and classifying the relevant variable. This article presents two approaches and compares them with each other: on the one hand, a rule-based country-of-birth classification using reference files and, on the other, the use of machine learning methods.

1

Einleitung

Die amtliche Wanderungsstatistik weist die Anzahl der Zu- und Fortzüge über Gemeindegrenzen nach verschiedenen demografischen Merkmalen nach. Das Merkmal „Geburtsstaat“ ist seit 2008 Erhebungsmerkmal der Wanderungsstatistik nach dem Bevölkerungsstatistikgesetz¹. Allerdings wird diese Angabe bisher aufgrund unzureichender Datenqualität lediglich für Schätzungen nach Staatengruppen für internationale Lieferverpflichtungen genutzt (Mundil/Grobecker, 2011; Carow und andere, 2019). Ein wesentliches Problem für die Datenqualität besteht darin, dass in mehr als 50 % der Lieferdaten aus dem Meldewesen das Merkmal „Geburtsstaat“ nicht befüllt ist. Zur hohen Zahl an fehlenden Angaben trägt vor allem bei, dass gemäß den melderechtlichen Vorgaben der Geburtsstaat nur im Falle einer Geburt im Ausland anzugeben ist. Somit wird das Feld für in Deutschland geborene Personen in der Regel leer gelassen. Hierdurch entfällt die Möglichkeit zu unterscheiden, ob der Geburtsstaat „Deutschland“ ist oder der Staat nicht bekannt ist beziehungsweise die Eintragung fälschlicherweise nicht erfolgte. Die Angabe zum Geburtsstaat kann somit nicht ungeprüft übernommen werden, sondern sie muss plausibilisiert und gegebenenfalls neu kodiert werden. Hierfür bietet sich insbesondere das Merkmal „Geburtsort“ an, das dabei helfen kann, von der Angabe des Ortes auf den Staat zu schließen (zum Beispiel Damaskus → Syrien).

Im Jahr 2019 wurden Methoden zur automatisierten Plausibilisierung und Signierung des Merkmals „Geburtsstaat“ untersucht. Daran beteiligt waren die Arbeitsbereiche im Statistischen Bundesamt, die die Wanderungsstatistik sowie das Maschinelle Lernen und die Imputationsverfahren verantworten. Ziel war, fehlende Angaben auf Mikrodatenebene einzusetzen und eine Veröffentlichung der Wanderungen nach dem Geburtsstaat zu ermöglichen, die den hohen Qualitätsanspruch der amtlichen Statistik erfüllt.

1 Gesetz über die Statistik der Bevölkerungsbewegung und die Fortschreibung des Bevölkerungsstandes (Bevölkerungsstatistikgesetz – BevStatG) vom 20. April 2013 (BGBl. I Seite 826), das zuletzt durch Artikel 9 des Gesetzes vom 18. Dezember 2018 (BGBl. I Seite 2639) geändert worden ist.

Der vorliegende Aufsatz stellt die wesentlichen Ergebnisse dieser Untersuchungen vor. Er zieht zwei ausgewählte Methoden zur Kodierung des Geburtsstaats heran: eine regelbasierte Signierung sowie die Anwendung von maschinellen Lernverfahren. Kapitel 2 gibt einen Überblick über die Rahmenbedingungen für die Signierung. Auf die regelbasierte Signierung des Geburtsstaats geht Kapitel 3 ein, Kapitel 4 behandelt die Signierung mit Verfahren des maschinellen Lernens. Kapitel 5 vergleicht die Ergebnisse beider Methoden miteinander, bewertet die Verfahren im Hinblick auf ihre Einsetzbarkeit und analysiert die Unterschiede in den Ergebnissen. Die Ergebnisse des Vergleichs fasst Kapitel 6 zusammen und gibt einen Ausblick auf das weitere Vorgehen zur Implementierung der Kodierung.

2

Rahmenbedingungen für die Signierung

2.1 Überblick zur Wanderungsstatistik

Die Wanderungsstatistik umfasst alle durch die Meldebehörden registrierten Zu- und Fortzüge mit Verlegung der Haupt- oder alleinigen Wohnung über die Gemeindegrenze. Für das Berichtsjahr 2018 wurden etwa 6,7 Millionen Wanderungsfälle ermittelt (rund 1,6 Millionen Zuzüge aus dem Ausland, 1,2 Millionen Fortzüge in das Ausland und 3,9 Millionen Binnenumzüge). Dies stellt die Größenordnung der jährlich zu signierenden Wanderungsfälle dar.

Der Merkmalsumfang der Wanderungsstatistik begrenzt grundsätzlich die Variablen, die bei der Kodierung berücksichtigt werden können. Darüber hinaus ist allerdings in weiteren Bewegungsstatistiken zu Geburten und Sterbefällen ebenfalls der Einsatz einer Kodierungsfunktion für den Geburtsstaat vorgesehen. Daher wird eine möglichst universelle Einsetzbarkeit der Kodierungsfunktion angestrebt. Die Erhebungsmerkmale der Wanderungsstatistik nach dem Bevölkerungsstatistikgesetz zeigt [Übersicht 1](#) auf Seite 100.

Übersicht 1

Merkmalsumfang der Wanderungsstatistik nach §4 Bevölkerungsstatistikgesetz

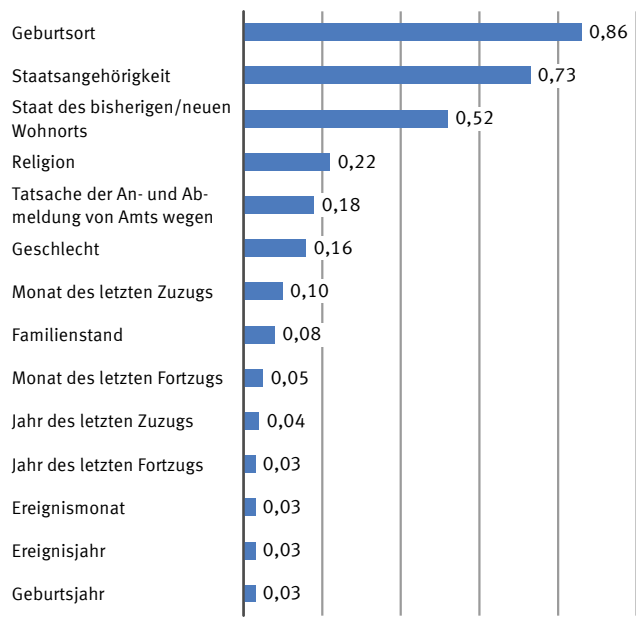
Merkmal	Skalierung
Ereignisdatum (Tag des Einzugs, Auszugs oder Wechsels des Wohnungsstatus)	Datumswert
Bisheriger Wohnort	› Staat (einschließlich unbekannt), 3-stelliger Schlüssel › Amtlicher Gemeindeschlüssel (AGS), 8-stelliger Schlüssel
Neuer Wohnort	› Staat (einschließlich unbekannt), 3-stelliger Schlüssel › Amtlicher Gemeindeschlüssel (AGS), 8-stelliger Schlüssel
Geschlecht	› männlich › weiblich › divers › ohne Angabe
Geburtsdatum	Datumswert
Familienstand	Feste Codeliste
Religion	Feste Codeliste
Staatsangehörigkeit	3-stelliger Schlüssel
Geburtsort	Freitextfeld
Geburtsstaat	3-stelliger Schlüssel
Datum des letzten Zuzugs (bei Fortzügen in das Ausland)	Datumswert
Datum des letzten Fortzugs (bei Zuzügen aus dem Ausland)	Datumswert
Tatsache der An- und Abmeldung von Amts wegen	› nicht zutreffend › Anmeldung von Amts wegen › Abmeldung von Amts wegen

Aus fachlicher Sicht bieten sich für die Signierung neben der Freitextangabe zum Geburtsort insbesondere die Staatsangehörigkeit sowie der Staat des bisherigen Wohnorts und der Staat des neuen Wohnorts an. Dies gilt unter der Annahme, dass viele Personen im Land ihrer Staatsangehörigkeit geboren sind oder aus ihrem Geburtsstaat zuziehen oder dorthin abwandern.

Um einen quantitativen Überblick über Zusammenhänge der möglichen erklärenden Merkmale und des Zielmerkmals zu gewinnen, ist Cramér's V (Cramér, 1946) ein geeignetes Maß. [↗ Grafik 1](#) stellt die Erhebungsmerkmale der Wanderungsstatistik im Berichtsjahr 2018 und ihren Zusammenhang zum gelieferten Geburtsstaat anhand Cramér's V dar. Hierbei zeigt sich, dass der Geburtsort am stärksten mit dem Geburtsstaat korreliert ist. Erwartungsgemäß weisen die Merkmale Staatsangehörigkeit und Staat des bisherigen und neuen Wohnorts auch einen starken quantitativen Zusammenhang mit dem Geburtsstaat auf, der Zusammenhang mit der Staatsangehörigkeit ist dabei stärker ausgeprägt. Weitere Merkmale (Religion, Tatsache der An- beziehungsweise Abmeldung von Amts wegen, Geschlecht) weisen einen geringeren Zusammenhang mit dem Geburtsstaat auf.

Grafik 1

Zusammenhang zwischen dem Merkmal "Geburtsstaat" und weiteren Erhebungsmerkmalen der Wanderungsstatistik 2018
Cramér's V



Cramér's V liegt immer zwischen 0 und 1.

2020 - 01 - 0202

2.2 Das Merkmal „Geburtsort“

Der Geburtsort ist technisch gesehen ein Pflichtmerkmal in den Datenlieferungen an die Statistik und sollte dementsprechend immer vorhanden sein. Die Nutzung des Merkmals „Geburtsort“ birgt allerdings mehrere Schwierigkeiten: Zum einen liegt keine umfassende Liste über alle Orte der Welt und den jeweils zugehörigen Staat vor. Zum anderen handelt es sich im Datensatz des Meldewesens um eine Freitextangabe, sodass unbrauchbare Angaben nicht ausgeschlossen werden können. Beispielsweise werden verschiedene Varianten von „unbekannt“ und „ohne Angabe“ geliefert. Weiterhin ist die Schreibweise des Geburtsorts nicht standardisiert. Ein Ort kann in vielen verschiedenen Schreibweisen sowie mit Schreibfehlern oder Zusatzinformationen geliefert werden. Es werden auch historische oder ausländische Ortsbezeichnungen geliefert. Eine vollumfassende Auflistung aller möglichen Schreibweisen für alle Orte zu erstellen, ist praktisch unmöglich.

Des Weiteren kann eine Ortsbezeichnung in mehreren Staaten vorkommen. „Neustadt“ beispielsweise gibt es als Gemeindename, Ortsteil oder historische Bezeichnung in vielen Staaten. Anhand der Ortsangabe lässt sich in diesen Fällen der Geburtsstaat nicht eindeutig ermitteln, es müssen weitere Informationen herangezogen werden.

2.3 Technischer Rahmen

Die Signierung soll im Zuge der Datenaufbereitung im Fachverfahren der Wanderungsstatistik durch die Statistischen Ämter der Länder erfolgen. So kann sichergestellt werden, dass plausibilisierte Geburtsstaatsangaben für alle Ämter des Statistischen Verbunds² zur Verfügung stehen. Es handelt sich dabei um eine Java-Anwendung mit Zentraler Produktion (das sogenannte OVIS-Anwendungsrahmenwerk; Statistisches Bundesamt, 2020). Meldungen zu Wanderungsfällen sowie Korrekturen werden laufend von den Meldebehörden übermittelt und in den Statistischen Landesämtern plausibilisiert. Das Signierungsverfahren in einem bestehenden Fachverfahren umzusetzen und in die Arbeitsabläufe zu integrieren, stellt eine besondere Herausforderung dar.

² Das sind die Statistischen Ämter des Bundes und der Länder.

Unter Berücksichtigung dieser Rahmenbedingungen wurden zwei grundsätzliche Ansätze für die Kodierung betrachtet: die regelbasierte Signierung sowie die Signierung mit maschinellen Lernverfahren. Innerhalb dieser groben Kategorien wurden jeweils verschiedene Modelle geprüft und die Ergebnisse des besten regelbasierten Modells dem besten Modell auf Basis maschinellen Lernens gegenübergestellt. Die Signierung wurde für das Berichtsjahr 2018 der Wanderungsstatistik durchgeführt.

3

Regelbasierte Signierung

Als regelbasierte Signierung des Geburtsstaats wird in diesem Aufsatz eine deterministische Signierungsmethode bezeichnet: Sie überprüft die Interplausibilität des Geburtsstaats mit dem Merkmal „Geburtsort“ anhand von Leitdateien und imputiert fehlende oder unplausible Ausprägungen des Merkmals „Geburtsstaat“ nach fest definierten Regeln. Um das Merkmal zuverlässig zu signieren, ist eine möglichst vollständige und korrekte Leitdatei entscheidend.

Folgende Leitdateien wurden für die Untersuchung verwendet:

- › Das Ortsverzeichnis des Zensus 2011 deckt weltweit Orte sowie deren Schreibweisen ab, allerdings nur solche, die in den deutschen Melderegistern zum Zeitpunkt des Zensus 2011 enthalten waren.
- › Das Gemeindeverzeichnis der Statistischen Ämter des Bundes und der Länder bildet monatlich alle aktuellen deutschen Gemeinden ab.
- › Das historische Staatenleitband besteht aus rund 1 100 aktuellen sowie historischen Gebiets- und Staatenbezeichnungen.
- › Die Unbekannt-Liste enthält verschiedene Schreibweisen von „unbekannten“ Geburtsortseinträgen (zum Beispiel „nicht bekannt“).

Zusammen decken diese Leitdateien rund 2,2 Millionen unterschiedliche Geburtsort-Staat-Kombinationen ab. Für eine produktive Nutzung eines regelbasierten Verfahrens ist es erforderlich, weitere Ort-Staat-Dateien hinzuzuziehen, die Qualität dieser Dateien zu sichern und sie zu einer Datei zusammenzufügen.

Zusätzlich wurde neben einem direkten Abgleich der gelieferten beziehungsweise standardisierten Geburtsortsangabe mit den Leitdateien noch der Einsatz eines auf dem Jaro-Winkler-Distanzmaß (Jaro, 1989) basierenden Ähnlichkeitsabgleichs untersucht. Ziel war, eine Signierung bei abweichenden Schreibweisen zu erreichen. Dieser Ansatz wurde nicht weiterverfolgt, da die Suche nach gleichen Ortsnamen nach der Standardisierung bereits sehr hohe Zuordnungsquoten lieferte. Die wenigen zu erwartenden zusätzlichen Zuordnungen und die in den Tests relativ hohen Laufzeiten³ waren dafür ausschlaggebend.

³ In den Tests wurde darauf verzichtet, eine Blocking-Strategie (Blakely/Salmond, 2002) für die Suche in den Leitdateien zu implementieren, was die erheblichen Laufzeiten erklärt. Wäre die Anzahl der zu erwartenden zusätzlichen Zuordnungen größer gewesen, hätte die Umsetzung einer Blocking-Strategie die Laufzeiten erheblich reduziert.

Im ausgewählten Signierungsverfahren werden der Geburtsort, der gelieferte Geburtsstaat, die Staatsangehörigkeit und eine Ort-Staat-Häufigkeitsverteilung berücksichtigt. Somit werden lediglich die zwei am stärksten mit dem Geburtsstaat zusammenhängenden Merkmale verwendet (siehe Abschnitt 2.1). Ein Ziel hierbei ist, das Schätzverfahren relativ einfach und gleichzeitig wenig fehleranfällig zu entwickeln, indem nur starke und intuitiv nachvollziehbare Zusammenhänge einfließen. Ein weiteres Ziel ist, die Kodierungsfunktion möglichst universell einsetzen zu können. Dies soll erreicht werden, indem nur Merkmale verwendet werden, die für alle Datensätze der Wanderungsstatistik sowie weiterer Bevölkerungsstatistiken vorliegen.

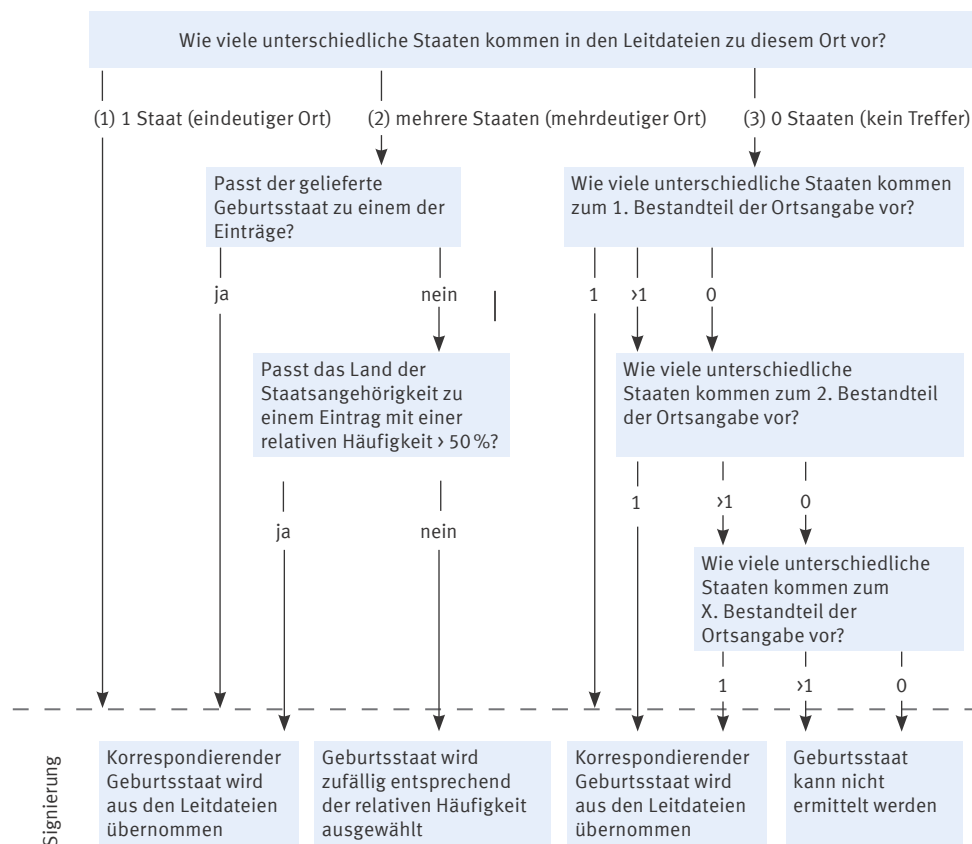
Eine Vorstufe der Signierung unterzieht die Geburtsortsangabe einer automatisierten Standardisierung hinsichtlich Groß-/Kleinschreibung, diakritischen Zeichen und Sonderzeichen sowie Leerzeichen. Das Signie-

rungsverfahren setzt unterschiedliche Regeln ein, abhängig davon, ob eine Ortsangabe (1) in den Leitdateien mit einem einzigen Staat vorkommt (= eindeutige Orte), (2) mit mehreren Staaten vorkommt (= mehrdeutige Orte) oder (3) gar nicht vorkommt. [Grafik 2](#)

Ist die Staatszuordnung eindeutig, erfolgt die Signierung mit diesem Geburtsstaat. Ist die Geburtsortsangabe hingegen mehrdeutig, sind weitere Informationen heranzuziehen. Zuerst wird die gelieferte Angabe zum Geburtsstaat überprüft. Wurde der Datensatz bereits mit einer gültigen Angabe zum Geburtsstaat geliefert und ist die Kombination Ort-Staat in den Leitdateien referenziert, wird angenommen, dass der gelieferte Geburtsstaat

Grafik 2

Ablauf der regelbasierten Signierung



korrekt ist. Der Geburtsstaat wird entsprechend signiert. Kann eine solche Signierung nicht erfolgen, wird als Nächstes auf eine Ort-Staat-Häufigkeitsverteilung auf Basis der Lieferdaten früherer Berichtsjahre zugegriffen. Hierbei wird zunächst geprüft, ob das mit der Staatsangehörigkeit übereinstimmende Land mit einer Häufigkeit größer 50 % auftritt. So wird angenommen, dass Wandernde zwar nicht immer die Staatsangehörigkeit des Landes aufweisen, in dem sie geboren sind, dass die Wahr-

scheinlichkeit dafür jedoch sehr hoch ist, wenn ihre Staatsangehörigkeit zum häufigsten Staat für diesen Geburtsort passt. (Zum Beispiel kann eine Person mit französischer Staatsangehörigkeit außerhalb Frankreichs geboren sein, dies ist aber sehr unwahrscheinlich, sofern ihr Geburtsort Paris ist. Das Gleiche gilt aber nicht für eine Person mit US-amerikanischer Staatsangehörigkeit mit dem Geburtsort Paris – obwohl es eine Stadt namens „Paris“ in den Vereinigten Staaten gibt.) Konnte immer noch kein Geburtsstaat für diesen mehrdeutigen Ort signiert werden, so wird der Staat aus den mit dem Ort referenzierten Staaten zufällig entsprechend der prozentualen Häufigkeit imputiert. So wird eine plausible Verteilung nach Staaten ermittelt, obwohl eine personenscharfe Zuweisung nicht sichergestellt ist.

Kommt die Ortsangabe gar nicht in den Leitdateien vor, wird sie nach der Position von fest definierten Zerlegungstriggern⁴ in mehrere Bestandteile aufgeteilt und diese Bestandteile werden mit den Leitdateien abgeglichen. Ein Beispiel dafür ist die Aufteilung von „Frankfurt a. M. (Hessen)“ in „Frankfurt a. M.“ und „Hessen“. Sobald ein eindeutiger Treffer gefunden wird, wird der Geburtsstaat signiert; weitere Ortsnamensbestandteile werden nicht weiter berücksichtigt. Gibt es nach Ab-

Tabelle 1

Regelbasierte Signierung der Wanderungen 2018 nach Zuweisungsschritten

Abgleich Leitdateien	Grund der Signierung	Ergebnis der Signierung	Signierte Datensätze	
			Anzahl	%
(1) Eindeutiger Ort	Der Geburtsort kommt eindeutig in den Leitdateien (ohne Unbekannt-Liste) vor.	bekannter Geburtsstaat	4 462 997	66,4
(2) Mehrdeutiger Ort	Die Kombination Geburtsort und gelieferter Geburtsstaat kommt in den Leitdateien vor.	bekannter Geburtsstaat	1 205 491	17,9
(2) Mehrdeutiger Ort	Die Kombination Geburtsort und Land der Staatsangehörigkeit kommt in den Leitdateien mit einer relativen Häufigkeit > 50 % vor.	bekannter Geburtsstaat	546 815	8,1
(2) Mehrdeutiger Ort	Der Geburtsstaat wird zufällig aus mehreren mit dem Geburtsort vorkommenden Staaten ausgewählt.	bekannter Geburtsstaat	80 799	1,2
(3) Unbekannter Ort	Ein Ortsnamensbestandteil kommt eindeutig in den Leitdateien vor.	bekannter Geburtsstaat	99 140	1,5
(1) Eindeutiger Ort oder (3) Unbekannter Ort	Der Geburtsstaat kann nicht ermittelt werden.	Geburtsstaat ist unbekannt	321 177	4,8
Insgesamt			6 716 419	100

gleich aller Bestandteile keinen oder keinen eindeutigen Treffer, so wird der Geburtsstaat als „unbekannt“ signiert. Einen Überblick über die Zuweisungsschritte der regelbasierten Signierung des Geburtsstaats in den Wanderungen 2018 gibt [Tabelle 1](#).

4

Maschinelles Lernen

Maschinelle Lernverfahren sind in der Lage, Muster in Datenbeständen zu erkennen. Überwachte Verfahren lernen Muster auf Basis von Eingabedaten, in denen der zu lernende Zusammenhang bereits vorhanden ist. Im hier diskutierten Fall liegen Geburtsstaatskodierungen aus Vorperioden vor; daher wurde geprüft, ob überwachte Lernverfahren geeignet sind, den Geburtsstaat zu signieren.

Bei der Auswahl der zu testenden Verfahren waren Zeit- und Hardware-Restriktionen zu berücksichtigen. Für die Tests in diesem Projekt stand ein Rechner mit 8 Threads und 32 Gigabyte zur Verfügung, als Programmiersprache R.

⁴ Das sind beispielsweise Schrägstriche, Klammern oder Kommas.

4.1 Der Trainingsdatensatz

Um eine Geburtsstaatssignierung mit maschinellen Lernverfahren vorzubereiten, wurde zunächst ein Trainingsdatensatz aus dem Jahresmaterial 2017 der Wanderungsstatistik erstellt. Ein Trainingsdatensatz sollte sehr zuverlässig sein, damit durch die maschinellen Lernverfahren keine falschen Muster gelernt werden. Um die Zuverlässigkeit der Ort-Staat-Zuordnungen zu verbessern, wurden die gelieferten Geburtsort-Geburtsstaat-Kombinationen anhand der in Kapitel 3 beschriebenen Leitdateien geprüft. Für den Trainingsdatensatz wurden aus dem Jahresmaterial 2017 diejenigen Datensätze entfernt, deren Geburtsort (1) nicht in den Leitdateien gefunden wurde oder (2) zu einer anderen Staatszuweisung als dem gelieferten Geburtsstaat geführt hat. Datensätze mit dem Geburtsstaat „unbekannt“ wurden nur dann beibehalten, wenn die Geburtsortsangabe offensichtlich unbrauchbar war. Das traf beispielsweise zu, wenn der Datensatz gar keine alphanumerischen Zeichen oder eine Variante von „unbekannt“ enthielt. Somit bestand der Trainingsdatensatz aus rund 6,2 Millionen Wanderungsdatsätzen.

4.2 Zeichenkettenähnlichkeit und N-Gramme

Eine wichtige Besonderheit bei Freitextfeldern sind Abweichungen in der Schreibweise (siehe Abschnitt 2.2). Um diese Abweichungen beim Training der Lernverfahren berücksichtigen zu können, wurden die Geburtsorte in N-Gramme zerlegt. N-Gramme (Kondrak, 2005) sind alle Fragmente eines Textes oder einer Zeichenkette mit N auf-

einanderfolgenden Elementen. Bigramme bestehen entsprechend aus zwei aufeinanderfolgenden Elementen.¹⁵

Für die vorliegende Anwendung wurden sämtliche Bigramme¹⁶ ermittelt, die in allen beobachteten Ausprägungen des Geburtsorts vorkommen. Anschließend wurde für jede Beobachtung ermittelt, wie oft jedes N-Gramm in der Ausprägung des Merkmals „Geburtsort“ vorkommt. Das Ergebnis ist eine Matrix mit der Anzahl der Beobachtungen und der Anzahl der Bigramme als Dimension, deren Elemente nichtnegative ganze Zahlen sind. Für das Lernen macht man sich zunutze, dass ähnliche Zeichenketten auch ähnliche Ausprägungen der Matrixelemente haben (→ Tabelle 2). Um diese Bigramme wurde das Trainingsmaterial ergänzt.

4.3 Maschinelle Lernverfahren

Für die Tests mit maschinellen Lernverfahren kamen in Anbetracht der Hardware-Restriktionen Naïve-Bayes- sowie baumbasierte Klassifikatoren in Betracht. Als erklärende Variablen wurden zunächst alle im Datensatz vorliegenden Merkmale verwendet. Die Verfahren wurden zunächst anhand des Trainingsmaterials von 2017 und eines 15-%-Testdatensatzes getestet. Nachdem der Naïve-Bayes-Klassifikator (Webb, 2011) mit dem Trainingsdatensatz ohne Bigramm-Erweiterung sehr schnell relativ gute Ergebnisse lieferte (rund 87% korrekte Signierungen), gelang es nicht, mit dem um Bigramme erweiterten Datensatz bessere Ergebnisse zu erzielen.

5 Zum Beispiel sind „Pa“, „ar“, „ri“ und „is“ die Bigramme von „Paris“.

6 Es wurden Bigramme verwendet, weil sie relativ wenige zusätzliche Merkmale erzeugen. Die Bigramme wurden mit dem R-Paket stringdist (van der Loo, 2014) erstellt.

Tabelle 2

Beispielhafte Matrix von Schreibweisen und Bigrammen

Schreibweisen von Mumbai (standardisiert)	Bigramme																			
	MB	BA	MU	UM	AI	_M	AY	BO	I_	OM	M_	MS	SA	_S	AS	BI	BS	IA	Y_	
MUMBAI_M_S	1	1	1	1	1	1	0	0	1	0	1	0	0	1	0	0	0	0	0	
MUMBAI_MS	1	1	1	1	1	1	0	0	1	0	0	1	0	0	0	0	0	0	0	
MUMBAY	1	1	1	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	
MUMBIA	1	0	1	1	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	
MUMBSAI_MS	1	0	1	1	1	1	0	0	1	0	0	1	1	0	0	0	1	0	0	
BOMBAY	1	1	0	0	0	0	1	1	0	1	0	0	0	0	0	0	0	0	0	
BOMBAY_M_S	1	1	0	0	0	1	1	1	0	1	1	0	0	1	0	0	0	0	1	
BOMBAI	1	1	0	0	1	0	0	1	0	1	0	0	0	0	0	0	0	0	0	

Aus diesem Grund wurden weitere Tests mit baumbasierten Verfahren und speziell mit Random Forest durchgeführt.

Random Forest (Breiman, 2001) ist ein Algorithmus des überwachten Lernens, der auf Basis von Entscheidungsbäumen Klassifikations- und Regressionsmodelle anpasst. Dabei wird im Training mit den Merkmalsausprägungen und anhand der in den Trainingsdaten bekannten Klassen eine Reihe von Entscheidungsbäumen errechnet. Im Klassifikationsschritt werden die ermittelten Bäume verwendet, um anhand der Merkmalsausprägungen einer Beobachtung die unbekannte Klasse zu schätzen. Der Random-Forest-Algorithmus erzielte mit einem Trainingsdatensatz, der um Bigramme erweitert wurde, jedoch zusätzlich nur die Merkmale „Staatsangehörigkeit“ und „Staat des bisherigen Wohnorts“ enthält, erheblich bessere Ergebnisse als der Naïve-Bayes-Klassifikator.¹⁷ Hier wurden bezogen auf den Testdatensatz insgesamt 97 % korrekte Zuordnungen erreicht.¹⁸ Aufgrund dieses Ergebnisses wurde die Signierung des Geburtsstaats in der Wanderungsstatistik 2018 mit diesem Modell geprüft.

7 Auf weitere erklärende Merkmale musste aufgrund der bereits erwähnten Hardware-Beschränkungen verzichtet werden. Die Merkmale „Staatsangehörigkeit“ und „Staat des bisherigen Wohnorts“ wurden ausgewählt, weil sie einen relativ hohen Zusammenhang mit dem Geburtsstaat haben (siehe Grafik 1).

8 Das Modell wurde parallel auf 8 Kernen mit knapp 32 Gigabyte Arbeitsspeicher mit dem R-Paket ranger (Wright/Ziegler, 2017) in gut zwei Tagen angepasst.

5

Ergebnisse und Bewertung

5.1 Ergebnisse

Nachdem das regelbasierte Signierungsverfahren und ein Random-Forest-Modell vorlagen, wurden die Geburtsstaaten in den Wanderungsdaten des Jahres 2018 mit beiden Verfahren signiert und die Ergebnisse verglichen. Bei der Beurteilung einer Übereinstimmung zwischen geliefertem und signiertem Wert wurden fehlende Geburtsstaatsangaben mit „Deutschland“ gleichgesetzt, da gemäß der melderechtlichen Regelungen der Geburtsstaat nur im Falle einer Geburt im Ausland anzugeben ist (siehe Kapitel 1).¹⁹

➤ Tabelle 3 zeigt, dass beide Verfahren einen Großteil der Wanderungsfälle mit einem bekannten Geburtsstaat signieren. Bei der regelbasierten Signierung entspricht dies rund 95 %, bei der Signierung mit Random Forest 99 % der Wanderungsfälle. Beim Random-Forest-Ansatz verblieb somit in etwa 50 000 Fällen der Geburtsstaat mit „unbekannt“ signiert gegenüber 320 000 Fällen bei der regelbasierten Signierung.

Im Folgenden soll näherungsweise die Zuverlässigkeit der Signierungsergebnisse ermittelt werden. Dazu werden die rund 6,4 Millionen beziehungsweise 6,7 Millionen Fälle mit signierten Geburtsstaaten beider Verfahren

9 Ebenso werden Umkodierungen mit ehemaligen Staaten und deren Nachfolgestaaten (zum Beispiel der gelieferte Wert ist Sowjetunion und der signierte Wert Russland) sowie mit abhängigen Gebieten und deren Mutterstaat als Übereinstimmung ausgewiesen.

Tabelle 3

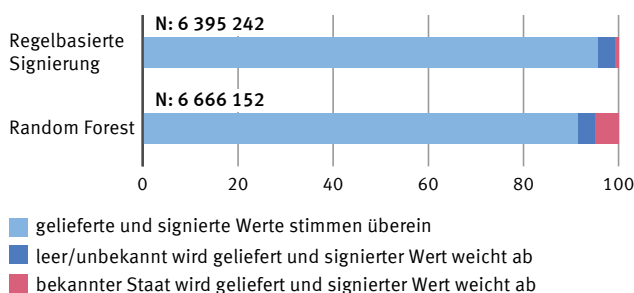
Zuweisungsquote der regelbasierten Signierung gegenüber Signierung mit Random Forest-Modell zum Geburtsstaat anhand der Wanderungsstatistik 2018

	Regelbasierte Signierung		Random Forest	
	Anzahl	%	Anzahl	%
Insgesamt	6 716 419	100	6 716 419	100
Geburtsstaat signiert	6 395 242	95,2	6 666 152	99,3
Gelieferte und signierte Werte stimmen überein	6 110 490	91,0	6 093 769	90,7
Leer/unbekannt wird geliefert und signierter Wert weicht ab	232 353	3,5	248 857	3,7
Bekannter Staat wird geliefert und signierter Wert weicht ab	52 399	0,8	323 526	4,8
„unbekannt“ signiert	321 177	4,8	50 267	0,7

ren betrachtet. Abgleiche dieser Signierungsergebnisse mit deren gelieferten Geburtsstaaten ergeben eine Größenordnung, in wie vielen Fällen die Signierung korrekt funktioniert. Hierunter fallen für beide Signierungsverfahren rund 6,1 Millionen Datensätze, für die die Signierung den gelieferten Geburtsstaat ergibt. Der signierte Geburtsstaat wird somit als wahrscheinlich korrekt angenommen. Hinzu kommen für beide Verfahren zwischen 230 000 und 250 000 Fälle, in denen der Geburtsstaat „leer“ oder „unbekannt“ geliefert wurde und durch das Signierungsverfahren auf einen gültigen Staat (außer Deutschland bei einem leeren Geburtsstaat) signiert wird. Unter diesen Fällen befinden sich wahrscheinlich viele Fälle, in denen die Eintragung fälschlicherweise nicht vorgenommen wurde (siehe Kapitel 1). Somit sollte in diesen Fällen in der Regel der signierte Geburtsstaat zuverlässiger sein als die Angabe „leer“, die auf eine Geburt in Deutschland hindeutet. Beim regelbasierten Verfahren wird in rund 50 000 Fällen ein anderer als der gelieferte bekannte Staat signiert, im Falle der Signierung mit Random Forest sind es mehr als 320 000 Fälle. Dabei handelt es sich wahrscheinlich meist um Fehlzusweisungen durch das Signierungsverfahren, wobei in Einzelfällen auch der gelieferte Staat falsch sein kann. Bei der regelbasierten Signierung liegt somit eine Fehlzusweisung in der Größenordnung von 0,8 % bezogen auf 6,4 Millionen signierte Fälle vor, beim Random-Forest-Verfahren sind es 4,9 % bezogen auf 6,7 Millionen signierte Fälle. Die höhere Zahl signierter Datensätze des Random-Forest-Modells (99,3 %), die aufgrund von weniger „unbekannt“-Zusweisungen gegenüber der regelbasierten Signierung zustande kommt, geht offenbar mit mehr Fehlzusweisungen in ähnlicher Größenordnung einher. ➤ Grafik 3

Grafik 3

Abgleich gelieferter und signierter Geburtsstaaten bei erfolgter Signierung in %



2020 - 01 - 0204

Hintergrund der unterschiedlichen Zuweisungsquoten ist, dass das regelbasierte Verfahren bei ergebnisloser Suche in den Leitdateien keine Zuweisung vornimmt, während das maschinelle Lernverfahren den Verteilungen im Trainingsmaterial entsprechend in der Regel einen Staat zuweist.

➤ **Tabelle 4** vergleicht nun detailliert die Ergebnisse der regelbasierten Signierung, der Signierung des Random-Forest-Modells und den für das Berichtsjahr 2018 von den Meldeämtern an die Statistik gelieferten Geburtsstaat.

In fast 88 % der Wanderungsfälle 2018 wurde die gelieferte Geburtsstaatsangabe durch beide Signierungsverfahren bestätigt. Die Angabe ist somit mit hoher Wahrscheinlichkeit korrekt. In knapp 4 % der Fälle stimmen beide Signierungsverfahren überein, ermitteln aber einen anderen Geburtsstaat als den gelieferten. Der gelieferte Staat wird somit als unplausibel angenommen. Es handelt sich hierbei insbesondere um „leer“ gelieferte Geburtsstaatsfelder, die nicht auf Deutschland kodiert werden (siehe Zeile B1). Deutlich seltener kommt es vor, dass ein gelieferter Staat von beiden Verfahren als unbekannt kodiert wird (B2) oder dass ein ausländischer Staat von beiden Verfahren auf einen anderen gültigen Staat umgeschlüsselt wird (B3).

Etwa 8 % der Wanderungsfälle weisen eine abweichende Signierung durch das regelbasierte Verfahren und das Random-Forest-Modell auf. Darunter sind etwa 3 % der Wanderungsfälle, für die die regelbasierte Signierung den gelieferten Geburtsstaat ermittelt hat. Das Ergebnis der regelbasierten Signierung wird somit als plausibel und das Ergebnis der Signierung mit dem Random-Forest-Modell als wahrscheinlich falsch angenommen. In weiteren 3 % der Wanderungsfälle ergibt die Signierung mit Random Forest den gelieferten Geburtsstaat. Die regelbasierte Signierung weicht hiervon häufig ab, da das Verfahren keinen Geburtsstaat ermitteln konnte (169 707 Datensätze, das entspricht 83 %). Für rund 2 % der Datensätze liefert der Vergleich unklare Ergebnisse. In den meisten Fällen weichen der gelieferte Staat sowie beide Signierungsergebnisse voneinander ab.¹⁰ Ohne manuelle Prüfung ist in diesen Fällen die Richtigkeit

¹⁰ Weiterhin können hierunter Fälle ehemaliger Staaten oder abhängiger Gebiete fallen, beispielsweise wenn der Staat „Jugoslawien“ geliefert wurde und beide Signierungsverfahren unterschiedliche Nachfolgestaaten von Jugoslawien ermitteln.

Tabelle 4

Wanderungsstatistik 2018: Gelieferter Geburtsstaat gegenüber Signierungsergebnissen

Gruppe	Gelieferter Wert = Random Forest	Gelieferter Wert = regelbasierte Signierung	Regelbasierte Signierung = Random Forest	Bedeutung	Anzahl	Anteil in %
Insgesamt					6 716 419	100
A	ja	ja	ja	Übereinstimmung aller drei Werte	5 905 527	87,93
B	nein	nein	ja	Gelieferter Wert wahr- scheinlich inkorrekt	256 127	3,81
B1	gelieferter Wert ist leer/unbekannt				204 163	3,04
B2	gelieferter Wert ist bekannt und signierter Wert ist unbekannt				25 686	0,38
B3	gelieferter und signierter Wert entsprechen bekanntem Schlüssel				26 278	0,39
C	nein	ja	nein	Regelbasierte Signierung wahrscheinlich korrekt	222 759	3,32
C1	Regelbasierte Signierung ergibt „Deutschland“				18 672	0,28
C2	Regelbasierte Signierung ergibt ausländischen Staat				202 616	3,02
C3	Regelbasierte Signierung ergibt unbekannten Staat				1 471	0,02
D	ja	nein	nein	Random Forest wahrscheinlich korrekt	204 717	3,05
D1	Random Forest ergibt „Deutschland“				25 415	0,38
D2	Random Forest ergibt ausländischen Staat				179 298	2,67
D3	Random Forest ergibt unbekannten Staat				4	0,00
E	Sonstiges			Keine Übereinstimmung	127 289	1,90

sowohl der erreichten Kodierungen als auch der gelieferten Werte nicht feststellbar.

5.2 Bewertung

Zusammenfassend liefern beide Signierungsverfahren ähnlich gute Ergebnisse. Beide Verfahren sind in der Lage, fälschlicherweise fehlende Geburtsstaatseinträge relativ zuverlässig zu identifizieren und können somit einen systematisch zu hohen Ausweis des Geburtsstaats „Deutschland“ vermeiden. Für beide Verfahren liegt die Zahl an unbekannten Zuordnungen nach der Signierung unter 5%. Die Verfahren liefern zu einem hohen Teil identische Ergebnisse, aber in rund 8% der Fälle abweichende Zuordnungen. Dies liegt zum Teil darin begründet, dass das Random-Forest-Modell fast immer eine Staatszuweisung trifft, wenn nicht recht sicher ein unbekannter Staat festgestellt wurde. Im Falle des regelbasierten Verfahrens hingegen wird „unbekannt“ zugewiesen, wenn der Ort beziehungsweise ein Bestandteil nicht in den Leitdateien gefunden wurde. Infolge dieses ersten Tests erweisen sich beide Verfahren als grundsätzlich für die Kodierung aus methodischer Sicht geeignet.

Neben der fachlichen Eignung müssen daher weitere Kriterien für die Entscheidung für ein Verfahren zum Einsatz in der Fachanwendung der Wanderungsstatistik herangezogen werden. [↗ Übersicht 2 auf Seite 108](#)

Beide Verfahren haben Potenzial, noch zuverlässigere Signierungsergebnisse zu erreichen. Die Quote der im regelbasierten Verfahren als „unbekannt“ signierten Fälle kann durch die Erstellung, Qualitätssicherung und laufende Aktualisierung einer umfangreichen Ortsdatei weiter sinken. In den Tests wurde beispielsweise mit dem Ortsverzeichnis des Zensus 2011 eine nicht aktualisierte Datei verwendet. Dies wirkt sich insofern aus, dass es seit 2011 vermehrt Zuwanderungen aus Staaten gegeben haben dürfte, für die bislang wenige Orte in der Datei erfasst sind. Beispielsweise umfasst die Datei nur wenige syrische Orte, während Syrien 2018 der vierthäufigste gemeldete Geburtsstaat für Wandernde ist. Das maschinelle Lernverfahren kann durch Wahrscheinlichkeitsschwellen relativ „unsichere“ Signierungen mit „unbekannt“ anstatt mit einem Geburtsstaat signieren. Auch die Erhöhung des Umfangs des Trainingsmaterials sowie das Training mit weiteren N-Grammen mit $N > 2$ kann geprüft werden. Der Einsatz anderer Lernverfahren, wie etwa Gradient-Boosting, kann ebenfalls getestet werden.

Übersicht 2

Bewertungskriterien für den Einsatz

	Regelbasierte Signierung	Signierung mit Methoden des maschinellen Lernens
Weitere Verbesserung der Verfahren	› Qualitätssicherung und Aktualisierung der Ortsdatei	› Umfang des Trainingsdatensatzes vergrößern › Einsatz von Wahrscheinlichkeitsschwellen › Tests weiterer N-Gramme › Tests mit anderen Verfahren
Einsetzbarkeit für andere Statistiken	› abhängig von Merkmalsverfügbarkeit grundsätzlich gegeben	› abhängig von Merkmalsverfügbarkeit grundsätzlich gegeben
Laufende Aufwände	› Pflege einer Ortsdatei	› Durchführung des Trainings
Risiken bei der technischen Umsetzung	› gering, da erforderliche IT-Methoden grundsätzlich bekannt	› technische Unwägbarkeiten hinsichtlich Performanz › geringe Erfahrungswerte der Implementierung solcher Verfahren in bestehende Fachanwendungen
Fachliche Risiken beim Einsatz der Methoden	› Signierung neuer Orte	› geringere Erklärbarkeit und Nachvollziehbarkeit der Signierung › Signierung neuer Orte

Das Projekt hat zudem gezeigt, dass wegen bestehender Hardware-Restriktionen einige maschinelle Lernverfahren nicht getestet und Modelloptimierungen nicht durchgeführt werden konnten. Dies ist zurzeit ein fachübergreifendes Problem, dessen Lösung erhebliches Potenzial für den Einsatz maschineller Lernverfahren birgt.


Beide Verfahren sind grundsätzlich auf andere Statistiken übertragbar, wobei die Verfügbarkeit von Merkmalen sowie Besonderheiten der jeweiligen Statistik berücksichtigt werden müssen. Für beide Verfahren ist ein laufender Aufwand zu erwarten: Für das regelbasierte Verfahren ist die Ortsdatei zu aktualisieren und hierfür fehlende Zuordnungen oder möglicherweise falsche Zuordnungen sind jährlich auszuwerten. Ein Verfahren auf Basis maschineller Lernverfahren muss regelmäßig neu trainiert werden, um aktuelle Entwicklungen einzubeziehen. Bei der technischen Umsetzung werden aktuell größere Risiken bei der Signierung mit Verfahren des maschinellen Lernens erwartet, da IT-seitig zurzeit noch Erfahrungswerte mit der Implementierung solcher Methoden in bestehende Fachanwendungen fehlen. Darüber hinaus wäre das Zustandekommen der Signierungsergebnisse im Einzelfall bei Random Forest weniger gut nachvollziehbar als im Falle der regelbasierten Signierung. Wegen der genannten Argumente wird für die Wanderungsstatistik das regelbasierte Signierungsverfahren ausgewählt.

6

Fazit und Ausblick

Der vorliegende Beitrag hat zwei Verfahren zur Kodierung des Geburtsstaats in der Wanderungsstatistik verglichen: ein regelbasiertes Verfahren auf Basis von Ort-Staat-Leitdateien sowie ein Random-Forest-Modell. Ziel war, ein Verfahren zu identifizieren, das auf Basis vorliegender Merkmale zuverlässig den Geburtsstaat ermittelt. Die Tests ergaben, dass beide Verfahren für die Kodierung des Geburtsstaats in der Wanderungsstatistik grundsätzlich geeignet sind. Weil die Risiken der IT-seitigen Umsetzung geringer sowie die Signierungsergebnisse besser nachvollziehbar sind, wird der Einsatz des regelbasierten Verfahrens in der Wanderungsstatistik und in weiteren Bevölkerungsstatistiken angestrebt. Eine Umsetzung in der Fachanwendung für die Wanderungsstatistik ist beabsichtigt.

Im Zuge der Analysen wurden für beide Verfahren noch Verbesserungspotenziale erkannt, die in die künftigen Arbeiten einfließen sollen: Das regelbasierte Verfahren signierte seltener einen Geburtsstaat als das Random-Forest-Modell, die erreichten Zuordnungen waren aber in der Regel zuverlässiger. Bis zum produktiven Einsatz der Signierungsfunktion ist geplant, eine Ortsdatei aus mehreren Ort-Staat-Referenzdateien zu erstellen, deren Qualität zu sichern und um fehlende Ort-Staat-Kombinationen zu ergänzen. Darüber hinaus soll die Ortsdatei

laufend gepflegt werden. Hierbei sollen Datensätze, für die bei der Signierung keine oder eine unplausible Zuweisung getroffen wurde, jährlich ausgewertet und in die Ortsdatei eingepflegt werden. Die Ergebnisse gegebenenfalls erfolgter manueller Nachprüfungen und Korrekturen durch die Statistischen Landesämter sollen ebenfalls laufend automatisiert in die Qualitätssicherung einfließen. Mit diesen Verbesserungen ist zu erwarten, dass Zuverlässigkeit und Vollständigkeit der Ortsdatei und damit der Geburtsstaatssignierungen im Zeitverlauf weiter zunehmen. Weiterhin kann untersucht werden, ob maschinelle Lernverfahren auch bei der Qualitätssicherung und Pflege der Ortsdatei eingesetzt werden können. Hiermit können abweichende Schreibweisen von Orten sowie aktuelle Entwicklungen, die in den Trainingsdaten bereits erkennbar sind, sich in der Ortsdatei aber noch nicht widerspiegeln, bei der Verbesserung der Ortsdatei berücksichtigt werden. Mit den vorgestellten Ergebnissen stehen zukunftsfähige und effiziente Werkzeuge für die Geburtsstaatssignierung zur Verfügung, die in der Wanderungsstatistik angewendet und auf weitere Bevölkerungsstatistiken übertragen werden können. 

LITERATURVERZEICHNIS

- Blakely, Tony/Salmond, Clare. *Probabilistic record linkage and a method to calculate the positive predictive value*. In: International Journal of Epidemiology. Ausgabe 6. Jahrgang 31, 2002, Seite 1246 ff. doi: 10.1093/ije/31.6.1246
- Breiman, Leo. *Random forests*. In: Machine Learning. Ausgabe 1/45, 2001, Seite 5 ff., doi: 10.1023/A:1010933404324
- Cramér, Harald. *Mathematical Methods of Statistics*. Princeton 1946, hier: Seite 282 ff. ISBN 0-691-08004-6
- Carow, Annelen/Mundil-Schwarz, Rabea/Vigneau, Elsa. [*Bevölkerung am üblichen Aufenthaltsort und Weiterentwicklung des Schätzverfahrens zur Langzeitmigration*](#). In: WISTA Wirtschaft und Statistik. Ausgabe 3/2019, Seite 65 ff.
- Jaro, Matthew A. *Advances in record-linkage methodology as applied to matching the 1985 census of Tampa, Florida*. In: Journal of the American Statistical Association. Ausgabe 406. Jahrgang 84, 1989, Seite 414 ff.
- Kondrak, Grzegorz. *N-Gram Similarity and Distance*. In: Consens, Mariano/Navarro, Gonzalo (Herausgeber). SPIRE 2005: String Processing and Information Retrieval. Lecture Notes in Computer Science 3772. Berlin/Heidelberg, 2005.
- Mundil, Rabea/Grobecker, Claire. [*Schätzverfahren zu Langzeitmigranten in Deutschland 2009, Teil 1: Deutsche Personen*](#). In: Wirtschaft und Statistik. Ausgabe 10/2011, Seite 967 ff.
- Statistisches Bundesamt. *OVIS-Anwendungsrahmenwerk. Datenaufbereitung*. [Zugriff am 14. April 2020]. Verfügbar unter: www.destatis.de
- van der Loo, Mark P.J. *The stringdist Package for Approximate String Matching*. In: The R Journal. Jahrgang 6. Ausgabe 1/2014, Seite 111 ff. [Zugriff am 14. April 2020]. Verfügbar unter: <https://CRAN.R-project.org/package=stringdist>.
- Webb, Geoffrey I. *Naïve Bayes*. In: Sammut, Claude/Webb, Geoffrey I. (Herausgeber). Encyclopedia of Machine Learning. Boston 2011.
- Wright, Marvin N./Ziegler, Andreas. *ranger: A Fast Implementation of Random Forests for High Dimensional Data in C++ and R*. In: Journal of Statistical Software. Band 77. 2017, Seite 1 ff., doi: 10.18637/jss.v077.i01.

RECHTSGRUNDLAGEN

Gesetz über die Statistik der Bevölkerungsbewegung und die Fortschreibung des Bevölkerungsstandes (Bevölkerungstatistikgesetz – BevStatG) vom 20. April 2013 (BGBl. I Seite 826), das zuletzt durch Artikel 9 des Gesetzes vom 18. Dezember 2018 (BGBl. I Seite 2639) 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 Juni 2020; Seite 41 korrigiert am 11.11.2020, Seite 1 korrigiert am 12.11.2020

Das Archiv älterer Ausgaben finden Sie unter www.destatis.de

Artikelnummer: 1010200-20003-4, ISSN 1619-2907

© Statistisches Bundesamt (Destatis), 2020

Vervielfältigung und Verbreitung, auch auszugsweise, mit Quellenangabe gestattet.