

Statistisches Bundesamt (Destatis) (Ed.)

**Periodical Part**

## WISTA – Wirtschaft und Statistik, Ausgabe 6/2020

WISTA – Wirtschaft und Statistik

**Provided in Cooperation with:**

Statistisches Bundesamt (Destatis), Wiesbaden

*Suggested Citation:* Statistisches Bundesamt (Destatis) (Ed.) (2020) : WISTA – Wirtschaft und Statistik, Ausgabe 6/2020, WISTA – Wirtschaft und Statistik, ISSN 1619-2907, Statistisches Bundesamt (Destatis), Wiesbaden, Vol. 72, Iss. 6

This Version is available at:

<https://hdl.handle.net/10419/228497>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

---

# WISTA

## Wirtschaft und Statistik

---

Prof. Dr. Walter Krämer	<b>Verleihung des Gerhard-Fürst-Preises 2020</b>
Luis Federico Flores, Christiane Loos	<b>Saisonbereinigung der monatlichen kassenmäßigen Steuereinnahmen</b>
Dr. Elena Schmidt	<b>Korrektur des Tätigkeitsschlüssels der Bundesagentur für Arbeit mithilfe maschineller Lernverfahren</b>
Dr. Hanna Brenzel, Kathrin Gebers	<b>Werkstattbericht: Georeferenzierung im Statistischen Verbund</b>
Sandra Schymura	<b>Beschäftigte und ihre Verdienste nach der zweiten Erhöhung des Mindestlohns</b>
Dr. Sandra Jung, Dr. Wolfhard Kaus	<b>Unternehmensstrukturstatistiken und Statistik für kleine und mittlere Unternehmen nach dem EU-Unternehmensbegriff</b>
Rainer Opfermann	<b>Personenbeförderung und Unfälle auf Binnenwasserstraßen</b>

---

**6 | 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
oder —	grundsätzliche Änderung innerhalb einer Reihe, die den zeitlichen Vergleich beeinträchtigt
/	keine Angaben, da Zahlenwert nicht sicher genug
()	Aussagewert eingeschränkt, da der Zahlenwert statistisch relativ unsicher ist

Abweichungen in den Summen ergeben sich durch Runden der Zahlen.

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

# INHALT

3	Editorial
4	Kennzahlen – Wichtige Indikatoren zur Coronakrise
8	Informationsangebote zur Coronakrise
10	Kurznachrichten
15	Sonderrubrik Deutsche EU-Ratspräsidentschaft <b>Deutsche EU-Ratspräsidentschaft im Bereich Statistik</b> <i>German EU Council Presidency in the field of statistics</i>
19	Prof. Dr. Walter Krämer <b>Verleihung des Gerhard-Fürst-Preises 2020</b> <i>The 2020 Gerhard Fürst Award</i>
23	Luis Federico Flores, Christiane Loos <b>Saisonbereinigung der monatlichen kassenmäßigen Steuereinnahmen</b> <i>Seasonal adjustment of the monthly cash tax revenue</i>
37	Dr. Elena Schmidt <b>Korrektur des Tätigkeitsschlüssels der Bundesagentur für Arbeit mithilfe maschineller Lernverfahren</b> <i>Correction of the occupational code number of the Federal Employment Agency by means of machine learning methods</i>
48	Dr. Hanna Brenzel, Kathrin Gebers <b>Workstattbericht: Georeferenzierung im Statistischen Verbund</b> <i>Workshop report: geo-referencing in the German official statistics network</i>

# INHALT

---

- |    |   |
|----|---|
| 58 | Sandra Schymura<br><b>Beschäftigte und ihre Verdienste nach der zweiten Erhöhung des Mindestlohns</b><br><i>Employees and their earnings after the second in the minimum wage increase</i>  |
| 68 | Dr. Sandra Jung, Dr. Wolfhard Kaus<br><b>Unternehmensstrukturstatistiken und Statistik für kleine und mittlere Unternehmen nach dem EU-Unternehmensbegriff</b><br><i>Structural business statistics and statistics on small and medium-sized enterprises based on the EU enterprise concept</i> |
| 82 | Rainer Opfermann<br><b>Personenbeförderung und Unfälle auf Binnenwasserstraßen</b><br><i>Passenger transport and accidents on inland waterways</i>  |

---

# EDITORIAL

---

Dr. Georg Thiel

---



## LIEBE LESERIN, LIEBER LESER,

wie viele andere Veranstaltungen konnte auch die Verleihung des diesjährigen Gerhard-Fürst-Preises am 19. November 2020 wegen der Corona-Pandemie nur virtuell stattfinden. Das Statistische Bundesamt hat in diesem Jahr vier wissenschaftliche Abschlussarbeiten prämiert. Sie haben nicht nur engen Bezug zur Arbeit der statistischen Ämter, sondern enthalten auch neue Anstöße für weitergehende Untersuchungen oder stellen relevante Verbesserungen gegenüber aktuellen Ansätzen dar. Die Preisträgerin und die Preisträger stellen ihre Arbeiten in den kommenden WISTA-Heften ausführlich vor, zuvor berichten wir in dieser Ausgabe über die Preisverleihung.

Neue Methoden oder methodische Weiterentwicklungen stehen im Fokus weiterer Beiträge: Im Zuge der Corona-Pandemie hat das Statistische Bundesamt erstmals im August 2020 kalender- und saisonbereinigte monatliche Steuereinnahmen veröffentlicht. Sie sind wichtige Indikatoren für die Finanzpolitik und die Konjunkturbeobachtung, gerade in Krisenzeiten. Den Einsatz eines überwachten maschinellen Lernverfahrens zur Korrektur des Tätigkeitsschlüssels der Bundesagentur für Arbeit mithilfe eines Datensatzes aus der Verdienststrukturerhebung beschreibt ein weiterer Artikel. Ein Werkstattbericht zur Georeferenzierung im Statistischen Verbund gibt einen Überblick über die Rahmenbedingungen, die angestoßenen Aktivitäten und die Analyseprojekte auf Basis integrierter statistischer und geografischer Informationen.

Zum Ende eines Jahres, das unser Leben durch die Coronakrise beispiellos geprägt hat, wünsche ich Ihnen trotz der nach wie vor notwendigen Einschränkungen des gesellschaftlichen und wirtschaftlichen Lebens erholsame Festtage und einen guten Start in das Jahr 2021. Bleiben Sie gesund!

A handwritten signature in black ink, reading "Georg Thiel". The signature is written in a cursive, flowing style.

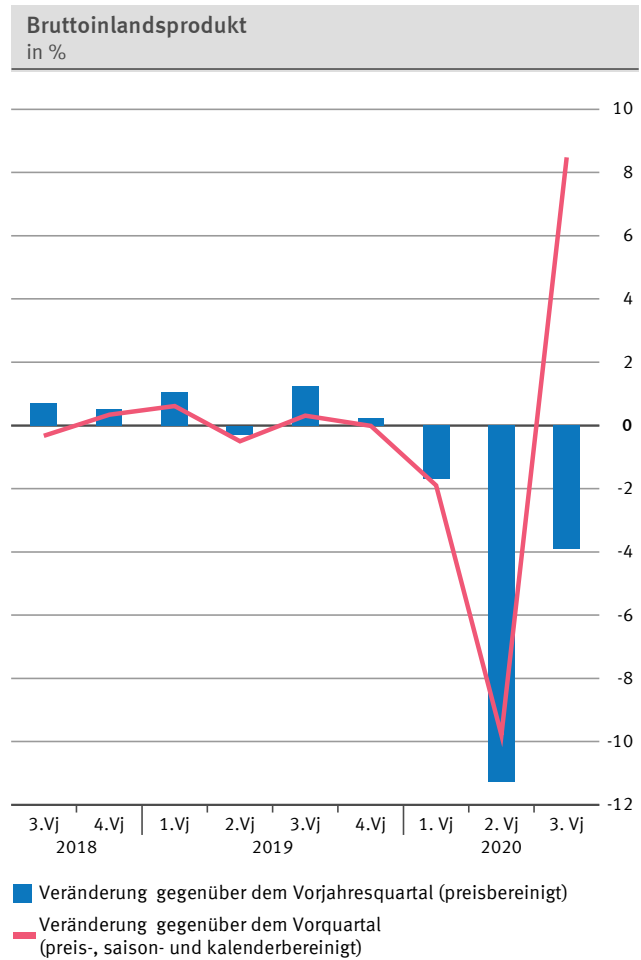
Präsident des Statistischen Bundesamtes

## Auswirkungen der COVID-19-Pandemie auf Wirtschaft und Gesellschaft

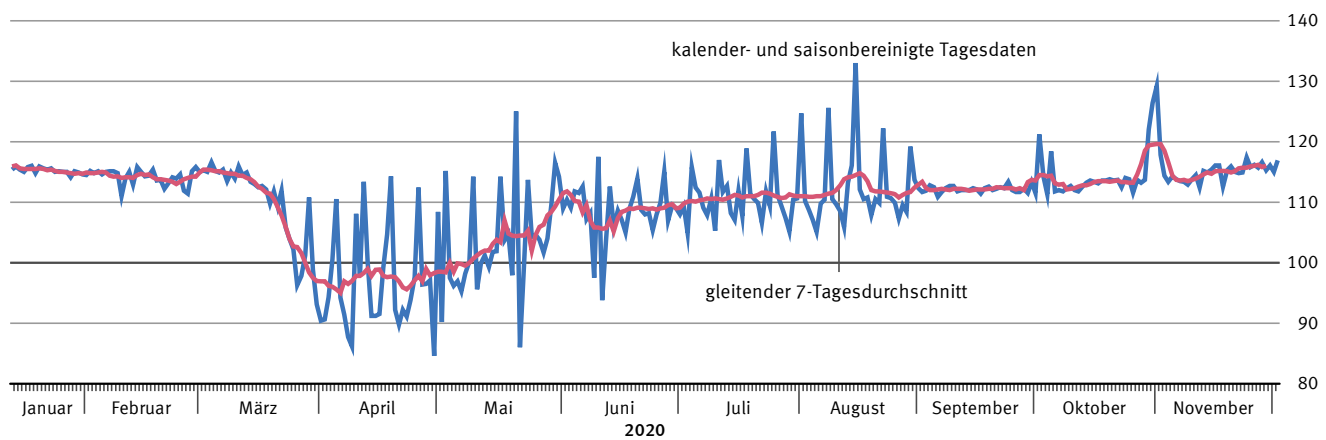
Die globale Ausbreitung des SARS-CoV-2-Virus und der dadurch verursachten Erkrankung COVID-19 seit dem Frühjahr 2020 trifft Gesellschaft und Wirtschaft stark. Die langfristigen Folgen der Corona-Pandemie sind noch nicht überschaubar.

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

Am 30. Oktober 2020 hat das Statistische Bundesamt zum zweiten Mal das Bruttoinlandsprodukt bereits 30 Tage nach Quartalsende veröffentlicht. Der [Podcast BIP t+30](#) informiert zu den Gründen, warum diese Umstellung erfolgte, und erläutert die methodischen Grundlagen der BIP-Schnellschätzung. Mit der BIP t+30-Schnellschätzung stehen hochaktuelle faktenbasierte Wirtschaftsdaten zur Verfügung, die in Krisenzeiten besondere Relevanz haben.



## Lkw-Maut-Fahrleistungsindex ab Januar 2020 2015 = 100

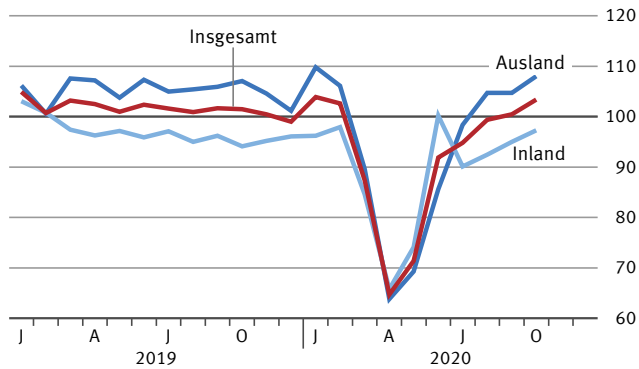


Quellen: Bundesamt für Güterverkehr, Deutsche Bundesbank, Statistisches Bundesamt

Stand: 10.12.2020

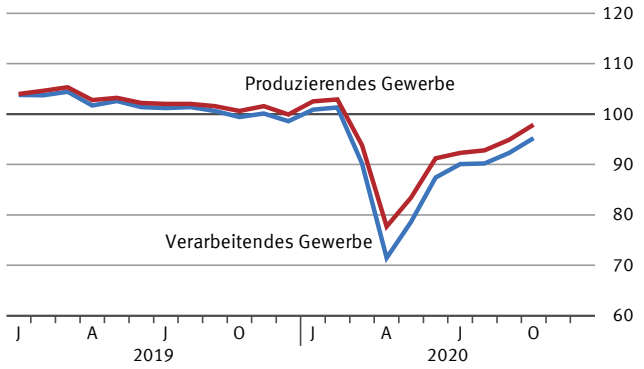
# Kennzahlen – Wichtige Indikatoren zur Coronakrise

**Auftragseingang im Verarbeitenden Gewerbe**  
Volumenindex 2015 = 100



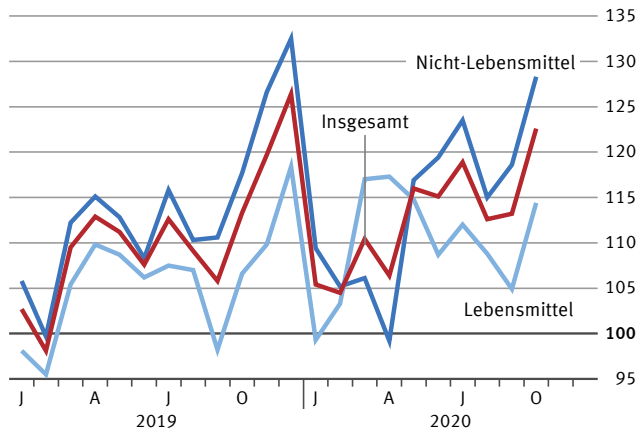
Kalender- und saisonbereinigter Wert nach dem Verfahren X13 JDemetra+. – Vorläufiges Ergebnis.

**Produktion im Produzierenden und Verarbeitenden Gewerbe**  
Index 2015 = 100

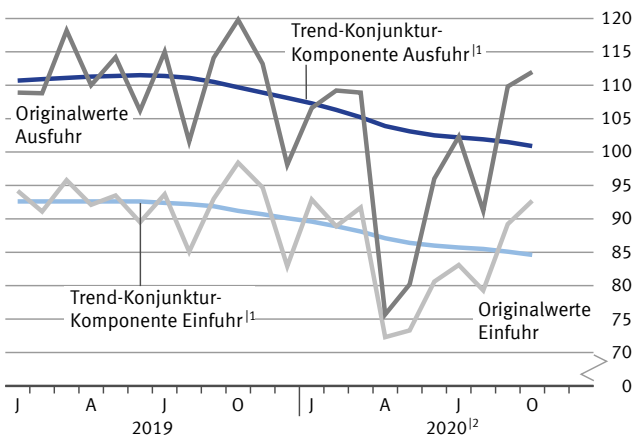


Kalender- und saisonbereinigter Wert nach dem Verfahren X13 JDemetra+. – Vorläufiges Ergebnis.

**Umsatz im Einzelhandel**  
2015 = 100



**Außenhandel**  
in Mrd. EUR



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

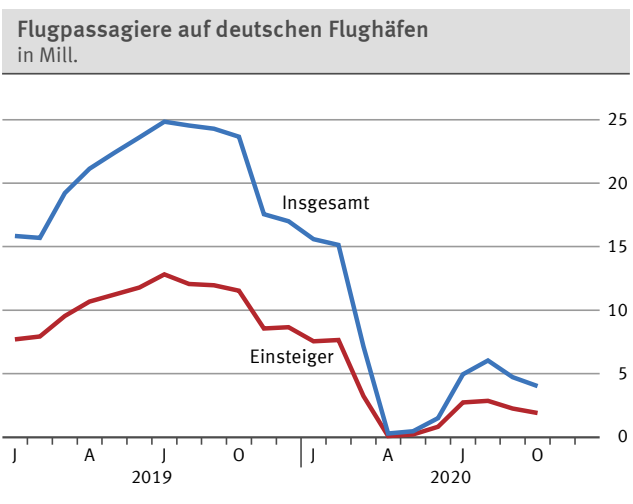
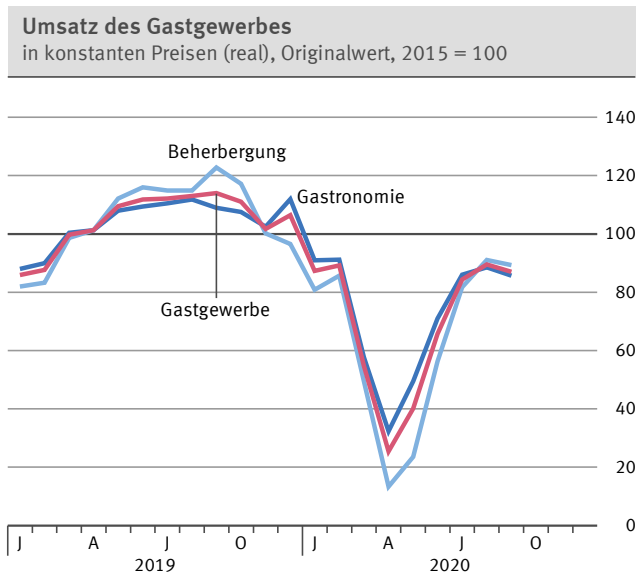
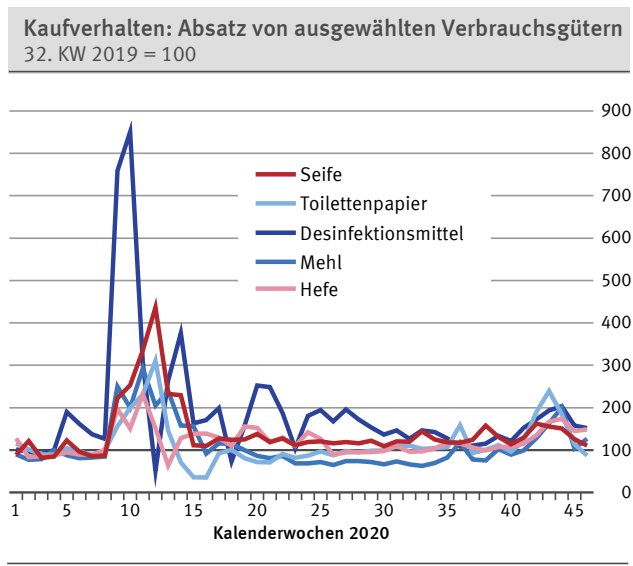
Stand: 10.12.2020



# Kennzahlen – Wichtige Indikatoren zur Coronakrise

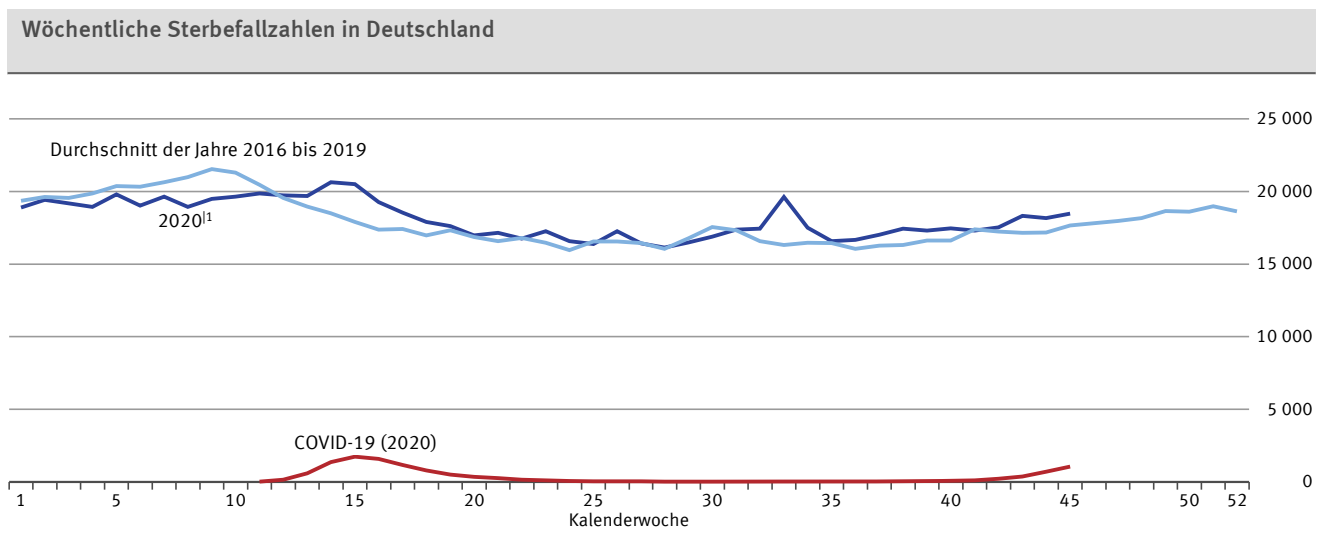
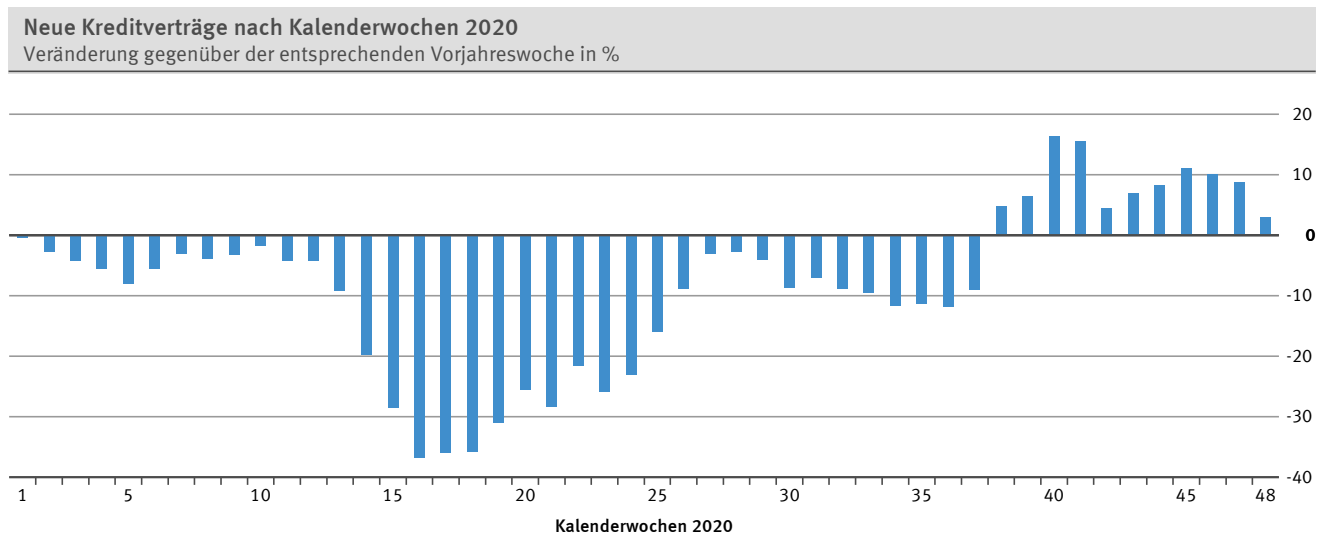
Verbraucherpreisindex 2015 = 100			
2019	2020		
Januar	103,4	Januar	105,2
Februar	103,8	Februar	105,6
März	104,2	März	105,7
April	105,2	April	106,1
Mai	105,4	Mai	106,0
Juni	105,7	Juni	106,6
Juli	106,2	Juli	106,1
August	106,0	August	106,0
September	106,0	September	105,8
Oktober	106,1	Oktober	105,9
November	105,3		
Dezember	105,8		

Veränderung zum Vorjahresmonat  
**105,9** ↓ **-0,2 %**



Stand: 10.12.2020

# Kennzahlen – Wichtige Indikatoren zur Coronakrise



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

Stand: 10.12.2020



### Corona-Statistiken – Auswirkungen der Pandemie

Die globale Ausbreitung des SARS-CoV-2-Virus und die damit verbundenen Maßnahmen zur Eindämmung der Pandemie treffen Gesellschaft und Wirtschaft stark. Die langfristigen Folgen der Corona-Pandemie sind noch nicht überschaubar. Auf der Sonderseite „Corona-Statistiken“ werden statistische Ergebnisse präsentiert, in denen sich bereits die Auswirkungen zeigen oder in denen Auswirkungen zu erwarten sind. Dazu gehören auch neue Statistiken und Auswertungen, die das Statistische Bundesamt kurzfristig im Rahmen der Coronakrise bereitgestellt hat, um den Bedarf an Zahlen am aktuellen Rand zu decken:

- ↳ Sonderauswertung zu Sterbefallzahlen
- ↳ Täglicher Lkw-Maut-Fahrleistungsindex
- ↳ Frühindikator zur Umsatzentwicklung
- ↳ Insolvenzbekanntmachungen
- ↳ Scannerdaten zur Analyse des Kaufverhaltens
- ↳ Kreditvergaben und Auskünfte für Online-Transaktionen
- ↳ Mobilitätsindikatoren
- ↳ [www.destatis.de/corona](http://www.destatis.de/corona)



### Krisenmonitor – Vergleich Corona- und Finanzmarktkrise

In der öffentlichen Diskussion wird häufig versucht, die erwarteten wirtschaftlichen Auswirkungen der Corona-Pandemie mit denen der globalen Finanz- und Wirtschaftskrise 2008/09 zu vergleichen. Zur Unterstützung dieser Diskussion hat das Statistische Bundesamt verschiedene Konjunkturindikatoren zusammengestellt, die die konjunkturellen Entwicklungen während der Corona- und während der Finanzmarktkrise vergleichen. Hierzu wird der Beginn der Corona-bedingten Rezession im ersten Quartal 2020 dem Anfang der Finanzmarktkrise im zweiten Quartal 2008 gegenübergestellt (Monat Januar 2020 zu April 2008). Die Zahlen für das Jahr 2020 werden laufend aktualisiert.

- ↳ [www.destatis.de/corona](http://www.destatis.de/corona)



### EXDAT – Experimentelle Daten

In der neuen Rubrik „EXDAT – Experimentelle Daten“ veröffentlicht das Statistische Bundesamt regelmäßig neue, innovative Projektergebnisse. Sie entstehen auf der Grundlage neuer Datenquellen und Methoden. Im Reifegrad unterscheiden sie sich von amtlichen Statistiken, insbesondere in Bezug auf Harmonisierung, Erfassungsbereich und Methodik. Doch die Corona-Pandemie hat gezeigt, dass experimentelle Daten gerade in Krisenzeiten eine wertvolle und unverzichtbare Ergänzung zum amtlichen Datenangebot sind, weil sie eine schnellere Berichterstattung ermöglichen.

↳ [www.destatis.de/exdat](http://www.destatis.de/exdat)



### EU-Monitor COVID-19

Die COVID-19-Pandemie hat in allen EU-Staaten schwerwiegende Folgen für Wirtschaft und Gesellschaft verursacht. Die Viruserkrankung und die damit einhergehenden Sicherheitsmaßnahmen werden auch noch längere Zeit spürbare Auswirkungen haben. Dieser Monitor zeigt anhand von interaktiven Grafiken, welche EU-Staaten besonders unter den Folgen leiden und wie sich die Lage entwickelt. Dargestellt werden verschiedene Indikatoren aus den Bereichen Wirtschaft und Preise, Industrie und Baugewerbe, Arbeitsmarkt, Dienstleistungen, Exportwirtschaft sowie Bevölkerung und Gesundheit.

↳ [www.destatis.de/corona](http://www.destatis.de/corona)



### Deutsche EU-Ratspräsidentschaft im Bereich Statistik

Deutschland hat vom 1. Juli bis 31. Dezember 2020 den Vorsitz im Rat der Europäischen Union inne. Die Programmplanung der deutschen EU-Ratspräsidentschaft im Bereich Statistik verfolgt das Ziel, die Erfahrungen und Herausforderungen aus der Corona-Pandemie in Verwaltungshandeln umzusetzen.

↳ [www.destatis.de/eu2020](http://www.destatis.de/eu2020)

---

# KURZNACHRICHTEN

---

## INFORMATIONEN ZUR CORONAKRISE

---

### Sonderauswertung zur Mobilität in Corona-Hotspots

Mobilfunkdaten können einen Hinweis darauf geben, wie stark sich das Mobilitätsverhalten der Bürgerinnen und Bürger in sogenannten Corona-Hotspots nach Inkrafttreten von Beschränkungsmaßnahmen zur Eindämmung der Covid-19-Pandemie verändert. Dahinter steht die Annahme, dass mit einer Verringerung der Mobilität auch die Anzahl sozialer Interaktionen und somit die Ansteckungsgefahr zurückgehen.

Eine Sonderauswertung experimenteller Daten des Statistische Bundesamtes zeigt, dass die Mobilität in den Landkreisen und kreisfreien Städten, die in der 42. Kalenderwoche (12. bis 18. Oktober 2020) eine kritische 7-Tage-Inzidenz aufwiesen, gegenüber der 38. Kalenderwoche (14. bis 20. September) um 13% zurückgegangen ist. Wenngleich auch Faktoren wie das Wetter und die touristische Aktivität die Mobilität der Bevölkerung beeinflussen, nahm die Mobilität in den weniger betroffenen Landkreisen und kreisfreien Städten im Vergleichszeitraum lediglich um 7% und damit etwa halb so stark ab.

Wie sich die Mobilität in den einzelnen Landkreisen und kreisfreien Städten verändert hat, zeigt eine [neue interaktive Karte](#). Das Angebot wird zunächst wöchentlich aktualisiert und schrittweise erweitert.

Die Ergebnisse basieren auf einer Sonderauswertung anonymisierter Mobilfunkdaten, die das Statistische Bundesamt insbesondere hinsichtlich ihres Nutzens für die kleinräumige und aktuelle Abbildung der Bevölkerung und ihrer Mobilität untersucht.

Bei diesen experimentellen Daten handelt es sich um neue, innovative Projektergebnisse, die auf der Grundlage neuer Datenquellen und Methoden entstehen. Im Reifegrad unterscheiden sie sich von amtlichen Statistiken, insbesondere in Bezug auf Harmonisierung, Erfassungsbereich und Methodik.

➤ [www.destatis.de/exdat](http://www.destatis.de/exdat)

### Health at a Glance: Europe 2020

Mit Blick auf die derzeit rasche Ausbreitung des COVID-19-Virus waren viele Länder gezwungen, ihre Maßnahmen zur Eindämmung der Pandemie zu verschärfen. Jetzt kommt es entscheidend darauf an, dass die Regierungen eine effektive Strategie für eine zukünftige Lockerung der Maßnahmen erarbeiten, um weitere Lockdowns zu vermeiden. Zu einer solchen Strategie gehören deutlich effektivere Maßnahmen zur Durchführung von Tests, zur Kontaktnachverfolgung und zur Isolierung der Betroffenen sowie in Bezug auf die soziale Distanzierung.

Die Veröffentlichung „Health at a Glance: Europe 2020“ der Organisation für wirtschaftliche Zusammenarbeit und Entwicklung (OECD) vergleicht Gesundheitsfaktoren und Gesundheitsstrukturen in 36 europäischen Ländern. Im Fokus stehen Angebotsumfang und Zugänglichkeit der europäischen Gesundheitssysteme sowie Gesundheitsausgaben und die Verbreitung von Risikofaktoren in der Bevölkerung. Ein Sonderkapitel analysiert die Widerstandsfähigkeit der Gesundheitssysteme im Lichte der COVID-19-Pandemie.

➤ [www.oecd-ilibrary.org](http://www.oecd-ilibrary.org)

### OECD Digital Economy Outlook 2020

Die Covid-19-Krise hat die Nutzung digitaler Technik in Unternehmen und Privathaushalten verstärkt. Manche Internet-Provider berichten, dass ihr Datenverkehr seit Krisenbeginn um 60 % angestiegen ist. Gleichzeitig wird es immer wichtiger, der digitalen Kluft international und regional entgegenzuwirken. Der Glasfaseranteil an den Breitbandanschlüssen im Festnetz in OECD-Ländern beispielsweise reicht von unter 5 % in Deutschland, Österreich, Belgien und einigen weiteren Ländern bis zu 82 % in Südkorea.

Die dritte Ausgabe des „OECD Digital Economy Outlook“ bietet einen ganzheitlichen Überblick über konvergierende Trends, politische Entwicklungen und Daten sowohl auf der Angebots- als auch auf der Nachfrageseite der digitalen Wirtschaft. Der Bericht veranschaulicht, wie sich die digitale Transformation auf Wirtschaft und Gesellschaft auswirkt und zeigt zudem, wie die COVID-19-Pandemie sowohl Chancen als auch Herausforderungen des digitalen Wandels verstärkt.

↳ [www.oecd-ilibrary.org](http://www.oecd-ilibrary.org)

### OECD Pensions Outlook 2020

COVID-19 hat einen beispiellosen Schock für die Arbeitsmärkte und Rentenregelungen in der ganzen Welt ausgelöst. Viele Länder haben ihre Systeme zur Erhaltung von Arbeitsplätzen und der Arbeitslosenunterstützung ausgeweitet, sodass Arbeitnehmerinnen und Arbeitnehmer ihre erworbenen Ansprüche in den öffentlichen Rentensystemen und bis zu einem gewissen Grad auch in den Rentensparplänen behalten können.

Die aktuelle Ausgabe 2020 des OECD Pensions Outlook untersucht eine Reihe von politischen Optionen, die den Regierungen helfen sollen, die Nachhaltigkeit und Widerstandsfähigkeit der Rentensysteme zu verbessern. Der Bericht befasst sich unter anderem damit, wie sich zwischen den kurz- und langfristigen Folgen der politischen Reaktionen auf COVID-19 abwägen lässt, wie die Angemessenheit des Renteneinkommens bestimmt und bewertet werden kann oder auch wie kapitalgedeckte Rentenvereinbarungen Einzelpersonen in nicht standardisierten Arbeitssituationen dabei unterstützen können, für den Ruhestand vorzusorgen.

↳ [www.oecd-ilibrary.org](http://www.oecd-ilibrary.org)

## IN EIGENER SACHE

---

### Neues Erklärvideo zum Verbraucherpreisindex online

Jeden Monat veröffentlicht das Statistische Bundesamt den aktuellen Verbraucherpreisindex. Wie er entsteht, wie er mit der Inflationsrate zusammenhängt und was es mit Warenkorb und Wägungsschema auf sich hat erklärt das neue Video „Verbraucherpreisindex und Inflation kurz erklärt“:

↳ [www.destatis.de](http://www.destatis.de)

## AUS EUROPA

---

### 44. Sitzung des AESS

In der Sitzung des Ausschusses für das Europäische Statistische System (AESS) am 8./9. Oktober 2020 informierte die deutsche EU-Ratspräsidentschaft über die beiden geplanten Konferenzen sowie über den Sachstand der gegenwärtigen Rechtsetzungsverfahren in der Ratsarbeitsgruppe Statistik. Mittlerweile wurde am 25./26. November 2020 auf Ebene der Generaldirektorinnen und Generaldirektoren das „High-Level Meeting on ESS communication and governance in a time of crisis“ als virtuelle Veranstaltung durchgeführt. Am 27. November 2020 fand eine digitale Konferenz statt, an der auch Vertreterinnen und Vertreter aus der Wissenschaft teilnahmen. Unter dem Titel „Lessons Learned from COVID-19 – Short-term Data Requirements for Long-term Decision-making“ wurden insbesondere Fragen zu den Themen Wirtschaft und Finanzen vor dem Hintergrund der Corona-Pandemie diskutiert. Begleitend dazu wurden hierfür auf der Website der deutschen EU-Ratspräsidentschaft im Bereich Statistik ([www.destatis.de/eu2020](http://www.destatis.de/eu2020)) Beiträge in unterschiedlichen Formaten bereitgestellt. Eine Podcast-Reihe beleuchtet den Einfluss der Corona-Pandemie auf verschiedene gesellschaftliche Aspekte. Eine interaktive Europakarte informiert über neu entwickelte Produkte und Projekte, mit denen die europäischen Nationalen Statistischen Ämter den Herausforderungen der Corona-Pandemie begegnen.

Die Kommission wird eine mögliche EU-Verordnung für einen Rahmen bezüglich der Etablierung und Funktionsweise gemeinsamer Europäischer Datenräume (Common European Data Spaces) entwerfen. Zentrale Elemente dieses Entwurfs sollen die Rolle eines „Data intermediaries“ (organisiert den Datenaustausch) und die Einrichtung eines „Data Innovation Board“ (unterstützt die Kommission bei Einrichtung der Governance-Strukturen) sein.

Eurostat legte dem AESS eine Rahmenverordnung über Statistiken zum landwirtschaftlichen Input und Output (SAIO = Statistics on Agricultural Input and Output) vor, die einzelstatistische Aspekte der Landwirtschaft in einer Verordnung subsumieren soll.

Noch während der deutschen EU-Ratspräsidentschaft sollten eine erste Version eines ESS Recovery Dashboards mit ausgewählten Covid-19 relevanten Indikatoren erstellt, ein Krisenprotokoll abgestimmt und ein Krisenstab für das Europäische Statistische System (ESS) etabliert werden. Ziel des Krisenprotokolls für das ESS ist es, Abstimmungsprozesse zielgerichtet und effizient zu organisieren. Es umfasst beispielsweise die Koordinierung und Abstimmung im ESS, Maßnahmen für die Statistikproduktion sowie die interne und externe Kommunikation. Das Krisenprotokoll wurde unter dem Namen „Wiesbaden Memorandum“ auf dem High-Level Meeting verabschiedet.

Mit dem von der EU-Kommission vereinbarten Green Deal wird es weiteren Bedarf an statistischen Daten geben, insbesondere unter den Aspekten Aktualität und regionale Gliederung, wobei allerdings die Nationalen Statistikämter nicht weiter belastet werden sollen. So soll das Monitoring möglichst auf bestehenden Daten aufbauen und ein detaillierter Aktionsplan dem AESS im Frühjahr 2021 vorgelegt werden. Zur wirksamen Messung der Green-Deal-Maßnahmen bestehen allerdings noch Datenlücken, zudem sollte zur Vermeidung einer weiter steigenden Belastung für die Auskunftspflichtigen und die Nationalen Statistikämter der Ansatz „One in, one out“ konsequent angewendet werden.

Zum Jahr 2025 planen die Vereinten Nationen, das Regelwerk zu den Volkswirtschaftlichen Gesamtrechnungen (System of National Accounts – SNA) und zur Zahlungsbilanzstatistik (Balance of Payments Manual – BPM) zu überarbeiten. Die stetige Weiterentwicklung der Lebensverhältnisse der Menschen muss sich stets mit

der Berechnung der Wirtschaftsleistung einer Volkswirtschaft decken. Wichtige Themen sind unter anderem Digitalisierung, Gutes Leben und Nachhaltigkeit. Sukzessive werden danach mit dem Europäischen System Volkswirtschaftlicher Gesamtrechnungen (ESVG) auch die europäischen Regelungen angepasst.

Eurostat informierte über den Sachstand der Reform der Internationalen Standardklassifikation der Wirtschaftszweige der Vereinten Nationen (ISIC) zur Gliederung von Wirtschaftsbereichen und Industriezweigen. Eine Arbeitsgruppe arbeitet an konkreten Vorschlägen, die in die Arbeiten der Vereinten Nationen einfließen sollen. Ziel ist, einen ersten Entwurf auf der jährlichen Sitzung der Statistischen Kommission der Vereinten Nationen im März 2021 vorzulegen. Die ISIC stellt eine wichtige Grundlage für das SNA dar, daher ist eine Umsetzung der geplanten Reformen vor Beginn der SNA-Revision wünschenswert.

## Konferenz NTTS 2021

Die Konferenz Neue Techniken und Technologien für die Statistik (NTTS 2021) wird vom 9. bis 11. März 2021 sowie mit zusätzlichen Satellitenveranstaltungen (beispielsweise einem Big Hackathon) am 8. und 12. März 2021 in Brüssel stattfinden. Ziel ist es, laufende Forschungs- und Innovationsprojekte in der amtlichen Statistik vorzustellen und die Vorbereitung neuer innovativer Projekte zu fördern und zu erleichtern. Folgende Themenblöcke sind geplant:

- › Vertrauenswürdige intelligente Statistik (insbesondere neue Datenquellen, Web Intelligence, Daten mobiler Geräte, intelligente und vertrauenswürdige Umfragen, Dateninnovation sowie Technologien zur Wahrung der Privatsphäre)
- › Rahmenregelungen, Software und Tools (Replizierbarkeit und Reproduzierbarkeit, Verwendung von R in offiziellen Statistiken, neue statistische Tools, Open Source- und Sharing-Codes, Unternehmensarchitektur und -standards sowie Datenarchitektur)
- › Wiederverwendung und Weitergabe von Daten (Datenplattformen, statistische Offenlegungskontrolle, Vertraulichkeit und Datenschutz sowie technische und rechtliche Aspekte im Zusammenhang mit der Verwendung vertraulicher Daten amtlicher Statistiken)

- › Kommunikation und Verbreitung (Visualisierungen, Crowdsourcing- und Bürgerstatistiken, Daten-Storytelling, Methoden zur Erfassung von Benutzereingaben sowie die Bewertung der Benutzerbedürfnisse und -zufriedenheit)

Das Statistische Bundesamt wird sich sowohl an der Konferenz als auch am Big Hackathon aktiv mit Beiträgen einbringen.

## NEUERSCHEINUNGEN

---

### Bildungsfinanzbericht 2020 erschienen

Der Bildungsfinanzbericht fasst die wichtigsten verfügbaren Informationen zu den Bildungsausgaben zusammen. Er ist Teil der Bildungsberichterstattung, die kontinuierlich datengestützte Informationen über Rahmenbedingungen, Input, Verläufe, Ergebnisse und Wirkungen von Bildungsprozessen bereitstellt. Den Bildungsfinanzbericht erstellt das Statistische Bundesamt jährlich im Auftrag des Bundesministeriums für Bildung und Forschung und der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland.

↳ [www.destatis.de](http://www.destatis.de)

### Kulturfinanzbericht 2020 erschienen

Der Kulturfinanzbericht ist eine Gemeinschaftsveröffentlichung der Statistischen Ämter des Bundes und der Länder. Er erscheint alle zwei Jahre im Auftrag der Kultusministerkonferenz und der Beauftragten der Bundesregierung für Kultur und Medien im Einvernehmen mit dem Deutschen Städtetag.

Der aktuelle Kulturfinanzbericht 2020 bietet einen breiten Überblick über die öffentliche Kulturfinanzierung in Deutschland sowie darüber, in welchem Umfang die privaten Haushalte Ausgaben für die kulturelle Teilhabe aufwenden. Das zentrale Berichtsjahr ist das Haushaltsjahr 2017, Haushaltsplanungen bis zum Jahr 2020 ergänzen die Zeitreihe bis zum aktuellen Rand. Der Bericht bietet damit eine Datengrundlage für Politik, Kul-

turinstitutionen, Kultusverwaltungen, Wissenschaft und die interessierte Öffentlichkeit.

↳ [www.destatis.de](http://www.destatis.de)

### Jahresgutachten des Sachverständigenrats

Der Sachverständigenrat zur Begutachtung der gesamtwirtschaftlichen Entwicklung legte im November 2020 sein Jahresgutachten 2020/21 vor. Es trägt den Titel „Corona-Krise gemeinsam bewältigen, Resilienz und Wachstum stärken“. Das Jahresgutachten enthält auch den jährlichen nationalen Produktivitätsbericht, für den der Sachverständigenrat als nationaler Ausschuss für Produktivität die Faktoren eines anhaltenden Produktivitätswachstums sowie die Voraussetzung zur Steigerung von Wettbewerbsfähigkeit und Produktivität analysiert.

↳ [www.sachverstaendigenrat-wirtschaft.de](http://www.sachverstaendigenrat-wirtschaft.de)

### OECD-Wirtschaftsberichte: Deutschland 2020

Die deutsche Wirtschaft ist 2020 aufgrund der Corona-Pandemie in eine tiefe Rezession geraten. Durch entschlossenes staatliches Handeln gelang es zwar, die Kapazitäten des Gesundheitssystems zu stärken und zugleich Arbeitsplätze und Unternehmen zu schützen. Auch wurden die Investitionen zur Bewältigung der strukturellen Herausforderungen aufgestockt, die sich aus der Energiewende und der digitalen Transformation ergeben. Es muss aber noch mehr getan werden, um den Investitionsstau im Infrastrukturbereich aufzulösen.

„OECD-Wirtschaftsberichte: Deutschland 2020“ untersucht die großen Herausforderungen, denen sich Deutschland stellen muss, beurteilt die kurzfristigen Aussichten und formuliert spezifische Politikempfehlungen. In der aktuellen Ausgabe geht der Bericht der Frage nach, wie die Vorteile der digitalen Transformation voll ausgeschöpft werden können.

↳ [www.oecd-ilibrary.org](http://www.oecd-ilibrary.org)



### Jenseits des BIP: Was bei der wirtschaftlichen und sozialen Entwicklung wirklich zählt

Die globale Wirtschaftskrise von 2008 hat in vielen Ländern wirtschaftlich und gesellschaftlich tiefe Spuren hinterlassen. Hätten jenseits des Bruttoinlandsprodukts umfassendere Daten zu Wohlbefinden und sozialer Lage zur Verfügung gestanden, hätte die Politik wohl entschlossener auf die Krise reagiert. Tatsächlich müssen neben dem Bruttoinlandsprodukt weitere Kenngrößen entwickelt werden, um das allgemeine Wohlergehen von Bevölkerungen zu erfassen.

Der OECD-Bericht „Beyond GDP“ aus dem Herbst 2018 ist jetzt auf Deutsch erschienen. Darin argumentieren die Co-Vorsitzenden der der OECD angegliederten Hochrangigen Sachverständigengruppe zur Messung von wirtschaftlicher Leistung und sozialem Fortschritt, Joseph E. Stiglitz, Jean-Paul Fitoussi und Martine Durand, dass Indikatoren-Dashboards zu entwickeln sind, die abbilden, was wirklich zählt: wer vom Wachstum profitiert, ob dieses Wachstum ökologisch nachhaltig ist, wie die Menschen ihr Leben empfinden und welche Faktoren für den Erfolg eines Menschen oder eines Landes ausschlaggebend sind.

↳ [www.oecd-ilibrary.org](http://www.oecd-ilibrary.org)

### Atlas of Sustainable Development Goals 2020 online

Die 17 Ziele für nachhaltige Entwicklung (Sustainable Development Goals – SDGs) sind politische Zielsetzungen der Vereinten Nationen. Sie sollen weltweit der Sicherung einer nachhaltigen Entwicklung auf ökonomischer, sozialer sowie ökologischer Ebene dienen.

Auf Basis der Datenbank World Development Indicators der Weltbank sowie einer Vielzahl relevanter Datenquellen von Wissenschaftlerinnen und Wissenschaftlern und anderen Forschenden weltweit wird der Atlas of Sustainable Development Goals erstellt. Der Atlas 2020 enthält interaktives Storytelling und Datenvisualisierungen zu den 17 Zielen für nachhaltige Entwicklung. Es werden Trends für ausgewählte Ziele innerhalb jedes Ziels hervorgehoben und Konzepte zur Messung einiger

SDGs vorgestellt. Wenn Daten verfügbar sind, werden auch die sich abzeichnenden Auswirkungen der COVID-19-Pandemie auf die SDGs hervorgehoben. Die mittlerweile dritte Ausgabe der Atlasreihe soll die Leserinnen und Leser ansprechend informieren und inspirieren, Fortschritte bei der Erreichung der SDGs zu entdecken, zu verstehen und zu visualisieren.

↳ <https://datatopics.worldbank.org/sdgate>

---

# DEUTSCHE EU-RATSPRÄSIDENTSCHAFT IM BEREICH STATISTIK

---

↳ **Schlüsselwörter:** EU-Ratspräsidentschaft – Ratsarbeitsgruppe Statistik – ESS Recovery Dashboard for Europe – Wiesbaden Memorandum – Digitale Konferenz

## ZUSAMMENFASSUNG

Im Rahmen der deutschen EU-Ratspräsidentschaft hat das Statistische Bundesamt seit dem 1. Juli 2020 für ein halbes Jahr den Vorsitz in der Ratsarbeitsgruppe Statistik inne, bereitet deren Sitzungen vor und leitet sie. Darüber hinaus verfolgt das Statistische Bundesamt das Ziel, das europäische Statistikprogramm aktiv mitzugestalten. Dieser Beitrag informiert über den aktuellen Stand der Vorhaben im Rahmen der deutschen EU-Ratspräsidentschaft 2020 im Bereich Statistik und gibt einen Ausblick auf die geplanten Inhalte und Veranstaltungen bis zum Jahresende.

↳ **Keywords:** EU Council Presidency – Council Working Party on Statistics – ESS Recovery Dashboard for Europe – Wiesbaden Memorandum – digital conference

## ABSTRACT

*As part of the German EU Council Presidency, the Federal Statistical Office has chairmanship of the Council Working Party on Statistics for six months from 1 July 2020, i.e. it prepares the meetings and chairs them. In addition to that, the Federal Statistical Office intends to play an active part in shaping the European Statistics Programme. This article provides information on the current status of the projects carried out during the German EU Council Presidency 2020 in the field of statistics and presents an outlook on planned content and events until the end of the year.*

### EU-Gesetzgebungsverfahren

---

Die Ratsarbeitsgruppe Statistik ist das zentrale Gremium für die Beratung von EU-Rechtsetzungsakten auf dem Gebiet der Statistik. Den Vorsitz in der Ratsarbeitsgruppe Statistik während der deutschen EU-Ratspräsidentschaft hat der Präsident des Statistischen Bundesamtes, Dr. Georg Thiel, inne. Der Vorsitzende ist dafür verantwortlich, die Beratungen des Rates über EU-Rechtsvorschriften voranzubringen.

Während der deutschen EU-Ratspräsidentschaft hat das Präsidentschaftsteam im Statistischen Bundesamt zwei Gesetzgebungsverfahren mit Bezug zur amtlichen Statistik betreut: Die Verordnung des Europäischen Parlaments und des Rates zur Aufstellung des Programms über den Binnenmarkt, die Wettbewerbsfähigkeit der Unternehmen, einschließlich der kleinen und mittleren Unternehmen, und die europäischen Statistiken (COM[2018]441) (kurz: Binnenmarktprogramm) legt den Finanzrahmen für den Binnenmarkt der Europäischen Union (EU) von 2021 bis 2027 fest. Für diesen Zeitraum ist erstmals auch das Europäische Statistische Programm im Binnenmarktprogramm enthalten. Das Dossier wurde in der Ratsarbeitsgruppe Wettbewerbsfähigkeit und Wachstum verhandelt. Die Ratsarbeitsgruppe Statistik und die deutsche EU-Ratspräsidentschaft im Bereich Statistik standen ihr beratend zur Seite. Aktuell werden die Verhandlungen mit dem Europäischen Parlament zu den statistischen Aspekten des Binnenmarktprogramms intensiv begleitet. Zum Europäischen Statistischen Programm innerhalb des Binnenmarktprogramms konnte die Ratspräsidentschaft mit dem Europäischen Parlament sowie der Europäischen Kommission eine Einigung erzielen.

Belange der amtlichen Statistik umfasst auch die Verordnung über Privatsphäre und elektrische Kommunikation (COM[2017]010), die sich aktuell in der Abstimmung zwischen den Mitgliedstaaten in der Ratsarbeitsgruppe Telekommunikation und Informationsgesellschaft befindet. Auch hier haben die Ratsarbeitsgruppe Statistik und die deutsche EU-Ratspräsidentschaft im Bereich Statistik die Interessen zur Weiterentwicklung der amtlichen Statistik aktiv in die Verhandlungen eingebracht.

### EU-Ratspräsidentschaft vor dem Hintergrund der Corona-Pandemie

---

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. Die Corona-Pandemie stellt eine außerordentliche Herausforderung für die EU dar und erzeugt neue Informationsbedarfe. Dies betrifft etwa Daten aus den Bereichen Gesundheit, Wirtschaft, Finanzen und Gesellschaft. Die Relevanz der amtlichen Statistik für nachhaltige und zeitnahe Entscheidungsfindung ist durch die Krise noch stärker als bisher in den Fokus gerückt.<sup>1</sup> Die Erkenntnisse aus der Krise bildeten daher die Grundlage für die Schwerpunkte und das Programm der deutschen EU-Ratspräsidentschaft im Bereich Statistik.

### ESS Recovery Dashboard for Europe

---

In ihrem Arbeitsprogramm hat sich die deutsche EU-Ratspräsidentschaft im Bereich Statistik zum Ziel gesetzt, einen koordinierten Ansatz im Europäischen Statistischen System (ESS) voranzubringen, um die Bereitstellung verlässlicher Statistiken von hoher Priorität sicherzustellen. Zusammen mit dem Statistischen Amt der Europäischen Union (Eurostat) entwickelt das Statistische Bundesamt deshalb derzeit das „ESS Recovery Dashboard for Europe“. Dieses soll die wirtschaftlichen und sozialen Entwicklungen infolge der Corona-Pandemie in den Mitgliedstaaten und der EU anwenderorientiert zeigen. Es bietet die Chance, ein ganzheitliches Bild von Wirtschaft und Gesellschaft darzustellen und ihre Entwicklungen im Laufe der Zeit zeitnah abzubilden. Informationsbedarfe können so komprimiert und adressatenorientiert gedeckt werden. Die Veröffentlichung des Dashboards mit einer ersten Auswahl von rund 20 Indikatoren erfolgte Mitte Dezember 2020, zusätzliche Indikatoren sollen das Dashboard später ergänzen.

---

1 Der in WISTA 4/2020 erschienene Artikel [„Das Europäische Statistische System als Krisenmanager – verlässliche Daten für Europa“](#) fasst zusammen, welche Lösungen die europäischen Staaten für die Herausforderungen der Corona-Pandemie entwickelt haben und wie sich die Zusammenarbeit im Europäischen Statistischen System angesichts der Krise gestaltet.

### Wiesbaden Memorandum

---

Ebenfalls als Teil des Arbeitsprogramms der deutschen EU-Ratspräsidentschaft im Bereich Statistik hat das Statistische Bundesamt zusammen mit Eurostat ein Krisenprotokoll für das ESS erarbeitet. Ziel eines solchen Krisenprotokolls ist es, Abstimmungsprozesse zwischen den nationalen Statistikämtern und Eurostat zielgerichtet und effizient zu organisieren. Auch in künftigen Krisensituationen soll das ESS schnell handlungsfähig sein. Hierfür ist sicherzustellen, dass alle Beteiligten genau wissen, was wann zu tun ist. Dies wird mit dem Krisenprotokoll für das ESS erreicht. Die Generaldirektorinnen und Generaldirektoren des ESS verabschiedeten das Krisenprotokoll beim High-level-Meeting im November 2020 als Wiesbaden Memorandum.

### High-level-Meeting der Generaldirektorinnen und Generaldirektoren des ESS

---

Am 25. und 26. November 2020 fand das „High-level meeting on ESS communication and governance in a time of crisis“ der Generaldirektorinnen und Generaldirektoren der Statistischen Ämter des ESS statt. Die ursprünglich in Wiesbaden geplante Veranstaltung wurde aufgrund der Corona-Pandemie virtuell durchgeführt. Die Generaldirektorinnen und Generaldirektoren tauschten sich zu Erfahrungen und Lehren aus der Corona-Pandemie in den Bereichen Governance, Kommunikation, Datenerhebung und -produktion, Qualität sowie Innovation aus. Beim Thema Kommunikation wurden insbesondere Herausforderungen und Chancen erörtert, die sich für das ESS aus der gegenwärtigen Krise ergeben, etwa hinsichtlich der Stärkung der Relevanz und der Marke „Amtliche Statistik“, der Möglichkeit von Innovationsentwicklungen in und durch die Krise und der Schaffung neuer strategischer Partnerschaften, vor allem mit Blick auf neue digitale Daten.

### Digitale Konferenz „Lessons Learned from COVID-19“

---


Unter dem Titel „Lessons learned from COVID-19“ richtete das Statistische Bundesamt eine Digitale Konferenz aus. Mittels verschiedener Formate wurde in einem europäischen Kontext aufgezeigt, welche Informationen die amtliche Statistik mit ihren Partnern als Grundlage für Entscheidungsfindungen auch in Krisenzeiten liefert. Darüber hinaus wurde für verschiedene Fachbereiche erläutert, welche weiteren Datenbedarfe in der Krise bestehen. Die Erfahrungen mit den Informationen der amtlichen Statistik während der Krise aus Nutzersicht wurden dabei ebenso in den Blick genommen wie die Perspektive der Statistikproduzenten. Eine Podcast-Reihe beleuchtete den Einfluss der Corona-Pandemie auf verschiedene gesellschaftliche Aspekte. Eine interaktive Europakarte mit Video-, Audio- und Textbeiträgen bietet Informationen zu den Erfahrungen der europäischen Nationalen Statistischen Ämter im Kontext der Corona-Pandemie.

Den Abschluss der Digitalen Konferenz bildete eine virtuelle Podiumsdiskussion am 27. November 2020. Unter dem Titel „Lessons Learned from COVID-19 – Short-term Data Requirements for Long-term Decision-making“ wurden insbesondere Fragen zu den Themen Wirtschaft und Finanzen vor dem Hintergrund der Corona-Pandemie diskutiert. Staatssekretär Dr. Markus Richter (Bundesministerium des Innern, für Bau und Heimat) eröffnete die Digitale Konferenz mit einer Keynote über die nationalen Lehren aus der Pandemie und innovative Informationsangebote. Die Generaldirektorin von Eurostat, Dr. Mariana Kotzeva, setzte in ihrer Keynote den europäischen Impuls zum Auftakt der Veranstaltung. Unter der Moderation von Stefan Schweinfest (Direktor der Statistischen Kommission der Vereinten Nationen) nahmen Prof. Dr. Isabel Schnabel (Mitglied des Executive Board der Europäischen Zentralbank), Prof. Dr. Dr. h.c. Lars P. Feld (Vorsitzender des Sachverständigenrats zur Begutachtung der gesamtwirtschaftlichen Entwicklung), Dominik Rozkrut (Präsident von Statistics Poland), Prof. Dr. Agnès Bénassy-Quéré (Chefökonomin des französischen Finanzministeriums) und Stefano Palmieri (Vorsitzender der Fachgruppe Wirtschafts- und Währungsunion, wirtschaftlicher und sozialer Zusammenhalt [ECO] im Europäischen Wirtschafts- und Sozialausschuss) an der Podiumsdiskussion teil. Die Teilnehmenden legten

dar, wie sich die Datenbedarfe in der Pandemie geändert haben, beispielsweise, dass Daten schneller aktualisiert werden müssten, um die Auswirkungen der Pandemie akkurat einschätzen zu können. Daraufhin wurde das Spannungsfeld Aktualität und Genauigkeit in der Debatte von verschiedenen Blickwinkeln beleuchtet. Ein wichtiger Teil der Debatte war, neben der Digitalisierung auch die Belastung der Befragten nicht aus den Augen zu verlieren. Dr. Georg Thiel beendete die Podiumsdiskussion und die Digitale Konferenz mit einem Fazit zur Arbeit der amtlichen Statistik.

### Weitere Informationen

---

Die eigens eingerichtete Internetseite [www.destatis.de/eu2020](http://www.destatis.de/eu2020) stellt alle Informationen zur deutschen EU-Ratspräsidentschaft im Bereich Statistik auch über das Ende der Ratspräsidentschaft hinaus bereit. Dort stehen auch die Beiträge zur Digitalen Konferenz „Lessons Learned from COVID-19“ zur Verfügung. 

# VERLEIHUNG DES GERHARD-FÜRST-PREISES 2020

Prof. Dr. Walter Krämer

↘ **Schlüsselwörter:** Gerhard-Fürst-Preis – amtliche Statistik – Nachwuchspreis – Straßensensordaten – Small-Area-Verfahren – Machine-Learning-Algorithmen – Datenfusion

## ZUSAMMENFASSUNG

Mit dem Gerhard-Fürst-Preis des Statistischen Bundesamtes werden Arbeiten in den Kategorien Dissertationen und Master- beziehungsweise Bachelorarbeiten ausgezeichnet, die theoretische Themen mit einem engen Bezug zum Aufgabenspektrum der amtlichen Statistik behandeln oder empirische Fragestellungen unter intensiver Nutzung von Daten der amtlichen Statistik untersuchen.

Das Statistische Bundesamt möchte mit diesen jährlichen wissenschaftlichen Auszeichnungen die Zusammenarbeit zwischen Wissenschaft und amtlicher Statistik weiter intensivieren. Zugleich soll der Preis junge Wissenschaftlerinnen und Wissenschaftler ermutigen, das vielfältige Datenangebot der amtlichen Statistik für ihre empirischen Forschungen ausgiebig zu nutzen.

↘ **Keywords:** Gerhard Fürst Award – official statistics – prize for young researchers – road sensors – small area estimation – machine learning algorithms – data fusion

## ABSTRACT

*The Federal Statistical Office grants the Gerhard Fürst Awards to doctoral, Master's and Bachelor's theses which deal with theoretical topics with a close link to the scope of duties of official statistics, or examine empirical questions with the intense use of data from official statistics.*

*The annual scientific awards of the Federal Statistical Office are aimed at intensifying the cooperation between the scientific community and the official statistical agencies. At the same time, the awards are intended to encourage junior scientists to make extensive use in their empirical research of the wide variety of data offered by official statistics.*



**Prof. Dr. Walter Krämer**

ist Direktor des Sonderforschungsbereichs „Statistical modelling of nonlinear dynamic processes“ an der Fakultät Statistik der Technischen Universität Dortmund. Seit dem Jahr 2014 ist er Vorsitzender des unabhängigen Gutachtergremiums, das die wissenschaftlichen Arbeiten zur Auszeichnung mit dem Gerhard-Fürst-Preis des Statistischen Bundesamtes empfiehlt.

## Einleitung

---

Auf Empfehlung eines [unabhängigen Gutachtergremiums](#) zeichnet das Statistische Bundesamt seit 1999 jährlich herausragende wissenschaftliche Arbeiten mit einem engen Bezug zur amtlichen Statistik mit dem [Gerhard-Fürst-Preis](#) aus. Für den Gerhard-Fürst-Preis 2020 hat das Gutachtergremium sowohl zwei Arbeiten in der Kategorie „Dissertationen“ als auch zwei Arbeiten in der Kategorie „Master-/Bachelorarbeiten“ als gleichermaßen herausragend und preiswürdig erachtet. Das Statistische Bundesamt hat am 19. November 2020 somit vier wissenschaftliche Arbeiten mit dem Gerhard-Fürst-Preis 2020 prämiert:

- › Für seine an der Universität Duisburg-Essen bei Professor Dr. Rainer Schnell entstandene Dissertation zum Thema „Correcting Survey Measurement Error With Big Data from Road Sensors Through Capture-recapture“ wurde Dr. Jonas Klingwort ausgezeichnet. Die theoretische Fundierung und erstmalige Verwendung von Straßensensordaten zur Korrektur von Befragungsdaten sahen die Gutachterinnen und Gutachter als besonders förderwürdig an.
- › Professorin Dr. Natalia Rojas-Perilla erhielt die Ehrung für ihre Dissertation „The Use of Data-driven Transformations and Their Applicability in Small Area Estimation“. Ihr Forschungsgebiet im Bereich der Analyse von Methoden der Small-Area-Verfahren hat große praktische Relevanz für statistische Ämter. Die Arbeit entstand an der Freien Universität Berlin bei Professor Dr. Timo Schmid.
- › In der Kategorie „Master-/Bachelorarbeiten“ wurde Daniel Haake für seine Masterarbeit zum Thema „Prognose von Wohnungseinbrüchen mit Hilfe von Machine-Learning-Algorithmen“ der Gerhard-Fürst-Preis 2020 zuerkannt. Diese Arbeit zeigt, dass mit dem ihr zugrundeliegenden Wahrscheinlichkeitsansatz künftige Wohnungseinbrüche innerhalb einer kürzeren Zeit prognostiziert werden können. Dies stellt eine wesentliche Verbesserung gegenüber aktuellen Ansätzen dar. Herr Haake verfasste seine Arbeit an der Hochschule Albstadt-Sigmaringen bei Professor Dr. Andreas Knoblauch.
- › Jannik Schaller erhielt den Gerhard-Fürst-Preis für seine Masterarbeit „Datenfusion von EU-SILC und HBS: Vergleich zwischen Random Hot-Deck und

Predictive Mean Matching im Rahmen einer Simulationsstudie“. Das in der Arbeit verwendete optimierte Datenfusionsverfahren konnte den Informationsgehalt der fusionierten Datenfiles deutlich verbessern. Die Arbeit entstand bei Professor Dr. Martin Messingschläger an der Otto-Friedrich-Universität Bamberg.

Die Preisverleihung konnte in diesem Jahr aufgrund der Corona-Pandemie nur virtuell stattfinden. Für das Statistische Bundesamt überreichte Dr. Daniel Vorgrimler, Leiter der Abteilung „Strategie und Planung, Internationale Beziehungen, Forschung und Kommunikation“ und Schriftleiter dieser Zeitschrift, der Preisträgerin und den Preisträgern ihre Urkunden. Die folgenden Laudationes auf die prämierten Arbeiten hielt der Vorsitzende des Gutachtergremiums, Professor Dr. Walter Krämer (Technische Universität Dortmund). Die Preisträgerin und die Preisträger berichten im Jahr 2021 in dieser Zeitschrift ausführlich über ihre Arbeiten.

Die [Kurzfassungen](#) der prämierten Arbeiten sowie weitere Details zur Verleihung des Gerhard-Fürst-Preises sind im Internetangebot des Statistischen Bundesamtes zu finden.

## Laudationes Gerhard-Fürst-Preis 2020

---

Liebe Preisträger, liebe Zuschauer-Gäste,

auch für mich als Vorsitzender der Jury ist es immer wieder eine Ehre, und in gewisser Weise auch eine Herausforderung, die jährlichen Gewinner des Gerhard-Fürst-Preises einer breiteren Öffentlichkeit vorzustellen. Im Moment ist diese Öffentlichkeit zwar auf die Teilnehmer dieser Videokonferenz beschränkt, aber die Laudationes werden ja dann auch in Wirtschaft und Statistik publiziert, sodass sich jeder einen Eindruck machen kann, was unser wissenschaftlicher Statistik-Nachwuchs in Bezug auf die Amtsstatistik für tolle neue Ideen hat.

Daran gab es auch dieses Jahr keinen Mangel. Die Jury hatte quasi ein Luxusproblem, was sich auch darin äußert, dass wir in beiden Kategorien – Bachelor-/Master- auf der einen und Doktorarbeiten auf der anderen Seite – jeweils zwei Arbeiten ausgezeichnet haben.

Bei den Bachelor-/Masterarbeiten haben uns die von Jannik Schaller aus Bamberg und Daniel Haake von der Hochschule Albstadt-Sigmaringen am besten gefallen.

Dass beide von ihren jeweiligen Hochschulen mit guten Noten ausgestattet worden sind, versteht sich fast von selbst. Aber bei beiden kommt noch ein gewisses Extra hinzu. Bei Herrn Haake ist das einmal die ungewöhnliche Art der Daten, die er untersucht: Und zwar handelt es sich um Wohnungseinbrüche, gemeldet an und gesammelt von dem Landeskriminalamt Baden-Württemberg für die Jahre 2010 bis 2018. Mithilfe dieser Daten und verschiedener Methoden des maschinellen Lernens gelingt es Herrn Haake, ein bisher bereits von der Polizei genutztes Prognoseverfahren für Wohnungseinbrüche deutlich zu verbessern. Wenn Sie also heute Abend nach Hause kommen und sehen: Aha, es wurde nicht eingebrochen, dann haben Sie das zum Teil also auch Herrn Haake zu verdanken. Aber Scherz beiseite: Gerade bei Einbruchsdelikten ist Prävention das A und O und hier sind Beiträge wie die von Herrn Haake höchst willkommen. Er war übrigens, und das ist aus meiner Sicht das zweite kleine Extra, selbst lange Zeit im Polizeidienst aktiv und weiß, wovon er redet. Praxisrelevanter jedenfalls geht es nicht.

Kein Mangel an Praxisrelevanz herrscht auch bei der zweiten von uns ausgezeichneten Masterarbeit von Janik Schaller mit dem Titel „Vergleich zwischen Random Hot-Deck und Predictive Mean Matching im Rahmen einer Simulationsstudie“. Hier geht es darum, wie man amtliche Daten zum gleichen Thema, aber aus verschiedenen Quellen, auf möglichst effiziente Weise zusammenführt und kombiniert. Konkret geht es hier um Daten zu Einkommen und Konsum. Da gibt es einmal die groß angelegte EU-weite sogenannte EU-SILC-Erhebung (für „European Union Statistics on Income and Living Conditions“) und auf der anderen Seite den sogenannten Household Budget Survey (HBS). Das sind periodische nationale und dann von Eurostat zusammengefasste Erhebungen betreffend die Ausgaben der privaten Haushalte für Waren und Dienstleistungen, die beiden letzten Erhebungen gab es 2010 und 2015. Beide Erhebungen, SILC und HBS, erheben auch Daten, die nur in einer der beiden vorkommen. Wirft man also die Stichproben zusammen, gibt es viele Lücken, und wie man die am besten füllt, wird von Herrn Schaller mittels sorgfältiger Simulationen untersucht. Die Amtsstatistik darf sich freuen.

Auch der Nutzen der beiden von uns ausgezeichneten Doktorarbeiten für die Amtsstatistik liegt auf der Hand. Sozusagen unmittelbar ins Herz dieser Amtsstatistik

zielt etwa die Doktorarbeit von Jonas Klingwort zum Thema „Correcting Survey Measurement Error With Big Data from Road Sensors Through Capture-recapture.“ Hier adressiert Herr Klingwort ein heißes Problem der aktuellen Amtsstatistik, nämlich inwieweit und unter welchen Voraussetzungen nicht amtliche Ablesedaten die Ergebnisse der Amtsstatistik verbessern könnten. Konkret geht es hier um den kommerziellen Güterverkehr auf dem holländischen Straßennetz: Wie viele Tonnenkilometer Bewegungen finden dort jährlich statt? Dazu macht man in den Niederlanden traditionell Umfragen bei Spediteuren, aber seit einigen Jahren sind auf holländischen Straßen auch Sensoren eingebaut, die darüberfahrende Lastwagen zusammen mit den Nutzlasten erfassen. Wenn man auf diese Weise in einem gegebenen Zeitraum eine Teilmenge aller Nutzfahrzeuge zusammen mit ihren Lasten erfasst, kann man aus deren Anteil in einer späteren Stichprobe auf die gesamten Tonnenkilometer rückschließen. Diese sogenannte Capture-Recapture-Methode ist über 200 Jahre alt und wurde bislang vor allem in der Biometrie zur Schätzung von Populationsgrößen entwickelt: Man markiert 1 000 Fische, rührt die gut auf den Weltmeeren um, und wenn man dann später bei einem Prozent einer Zufallsstichprobe die Markierung findet, hat man die Schätzung 100 000 für den gesamten Fischbestand. Das Pro und Kontra dieser Methode und ihre Verwendung in allen möglichen sozialen und naturwissenschaftlichen Kontexten wird von Herrn Klingwort sehr verdienstvoll ausgeführt, zusammen mit einer ausführlichen Diskussion zum Pro und Kontra von Big Data in der Amtsstatistik.

Die zweite von uns ausgezeichnete Doktorarbeit, die von Natalia Rojas-Perilla, adressiert einmal die Frage, wie, wann und warum Dateninputs für statistische Prozeduren zunächst einer Transformation unterzogen werden müssen, um gewünschte Eigenschaften wie Normalität oder Homoskedastie zu erzeugen. Das geschieht zunächst noch ohne Bezug auf spätere Anwendungen in der Amtsstatistik. Sehr verdienstvoll ist auch die Einbettung dieser Methoden in ein R-Paket. Eine Anwendung auf die Amtsstatistik erfolgt dann im zweiten Teil der Arbeit, wo diese Methoden für Small Area Estimation Einsatz finden, wie sie in der Amtsstatistik in den letzten Jahren immer wichtiger geworden ist. Konkrete Anwendung ist dabei die kleinräumige Erfassung von Indikatoren der Armut und der Ungleichheit. Der dritte Teil der Arbeit wendet sich einem nochmals spezielle-



ren Thema zu, nämlich der Frage, wie Zähldaten geeignet transformiert werden können. Insbesondere geht es hier darum, ob man von vornherein mit einer Poisson-Regression arbeiten oder zunächst die abhängige Variable transformieren und dann mit normaler Regression weitermachen soll, und immer mit einem Auge auf die Auswirkungen auf die Amtsstatistik. Also genau das, wofür der Gerhard-Fürst-Preis ursprünglich ins Leben gerufen worden ist.

Zusammenfassend kann ich also sagen, und ich wiederhole hier gerne, was ich schon öfters bei einer Gerhard-Fürst-Preisverleihung sagen durfte: Liebe Preisträger, wir freuen uns, dass es junge Menschen wie Sie gibt, die sich wissenschaftlich fundiert mit Problemen der Amtsstatistik auseinandersetzen. Sie alle haben den Preis verdient, herzlichen Glückwunsch! 🎉

# SAISONBEREINIGUNG DER MONATLICHEN KASSENMÄSSIGEN STEUEREINNAHMEN

Luis Federico Flores, Christiane Loos

↳ **Schlüsselwörter:** kassenmäßige Steuereinnahmen – Saisonbereinigung – Corona-Pandemie – Bundes-, Landes- und Gemeinschaftsteuer – Konjunkturstatistik

## ZUSAMMENFASSUNG

Im Zuge der Corona-Pandemie hat das Statistische Bundesamt erstmals im August 2020 die neuen kalender- und saisonbereinigten monatlichen kassenmäßigen Steuereinnahmen veröffentlicht. Sie sind wichtige Indikatoren für die Finanzpolitik und die Konjunkturbeobachtung in Deutschland und ermöglichen frühzeitig Aussagen zu Entwicklungstendenzen der kassenmäßigen Steuereinnahmen – die Ergebnisse sind durchschnittlich bereits 26 Tage nach Ende eines Monats verfügbar. Der Artikel erläutert die Struktur der zugrundeliegenden Daten und stellt die Herausforderungen der Saisonbereinigung der kassenmäßigen Steuereinnahmen und die angewandten Lösungsansätze vor. Abschließend wird die Entwicklung der bereinigten kassenmäßigen Steuereinnahmen im bisherigen Verlauf der Corona-Pandemie dargestellt.

↳ **Keywords:** *cash tax revenue – seasonal adjustment – corona pandemic – federal, Land and joint taxes – short-term economic statistics*

## ABSTRACT

*In the context of the corona pandemic, the Federal Statistical Office published the new calendar and seasonally adjusted monthly cash tax revenues for the first time in August 2020. They are important indicators for fiscal policy and short-term economic monitoring in Germany and they enable early assessment of trend movements in the cash tax revenues. On average, the results are available as early as 26 days after the end of the reporting month. This article explains the structure of the underlying data and presents the challenges involved with the seasonal adjustment of cash tax revenues as well as the approaches used to address them. Finally, the article discusses the development of the adjusted cash tax revenues during the course of the corona pandemic so far.*



**Luis Federico Flores**

ist Volkswirt und seit Dezember 2019 wissenschaftlicher Mitarbeiter im Referat „Konjunkturindizes, Saisonbereinigung“ des Statistischen Bundesamtes. Er beschäftigt sich schwerpunktmäßig mit Weiterentwicklungsprojekten im Bereich der Saisonbereinigung.



**Christiane Loos**

ist Soziologin und seit 2010 im Statistischen Bundesamt beschäftigt. Seit März 2018 leitet sie das Referat „Umsatzsteuer, Verbrauchsteuern, Steuerhaushalt, Realsteuervergleich“.

## 1

### Einleitung

Als wichtige Indikatoren für die Finanzpolitik und die Konjunkturbeobachtung in Deutschland veröffentlicht das Statistische Bundesamt vor dem Hintergrund der Corona-Pandemie seit August 2020 Ergebnisse zu kalender- und saisonbereinigten monatlichen kassenmäßigen Steuereinnahmen. Diese sind durchschnittlich 26 Tage nach Ende eines Monats verfügbar und ermöglichen somit frühzeitig Aussagen zu Entwicklungstendenzen der kassenmäßigen Steuereinnahmen. Die Saisonbereinigung wird partnerschaftlich mit der Deutschen Bundesbank durchgeführt.

Steuern sind die wichtigste Einnahmequelle von Bund, Ländern und Gemeinden, sie machen gut die Hälfte der Einnahmen des Staates aus. Bei den auch als Steuererwerb bezeichneten kassenmäßigen Steuereinnahmen handelt es sich um Geldbeträge, die in einem bestimmten Zeitraum in die Kassen der Gebietskörperschaften geflossen sind. Dabei ist nicht relevant, wann die Steuerschuld entstanden ist oder für welches Jahr die Zahlungen geleistet wurden, für das die Steuerpflicht bestand. Das Ist-Aufkommen eines bestimmten Berichtszeitraums enthält daher auch Vorauszahlungen, Abschlusszahlungen, Nachzahlungen, Säumniszuschläge und so weiter ohne Rücksicht auf eine periodengerechte Erfassung. Das Aufkommen kann auch durch Erstattungen gemindert sein.

Die Höhe des kassenmäßigen Steuererwerbs wird regelmäßig in unterschiedlicher Periodizität publiziert. Dazu stellt das Bundesministerium der Finanzen dem Statistischen Bundesamt die Ergebnisse für die Gemeinschaft-, Bundes- und Landessteuern monatlich für die Statistik zum Steuererwerb zur Verfügung. Die Höhe des kassenmäßigen Steuererwerbs aus Gemeindesteuern wird quartalsweise mit der Vierteljährlichen Kassenstatistik erhoben.<sup>1</sup>

Jährlich wiederkehrende Effekte, sogenannte Saisonwirkungen, beeinflussen die Entwicklung der kassenmäßigen Steuereinnahmen stark. Saisonwirkungen entstehen aus

1 Weitere Informationen zur Statistik des Steuererwerbs bieten der Qualitätsbericht (Statistisches Bundesamt, 2020) und die vierteljährlich erscheinende Fachserie 14 Reihe 4 „Steuerhaushalt“ unter [www.destatis.de](http://www.destatis.de).

saisonalen Verläufen der wirtschaftlichen Aktivitäten und den Steuerausgestaltungen. Solche Einflüsse können durch die Saisonbereinigung aus den Zeitreihenwerten herausgerechnet werden, was die Beurteilung aktueller Entwicklungstendenzen erleichtert.<sup>2</sup> Die komplexe Struktur der Originalwerte führt dazu, dass die Saisonbereinigung der kassenmäßigen Steuereinnahmen mit besonderen Herausforderungen verbunden ist.

Der Aufsatz beschreibt zunächst in Kapitel 2 die Struktur der den kassenmäßigen Steuereinnahmen zugrundeliegenden Daten. Die anschließenden Kapitel 3 und 4 erläutern methodische Aspekte der Saisonbereinigung sowie Herausforderungen und angewandte Lösungsansätze. Kapitel 5 stellt die Entwicklung der bereinigten kassenmäßigen Steuereinnahmen im bisherigen Verlauf der Corona-Pandemie dar, ein Fazit schließt den Beitrag ab.

## 2

### Die kassenmäßigen Steuereinnahmen

In der Statistik zu den kassenmäßigen Steuereinnahmen werden Ergebnisse zu etwa 40 verschiedenen Steuerarten veröffentlicht. Saisonbereinigt werden insgesamt 21 Steuerarten<sup>3</sup>, die alle Einnahmen aus Gemeinschaftsteuern und ausgewählte Einnahmen aus Bundes- und Landessteuern umfassen. Die auf Bund, Länder und Gemeinden zu verteilenden Gemeinschaftsteuern machen mit über 80% den weitaus größten Anteil an den zu bereinigenden kassenmäßigen Steuereinnahmen aus.

➤ **Tabelle 1**

Die Entwicklung der kassenmäßigen Steuereinnahmen – und damit des darin enthaltenen saisonalen und kalendarischen Musters – hängt grundsätzlich von zwei Einflussfaktoren ab: einerseits von den wirtschaftlichen und rechtlichen Vorgängen, an denen die jeweilige Steuer anknüpft, und andererseits von der gesetzlichen Ausgestaltung der Besteuerung. Die jeweilige Steuerart kann durch Festlegungen der Bemessungsgrundlagen, Steuersätze, Anmeldezeiträume, Abgabefristen oder

2 Zur Saisonbereinigung im Statistischen Bundesamt siehe Linz und andere (2018).

3 Nicht saisonbereinigt werden die Gemeindesteuern, für die nur vierteljährliche Ergebnisse vorliegen. Weiterhin werden einige Bundes- und Landessteuerarten wegen ihrer geringen Bedeutung sowie eines fehlenden Saisonmusters nicht saisonbereinigt.

**Tabelle 1**

Steuerarten, für die eine Saisonbereinigung durchzuführen ist

	Anteil an der Summe der kassenmäßigen Steuereinnahmen von Gemeinschaft-, Bundes- und Landessteuern 2019	Variationskoeffizient von Januar 1999 bis August 2020
	%	
<b>Gemeinschaftsteuern</b>	81,3	34
Lohnsteuer	30,4	28
+ Veranlagte Einkommensteuer	8,8	192
Nicht veranlagte Steuern vom Ertrag	3,2	77
+ Abgeltungsteuer	0,7	77
+ = Kapitalertragsteuer <sup>1</sup>	4,0	61
+ Körperschaftsteuer	4,4	189
= Steuern vom Einkommen <sup>1</sup>	47,6	53
Umsatzsteuer	25,3	23
+ Einfuhrumsatzsteuer	8,3	24
= Steuern vom Umsatz <sup>1</sup>	33,7	22
<b>Bundessteuern</b>	15,2	34
darunter:		
Versicherungsteuer	2,0	96
Tabaksteuer	2,0	36
Alkoholsteuer	0,3	38
Energiesteuer	5,6	56
Stromsteuer	0,9	29
Luftverkehrssteuer	0,2	32
Solidaritätszuschlag	2,7	48
<b>Landessteuern</b>	3,6	22
darunter: Biersteuer	0,1	18

<sup>1</sup> Diese Steuerart wird indirekt bereinigt.

Zahlungsfristen charakterisiert werden. Diese Festlegungen sind auch für die Vorgehensweise bei der Saisonbereinigung relevant und sollen daher im Folgenden kurz für ausgewählte Steuerarten beschrieben werden.<sup>4</sup> Ergänzend enthält die Tabelle den Variationskoeffizienten, der erste Hinweis auf das Ausmaß von Schwankungen des Steueraufkommens bei der betreffenden Steuerart gibt. Diese Schwankungen können verschiedene Ursachen haben und auch irreguläre Bewegungen sein, die sich nicht durch kalendarische oder Saisonmuster oder mittel- und langfristige Entwicklungstrends erklären lassen.

<sup>4</sup> Eine ausführliche Erläuterung aller Steuerarten findet sich in Bundesministerium der Finanzen (2019). [↗](#) Übersicht 1 am Ende des Kapitels beschreibt weitere, im Text nicht näher dargestellte Steuerarten, die ebenfalls saisonbereinigt werden.

## Lohnsteuer

Die Lohnsteuer ist eine besondere Erhebungsform der Einkommensteuer für natürliche Personen. Dabei wird der Arbeitslohn besteuert und die zu zahlende Steuer direkt vom Arbeitgeber vom Bruttolohn abgezogen und an das Finanzamt abgeführt. Dies erfolgt gesammelt für alle Beschäftigten eines Arbeitgebers zu festgelegten Zeitpunkten. Abhängig von der Höhe der Lohnsteuerzahlungen muss die Anmeldung im Monat nach dem Ende des Anmeldezeitraums monatlich, quartalsweise oder jährlich erfolgen und auch direkt gezahlt werden. Einfluss auf das in der Statistik ausgewiesene Aufkommen haben nicht nur steuerrechtliche Änderungen, sondern auch Änderungen der Höhe des Kindergeldes oder die Zulage zur Altersvorsorge (Riester-Rente). Letztere werden aus den Lohnsteuereinnahmen gezahlt und mindern diese. Der Variationskoeffizient ist mit 28 vergleichsweise gering, was auf eine geringe Volatilität der Einnahmen hindeutet. Insbesondere im Dezember zeigt sich ein hohes Maximum in der unterjährigen Entwicklung.

## Veranlagte Einkommensteuer

Die Veranlagung zur Einkommensteuer umfasst die Abgabe einer Steuererklärung mit anschließender Steuerfestsetzung in Form eines Steuerbescheids. Vor allem für Steuerpflichtige mit Gewinneinkünften, aber auch mit bestimmten Überschusseinkünften kann dies erst nach Ablauf des Veranlagungsjahrs erfolgen. Daher legt das zuständige Finanzamt die Höhe der Vorauszahlungen für das laufende Jahr auf Basis der Vorjahre fest. Es gibt vier Vorauszahlungstermine im Jahr, und zwar im März, Juni, September und Dezember. Zusätzlich prägen Nachzahlungen und Erstattungen aus den Steuerfestsetzungen durch die Abgabe einer Steuererklärung insbesondere von Steuerpflichtigen mit Einkünften aus nichtselbstständiger Arbeit das Steueraufkommen der veranlagten Einkommensteuer. Allerdings gibt es auch hier weitere Zahlungen, die das Aufkommen mindern. Dazu gehören die Investitionszulage und die Eigenheimzulage. Die Höhe dieser kassenmäßigen Steuereinnahmen eines Monats hängt also von vielen Einflussfaktoren ab, die in unterschiedlichen Perioden der Vergangenheit wirksam wurden. Der Variationskoeffizient für die Veranlagte Einkommensteuer ist mit 192 der

höchste im Vergleich zu allen betrachteten Steuerarten. Insbesondere durch die Erstattungen kommt es zu sehr hohen Schwankungen, sodass das gesamte Aufkommen zeitweise sogar negativ sein kann.

## Körperschaftsteuer

---

Die Körperschaftsteuer ist die Einkommensteuer für juristische Personen. Die Steuerfestsetzungen erfolgen über die Abgabe einer jährlichen Steuererklärung nach Ende des Veranlagungsjahres auf Basis der Jahresabschlüsse der Unternehmen. Auch hier werden auf Basis der Vorjahre Vorauszahlungen festgesetzt und zu den gleichen Terminen wie bei der veranlagten Einkommensteuer an das Finanzamt abgeführt. Wie schon die veranlagte Einkommensteuer weist auch die Körperschaftsteuer hohe Schwankungen der Einnahmen auf, was zu einem sehr hohen Variationskoeffizienten (189) führt.

## Umsatzsteuer

---

Die Umsatzsteuer (auch Mehrwertsteuer genannt) besteuert allgemein den gesamten privaten und öffentlichen Verbrauch und wird vom Endverbraucher getragen. Die Bemessungsgrundlage bilden grundsätzlich die Entgelte und Steuerschuldner sind die Unternehmen, da es technisch nicht möglich ist, die Umsatzsteuer bei den Konsumierenden zu erheben. Steuerpflichtige sind dazu verpflichtet, zehn Tage nach Ende des Voranmeldungszeitraums eine Umsatzsteuer-Voranmeldung abzugeben und die fällige Umsatzsteuer auch direkt abzuführen. Der Voranmeldungszeitraum ist generell das Quartal. Bei einer Steuerzahlung unter 1 000 Euro erfolgt die Anmeldung jährlich. Sollte allerdings die zu zahlende Steuer für das Vorjahr über 7 500 Euro liegen, dann muss die Voranmeldung monatlich erfolgen. Für die Steuerpflichtigen besteht die Möglichkeit, eine sogenannte Dauerfristverlängerung zu beantragen. Dadurch kann der Termin zur Abgabe der Voranmeldung um einen Monat verschoben werden. Von dieser Möglichkeit macht ein großer Teil der Unternehmen Gebrauch. Damit verschieben sich die Zahlungen der fälligen Umsatzsteuer ebenfalls um einen Monat. Nach Ablauf eines Kalenderjahres müssen die Steuerpflichtigen eine Steuererklärung abgeben, und zwar auch diejenigen, die nicht zur Abgabe einer Voranmeldung verpflichtet sind. Hierbei werden dann bereits getätigte Zahlungen verrechnet.

Die Schwankungen in den Werten der Umsatzsteuer sind vergleichsweise gering, was einen geringen Variationskoeffizienten von 23 zur Folge hat.

## Einfuhrumsatzsteuer

---

Die Einfuhrumsatzsteuer besteuert die Einfuhr von Gegenständen im Inland oder in Zollanschlussgebieten. Die Steuer bemisst sich am Zollwert, das ist meist der tatsächlich gezahlte oder zu zahlende Preis. Die Anmeldung muss bis zum 90. Tag nach Einfuhr erfolgen. Die Steuer ist spätestens am 10. Tag nach der Zollanmeldung fällig. Sollte ein sogenanntes Aufschubkonto bestehen (etwa für Unternehmen, die regelmäßig importieren), ist die Zahlung spätestens zum 16. des Folgemonats zu leisten. Wie die Umsatzsteuer weisen auch die Ergebnisse der Einfuhrumsatzsteuer vergleichsweise geringe Schwankungen (Variationskoeffizient: 24) auf.

## Energiesteuer

---

Die Energiesteuer besteuert den Verbrauch verschiedener Energieerzeugnisse. Die Steueranmeldung hat in der Regel im Folgemonat zu erfolgen, die Steuer wird einen Monat später fällig. Erdgaserzeugnisse können wahlweise auch jährlich bis Mai des Folgejahres angemeldet werden. Die Steuer ist dann im Juni des gleichen Jahres fällig. Wie schon bei vorhergehenden Verbrauchsteuern gibt es auch hier die Sonderregelung für den Monat Dezember mit erhöhten Steuereinnahmen, die zu vergleichsweise hohen Einnahmeschwankungen führt, insbesondere um den Jahreswechsel. Der Variationskoeffizient von 56 spiegelt dies wider.

# Saisonbereinigung der monatlichen kassenmäßigen Steuereinnahmen

## Übersicht 1

### Weitere ebenfalls saisonbereinigte Steuerarten

Nicht veranlagte Steuern vom Ertrag	Die nicht veranlagten Steuern vom Ertrag weisen vor allem die Steuereinnahmen durch die jährlichen Dividendenausschüttungen der Unternehmen aus. Von Jahr zu Jahr unterscheiden sich die Termine der Jahreshauptversammlungen und damit verbunden die Zeitpunkte der Dividendenausschüttungen. Daher schwanken die kassenmäßigen Steuereinnahmen aus nicht veranlagten Steuern vom Ertrag erheblich, was sich auch im Variationskoeffizienten für die nicht veranlagten Steuern vom Ertrag von 77 widerspiegelt.
Abgeltungsteuer	Zur Abgeltungsteuer zählen vor allem Zahlungen auf Zinserträge und Erträge aus Kapitalveräußerungen. Die Einnahmen aus dieser Steuer waren bis Ende 2014 durch eine hohe Volatilität mit einem ausgeprägt hohen Maximum im Januar jedes Jahres gekennzeichnet. Da jedoch in den letzten Jahren das Zinsniveau stark gefallen ist, hat sich die Entwicklung der Steuereinnahmen aus der Abgeltungsteuer stark gewandelt. Insbesondere die außergewöhnlich hohen Maxima im Januar eines Jahres sind weggebrochen, sodass der Variationskoeffizient von 77 deutlich niedriger ausfällt, wenn man nur die Werte ab 2015 betrachtet.
Versicherungsteuer	Mit der Versicherungsteuer werden Prämien- oder Beitragszahlungen aus Versicherungsverträgen besteuert. Generell erfolgt die Anmeldung monatlich, wobei je nach Höhe der Steuerzahlungen auch vierteljährliche oder jährliche Meldungen möglich sind. Die zu zahlende Steuer bemisst sich anhand der gezahlten Versicherungsentgelte. Der Variationskoeffizient für die Versicherungsteuer ist mit 96 sehr hoch im Vergleich zu den anderen zu bereinigenden Bundessteuern. Dies geht vor allem auf die außergewöhnlich hohen Maxima im Februar eines jeden Jahres zurück. In den restlichen Monaten sind die Schwankungen vergleichsweise niedrig.
Tabaksteuer	Die Tabaksteuer ist eine Verbrauchsteuer, die verschiedene Tabakwaren und gleichgestellte Erzeugnisse besteuert. Die Anmeldung erfolgt mit der Bestellung der Steuerzeichen. Die Zahlungsfristen unterscheiden sich je nach Art der Tabakwaren. Für Zigaretten und Rauchtobak fällt die Zahlung im Folgemonat an, für den Dezember gibt es eine Sonderregelung: Für Steuerzeichen, die bis zum 15. Dezember bezogen wurden, werden die Steuern noch im Dezember gezahlt. In den (nicht saisonbereinigten) kassenmäßigen Steuereinnahmen ist daher ein Ausschlag für den Dezember nach oben zu erkennen. Für Zigarren und Zigarillos, die mengenmäßig nicht mit den Zigaretten vergleichbar sind, liegt die Frist zur Zahlung erst im zweiten Folgemonat.
Alkoholsteuer (bis 2017: Branntweinsteuer)	Bis zur Aufhebung des Branntweinmonopols am 31. Dezember 2017 wurde die Branntweinsteuer erhoben, seitdem die Alkoholsteuer. Besteuert werden Alkohol beziehungsweise Branntwein und alkohol-beziehungsweise branntweinhaltige Waren. Die Steuer entsteht durch die Überführung der Erzeugnisse in den freien Verkehr und ist monatlich im Folgemonat anzumelden. Die zu zahlende Steuer wird dann spätestens im zweiten Folgemonat nach der Steuerentstehung fällig. Bis zum Jahr 2006 gab es für die damals geltende Branntweinsteuer eine ähnliche Sonderregelung für den Monat Dezember wie bei der Tabaksteuer. Dies hat jeweils im Dezember zu sehr hohen Steuereinnahmen geführt, denen im Folgemonat Januar dann ein starker Rückgang folgte. Diese Regelung wurde ab Januar 2007 abgeschafft, sodass die Sonderentwicklung zum Jahreswechsel nicht mehr zu beobachten ist. Zudem wurden die Fälligkeitsfristen innerhalb des zweiten Folgemonats verschoben. Beides hatte auch Auswirkungen auf die Schwankungen der Ergebnisse im Jahresverlauf, die bis dahin hoch waren und in der Folge deutlich abgenommen haben.
Stromsteuer	Mit der Stromsteuer wird aus dem Versorgungsnetz entnommener Strom besteuert. Die Anmeldung erfolgt wahlweise monatlich oder jährlich. Erfolgt die Anmeldung monatlich, so muss dies spätestens im Folgemonat geschehen, in dem auch die Zahlung fällig wird. Bei jährlicher Anmeldung ist dann eine monatliche Vorauszahlung fällig, die ebenfalls im Folgemonat zu zahlen ist. Die Jahressteuerschuld muss spätestens bis Mai des Folgejahres angemeldet werden und die Steuer wird spätestens im Juni fällig. Dabei werden die Vorauszahlungen verrechnet, wodurch es auch zu Nachzahlungen und Erstattungen kommen kann. Die Schwankungen sind insbesondere in den letzten zehn Jahren stark zurückgegangen, deshalb ist der Variationskoeffizient (29) ebenfalls gering.
Luftverkehrssteuer	Die Luftverkehrssteuer besteuert die Abflüge von deutschen Flughäfen. Die Steuersätze richten sich dabei nach unterschiedlichen Distanzklassen, sie wurden zuletzt im April 2020 im Rahmen des Klimaschutzprogramms 2030 deutlich angehoben. Die Anmeldung durch die Luftverkehrsunternehmen als Steuerschuldner hat monatlich zu erfolgen. Wie schon bei der Tabak-, Energie- und früheren Branntweinsteuer gibt es auch hier die Sonderregelung für Dezember, die in diesem Monat zu sehr hohen und im Januar dann zu sehr niedrigen Einnahmen führt.
Solidaritätszuschlag	Der Solidaritätszuschlag wurde zur Finanzierung der Deutschen Einheit 1995 eingeführt. Die Bemessungsgrundlage sind die verschiedenen Steuern vom Einkommen, mit deren Zahlung auch der Solidaritätszuschlag abgeführt wird. Entsprechend hat deren Streuungsintensität Einfluss auf den Solidaritätszuschlag, für den der Variationskoeffizient von 48 auf eine hohe Streuung hindeutet.
Biersteuer	Die Biersteuer wird wie alle anderen Verbrauchsteuern durch Abgabe einer Steueranmeldung beim zuständigen Hauptzollamt erhoben; sie ist monatlich im Folgemonat abzugeben. Versteuert werden Bier und Biermischungen, die in den steuerrechtlich freien Verkehr überführt wurden. Ebenfalls im Folgemonat ist dann die Biersteuer zu entrichten. Die Streuung der Einnahmen aus der Biersteuer ist die niedrigste aller betrachteten Steuerarten, was allerdings auch an den geringeren monatlichen Einnahmen liegt.

### 3

## Warum Saisonbereinigung?

Die Entwicklung der kassenmäßigen Steuereinnahmen spiegelt einerseits die wirtschaftliche Aktivität der besteuerten Subjekte wider und ist ein früh verfügbarer Konjunkturindikator. Andererseits gibt sie Auskunft über den finanziellen Spielraum der Gebietskörperschaften und dessen aktuelle Entwicklung und ist damit ein wichtiger Indikator für die Finanzpolitik.

Allerdings beeinflussen jährlich wiederkehrende sowie kalenderbedingte Effekte die Entwicklung der kassenmäßigen Steuereinnahmen in starkem Umfang, was die Beurteilung aktueller Entwicklungstendenzen erschwert. Saisoneffekte können dabei aus den unterjährig verlaufenden der wirtschaftlichen Aktivitäten oder durch die steuerrechtlichen Ausgestaltungen entstehen. Zum Beispiel sind bei der Körperschaftsteuer Saisonhöhepunkte zu den Vorauszahlungsterminen im März, Juni, September und Dezember eines Jahres zu beobachten. Solche Einflüsse können mithilfe von Saisonbereinigungsverfahren ausgeschlossen werden. Dabei sollen Schwankungen, die jährlich in denselben Jahreszeiten in ähnlicher Intensität wiederkehren, aus den Zeitreihenwerten herausgerechnet werden. Weiterhin sollen mithilfe der Kalenderbereinigung auch Schwankungen, die aufgrund der jeweiligen Kalenderkonstellation zu erwarten sind, aus den Ergebnissen entfernt werden. Kalendarische Effekte ergeben sich bei den Steuereinnahmen beispielsweise durch die unterschiedliche Lage beweglicher Feiertage und damit einhergehend unterschiedlich vielen Arbeitstagen für gleichnamige Monate. Die kalender- und saisonbereinigten Ergebnisse ermöglichen eine bessere Beurteilung der konjunkturellen Entwicklung und der Effekte von Änderungen in den steuerrechtlichen Ausgestaltungen (zum Beispiel die Anpassung eines Steuersatzes).

Um die Entwicklung der kassenmäßigen Steuereinnahmen zu beurteilen, erfolgte vor ihrer Veröffentlichung in saisonbereinigter Form der Vorjahresvergleich. Das ist ein einfacher Weg, saisonale Schwankungen aus der Betrachtung auszuschließen. Dabei wird der (unbereinigte) Originalwert der aktuellen Periode mit dem entsprechenden Wert desselben Monats im Vorjahr verglichen. Die Ergebnisse von Vorjahresvergleichen können aber in eine andere Richtung weisen als Vormonats-

vergleiche auf der Basis von saisonbereinigten Ergebnissen. Das ist zum Beispiel der Fall, wenn ein Abschwung zu Ende geht und der Vormonatsvergleich der saisonbereinigten Ergebnisse bereits eine Tendenz nach oben anzeigt, während das aktuelle unbereinigte Ergebnis noch unter dem Vorjahreswert liegt. Vorperiodenvergleiche auf der Basis von saisonbereinigten Ergebnissen ermöglichen daher grundsätzlich eine zeitnähere Konjunkturanalyse als Vorjahresvergleiche.

### 4

## Vorgehen bei der Saisonbereinigung

Für die Saisonbereinigung wird im Statistischen Bundesamt das Verfahren X13 in JDemetra+ verwendet.<sup>15</sup> Die erste Stufe des Verfahrens besteht aus einer RegARIMA-Modellierung, bei der die Zeitreihe an den Rändern um Schätzwerte verlängert sowie gegebenenfalls um Ausreißer- und Kalendereinflüsse bereinigt wird.<sup>16</sup> Mit Kalenderregressoren können kalendarische Effekte und mit Ausreißerregressoren außergewöhnliche Zeitreihenentwicklungen, etwa die Auswirkung von Steuererhöhungen, explizit modelliert werden. Diese beiden Effekte fließen dann nicht in das Saisonmuster, sodass dieses besser schätzbar ist. Die Ausreißereinflüsse bleiben jedoch in den saisonbereinigten Ergebnissen enthalten. Vereinfacht dargestellt handelt es sich um eine iterative Glättung der Originaldaten anhand von Trend- und Saisonfiltern, die aus gewichteten gleitenden Durchschnitten berechnet werden. Die Saisonfaktoren eines Monats ergeben sich aus der Glättung der Abweichungen der Originalwerte von den Trendwerten des betreffenden Monats. Die Trendwerte sind in der ersten Iterationsstufe ihrerseits Glättungen der Originalwerte. In späteren Iterationsstufen bezieht sich die Glättung auf die (vorläufigen) saisonbereinigten Reihen.

Für die Kalenderbereinigung werden Kalenderregressoren gebildet. So wird zum Beispiel der Regressor zur Berücksichtigung der Arbeitstage berechnet, indem zunächst die Feiertage und die Wochenendtage von der

5 Eine ausführliche Beschreibung der Vorgehensweise bei der Saisonbereinigung im Statistischen Bundesamt findet sich in Linz und andere (2018).

6 Die Verlängerung der Zeitreihe um geschätzte Werte wird unter anderem für die zweite Stufe des Verfahrens benötigt, bei welcher gleitende Durchschnitte berechnet werden.

Zahl der Tage im aktuellen Monat abgezogen werden. Feiertage, die nur in bestimmten Bundesländern gesetzlich festgelegt sind, werden mit dem Anteil der Arbeitnehmerinnen und Arbeitnehmer dieser Bundesländer an der Gesamtzahl der Arbeitnehmerinnen und Arbeitnehmer in Deutschland gewichtet.<sup>7</sup> Da jedoch nicht die Anzahl der Arbeitstage selbst, sondern die Abweichung dieser von ihrem Mittelwert von Interesse ist, wird der Regressor als Differenz zwischen der Anzahl der Arbeitstage des Monats und einem langjährigen monatspezifischen Mittel berechnet. Mit der RegARIMA-Modellierung kann dann der Einfluss eines zusätzlichen Arbeitstags auf die Höhe des Originalwerts aus den Vergangenheitswerten der betreffenden Statistik geschätzt werden. Diese Schätzwerte werden verwendet, um Kalenderfaktoren zu berechnen. Mithilfe dieser Faktoren wird der Einfluss von Verschiebungen bei der Zahl der Arbeitstage im jeweiligen Monat aus den Originaldaten herausgerechnet. Solche Verschiebungen können einen deutlichen Einfluss auf die monatlichen Steuereinnahmen haben. Zum Beispiel führte vor der Coronakrise ein zusätzlicher Arbeitstag in Deutschland durchschnittlich zu einem Anstieg des Luftverkehrsteueraufkommens um 2,8% in dem entsprechenden Monat. Gemäß europäischer Richtlinien (Eurostat, 2015) berücksichtigt die Kalender- und Saisonbereinigung den Einfluss von Brückentagen, Schulferienverschiebungen oder der Witterung nicht, er bleibt im saisonbereinigten Ergebnis sichtbar.

Die Saisonbereinigung der kassenmäßigen Steuereinnahmen ist mit besonderen Herausforderungen verbunden: Einerseits schlagen sich die wirtschaftlichen Aktivitäten der Akteure und damit Veränderungen bei deren Steuerschuld oft erst zeitversetzt in entsprechenden Änderungen der kassenmäßigen Steuereinnahmen nieder. Wegen der jeweiligen Anmelde-, Abgabe- und Zahlungsfristen, Voraus- und Nachzahlungen können sich zum Beispiel Kalendereffekte, die in einem bestimmten Monat die wirtschaftliche Aktivität beeinflussen, in verschiedenen späteren Perioden bei den kassenmäßigen Steuereinnahmen auswirken. Um solche Verschiebungen bei der Auswirkung von Kalendereffekten zu berücksichtigen, werden bei der Kalendereinigung der kassenmäßigen Steuereinnahmen zusätzlich zu den Standardregressoren auch zeitlich verschobene Varian-

ten der Kalenderregressoren eingesetzt. Andererseits wirken sich die häufig vorkommenden Änderungen des Steuerrechts zum Teil erst nach mehreren Jahren auf die Entwicklung der kassenmäßigen Steuereinnahmen aus. Diese Effekte werden mit verschiedenen Ausreißertypen und unter Umständen nicht nur in der Periode der Gesetzesänderungen, sondern auch in nachfolgenden Perioden modelliert.

Die folgenden Abschnitte erläutern die Vorgehensweise bei der Saisonbereinigung der kassenmäßigen Steuereinnahmen für drei ausgewählte Steuerarten beispielhaft: die Alkoholsteuer, die veranlagte Einkommensteuer und die Umsatzsteuer.

### Alkoholsteuer (bis 2017: Branntweinsteuer)

Steuerrechtsänderungen können das Saisonmuster in der Entwicklung der kassenmäßigen Steuereinnahmen grundlegend ändern. [↘ Grafik 1](#) zeigt die Entwicklung der unbereinigten Alkohol- beziehungsweise Branntweinsteuer im Zeitraum Januar 1999 bis August 2020. Bis Januar 2007 weisen die Einnahmen bei dieser Steuerart ein ausgeprägtes Saisonmuster auf, mit Höhepunkten im Dezember und Tiefpunkten im Februar. Ab Februar 2007 sind die Saisonauschläge weniger stark ausgeprägt und die Saisonhochs treten im Januar und Februar auf. Die Änderung der Entwicklung dieser Zeitreihe ist auf eine Anpassung der Fälligkeitsfristen sowie den Wegfall von Sonderregelungen für den Dezember zurückzuführen. Bis 2006 war die im November entstandene Steuer bereits im Dezember zu entrichten. Danach war die Alkoholsteuer in bestimmten Fällen der Steuerentstehung erst im zweiten auf die Steuerentstehung folgenden Monat zu zahlen.

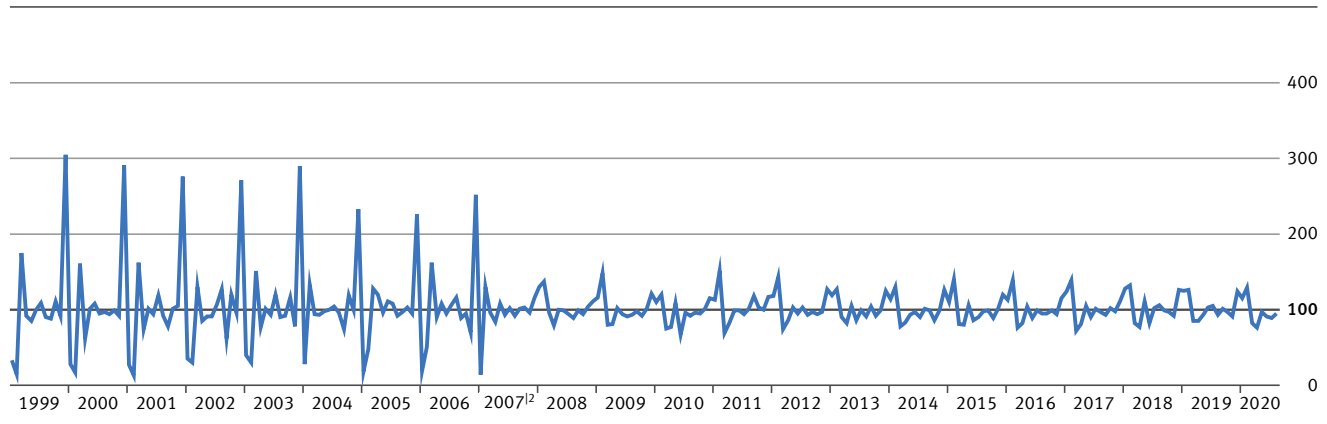
[↘ Grafik 2](#) illustriert die Änderung des Saisonmusters über die Zeit durch die Steuerrechtsänderung. Sie zeigt die rohen Saisonfaktoren der Alkoholsteuer als Datenpunkte, berechnet für den gesamten Zeitraum Januar 1999 bis August 2020 und definiert als Abweichungen der (vorab bereinigten) Originalwerte vom Trend. Ein roher Saisonfaktor von 0,97 für August 2020 bedeutet zum Beispiel, dass der vorab bereinigte Originalwert der Alkoholsteuer in diesem Monat 3% unter dem Trendzyklus-Wert lag. Die rohen Saisonfaktoren sind zuerst nach Monaten und dann nach Jahren sortiert. Zum

<sup>7</sup> Eine ausführliche Darstellung findet sich in Deutsche Bundesbank (2012).



**Grafik 1**

Entwicklung der nicht saisonbereinigten Alkoholsteuer<sup>1</sup>  
2015 = 100



1 Bis 2017 Branntweinsteuer.  
2 Ab 2007 Anpassung der Fälligkeitsfristen und Wegfall von Sonderregelungen für den Dezember.

2020 - 0545

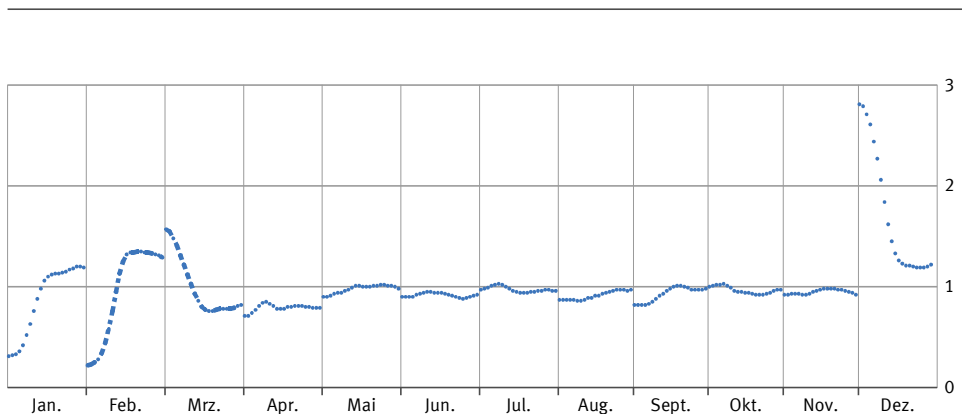
Beispiel sind für den Monat Januar die rohen Saisonfaktoren für alle Jahre des Betrachtungszeitraums als Punkte direkt hintereinander eingezeichnet. Es zeigt sich, dass die rohen Saisonfaktoren für die Monate Januar und Februar von Jahr zu Jahr stark ansteigen, während sich die rohen Saisonfaktoren für die Monate März und Dezember über die Jahre hinweg in die entgegengesetzte Richtung bewegen.

seits nah bei den rohen Saisonfaktoren liegen und andererseits über die Jahre hinweg relativ stabil sein. Durch die starke Änderung des Saisonmusters über die Zeit ist die Ableitung der geglätteten Saisonfaktoren stark erschwert. Daher wurde für die Durchführung der Saisonbereinigung die Zeitreihe aufgeteilt, und zwar in den Zeitraum Januar 1999 bis Januar 2007 sowie in den Zeitraum ab Februar 2007. Der Vorzug einer getrennten Saisonbereinigung ist in [Grafik 3](#) durch die rote Zeitreihe dargestellt, welche vergleichsweise weniger Ausschläge aufweist. Dies ist beispielsweise jeweils im Februar der Jahre 2004, 2005, 2006 und 2007 deutlich zu erkennen.

Bei der Saisonbereinigung werden mittels gleitender Durchschnitte aus den rohen Saisonfaktoren geglättete Werte gebildet, die den typischen Saisonausschlag eines Monats repräsentieren sollen. Sie sollen einer-

**Grafik 2**

Rohe Saisonfaktoren der Alkoholsteuer für den Zeitraum Januar 1999 bis August 2020

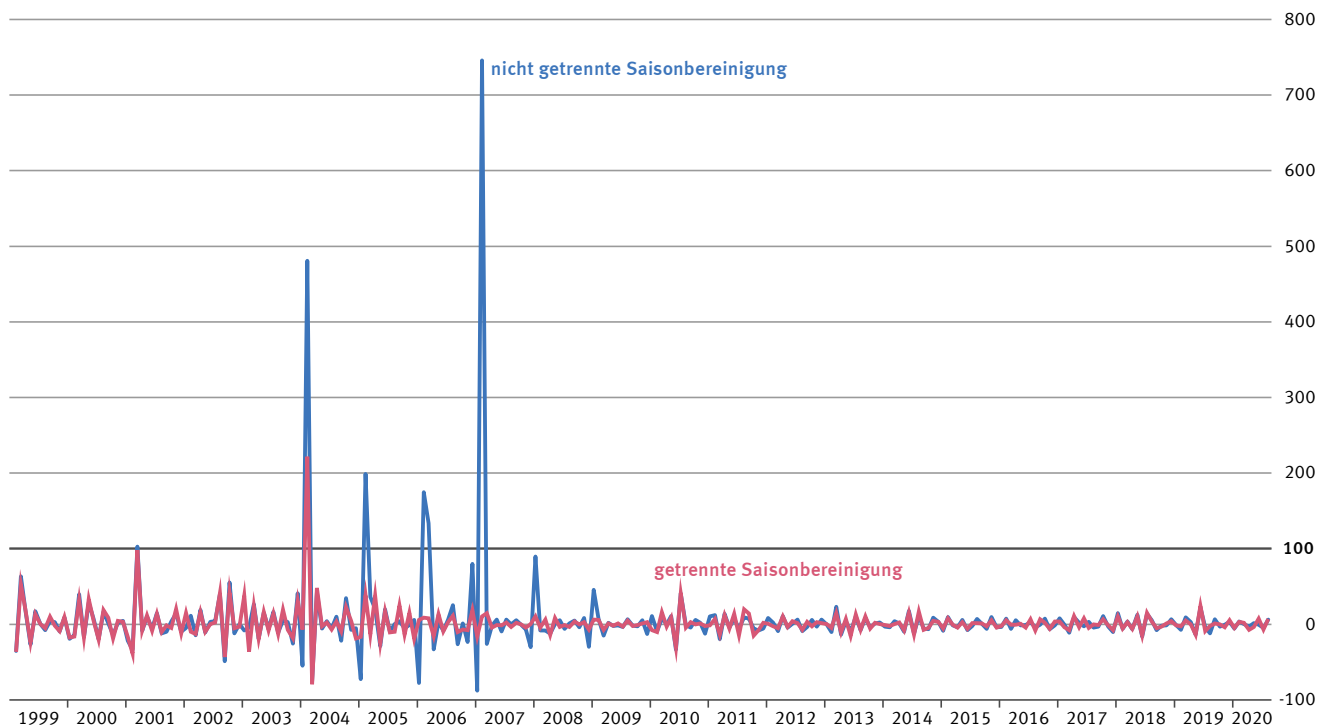


2020 - 0546

**Grafik 3**

Vormonatsraten der Alkoholsteuer (saisonbereinigt)

2015 = 100



2020 - 0547

### Veranlagte Einkommensteuer

↳ Grafik 4 zeigt die Entwicklung der veranlagten Einkommensteuer. Diese weist ein ausgeprägtes Saisonmuster mit Hochpunkten am Ende der Quartale auf, da die Vorauszahlungen in den Monaten März, Juni, September und Dezember stattfinden. Die Zeitreihe enthält negative Werte aufgrund der Steuererstattungen sowie der steigenden Zahlungen der Eigenheimzulage. Außerdem zeigt sie eine steigende Volatilität über den Zeitverlauf.

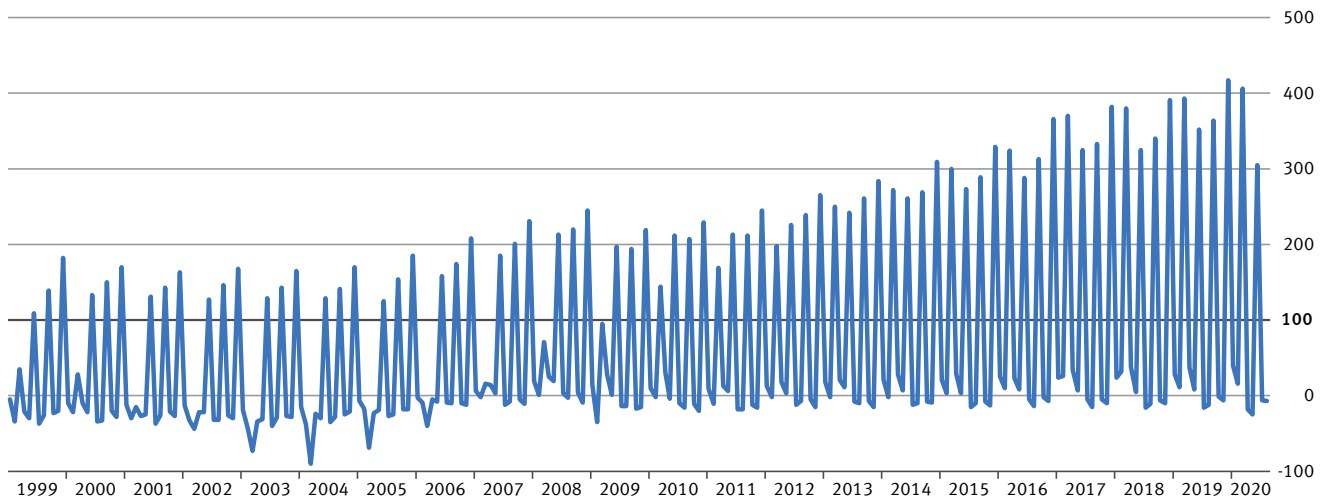
↳ Grafik 5 zeigt das Saisonmuster dieser Steuerart: Die Saisonhochs liegen am Ende der Quartale, also in den Monaten März, Juni, September und Dezember. Im März ändert sich jedoch der Ausschlag der rohen Saisonfaktoren über die Jahre hinweg: Sie sinken bis zum Jahr 2004 und steigen anschließend sehr stark an. Grund sind gesetzliche Änderungen: Von 1995 bis zum Jahresende 2005 wurden zur Herstellung oder Anschaffung einer Wohnung Eigenheimzulagen gewährt, die ausschließ-

lich im März aus den kassenmäßigen Steuereinnahmen der veranlagten Einkommensteuer ausbezahlt wurden. Seitdem wurden keine neuen Anträge angenommen und die Zahlungen sind dementsprechend zurückgegangen. Damit ist das Volumen der Einnahmen für März gestiegen und sind die steigenden Saisonausschläge erklärt.

Der Lösungsansatz zur präzisen Modellierung des Saisonmusters bei Änderungen der Rahmenbedingungen der Steuerarten ist wieder eine in zwei Zeiträume getrennte Saisonbereinigung: einerseits von Januar 1999 bis Dezember 2001 und andererseits ab Januar 2002. Aufgrund der geringen Datenlage für den erstgenannten Zeitraum unterliegt die Berechnung der Saisonfaktoren den Daten von Januar 1999 bis Dezember 2007, um das sich ändernde Saisonmuster im März am besten zu schätzen. Die Ergebnisse sind ähnlich wie für die Alkoholsteuer in Grafik 3 und daher nicht explizit dargestellt.

**Grafik 4**

Entwicklung der veranlagten Einkommensteuer<sup>1</sup>  
2015 = 100

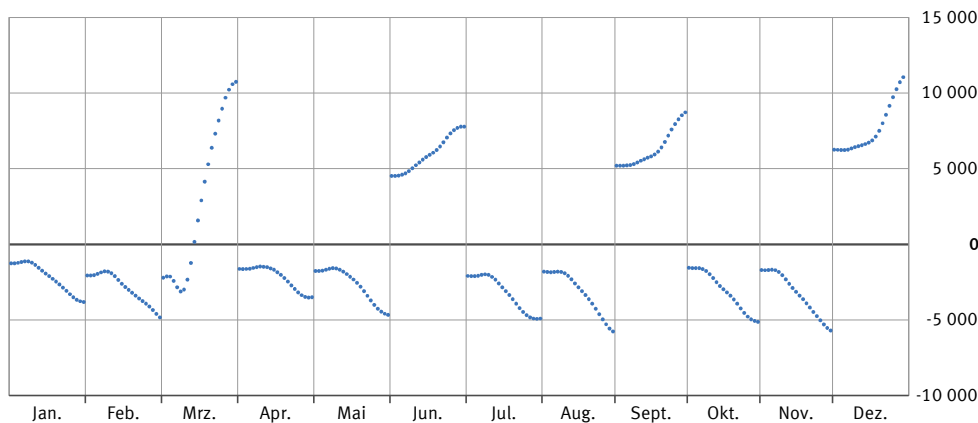


1 Nicht saisonbereinigte Werte.

2020 - 0548

**Grafik 5**

Rohe Saisonfaktoren in der veranlagten Einkommensteuer für den Zeitraum 1999 bis 2020



2020 - 0549

## Umsatzsteuer

Während wichtige ökonomische Variablen, zum Beispiel die monatliche Industrieproduktion, hauptsächlich von der Anzahl der Werktagen von Montag bis Freitag beeinflusst werden, hängt der Handelsumsatz vor allem von

der Anzahl an Tagen ab, an denen die Geschäfte geöffnet sind. Aus diesem Grund wird für die Umsatzsteuer, die vor allem im Handel anfällt, ein Kalenderregressor auf Basis von Verkaufstagen verwendet.

Aufgrund der Beziehung zwischen der Entstehung der Steuerschuld und der Einnahme aus der Steuer kann

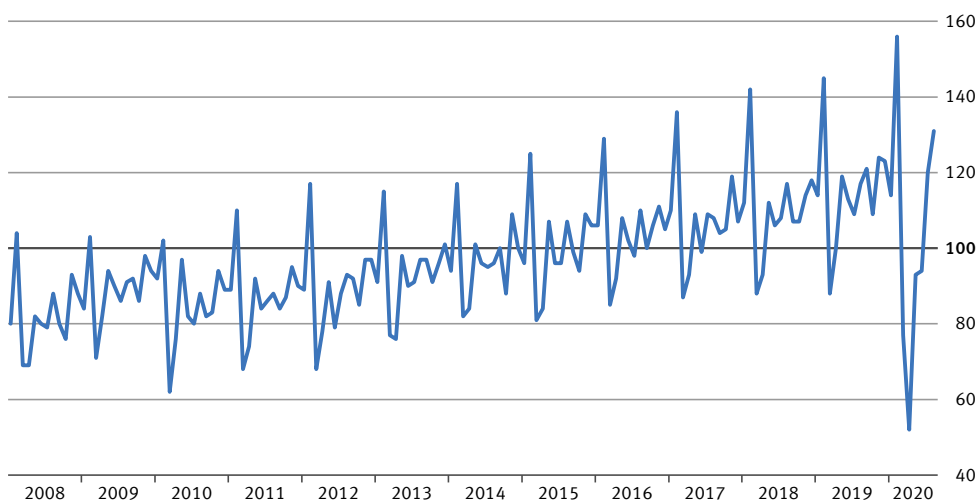
sich die Auswirkung der Kalenderkonstellationen bei den Einnahmen gegenüber der Entstehung verzögern. Die Zahlungsfrist der Umsatzsteuer liegt am zehnten Tag nach Ablauf jedes Voranmeldungszeitraums, bei Dauerfristverlängerungen einen Monat später.

Somit folgt die Entwicklung des Umsatzsteueraufkommens der wirtschaftlichen Entwicklung mit einer Verzögerung von ein bis zwei Monaten. Beispielsweise ist dann für die Berechnung von Kalendereffekten im Februar die Anzahl der Verkaufstage im Januar und Dezember relevant. Bei der Umsatzsteuer werden daher zeitlich verschobene Varianten von Kalenderregressoren mit einer Verschiebung von ein und zwei Monaten verwendet.<sup>8</sup> Darüber hinaus finden spätestens am 10. Februar Sondervorauszahlungen statt, weshalb bei der Umsatzsteuer positive saisonale Ausschläge im Februar zu beobachten sind. [↘ Grafik 6](#)

8 Zeitlich verschobene Varianten von Kalenderregressoren auf Basis von Arbeitstagen und Verkaufstagen werden zusätzlich für die folgenden Steuerarten verwendet: Einfuhrumsatzsteuer, Tabaksteuer, Alkoholsteuer, Energiesteuer und Biersteuer.

### Grafik 6

Entwicklung des Umsatzsteueraufkommens<sup>1</sup>  
2015 = 100



1 Nicht saisonbereinigt.

## 5

### Ergebnisse

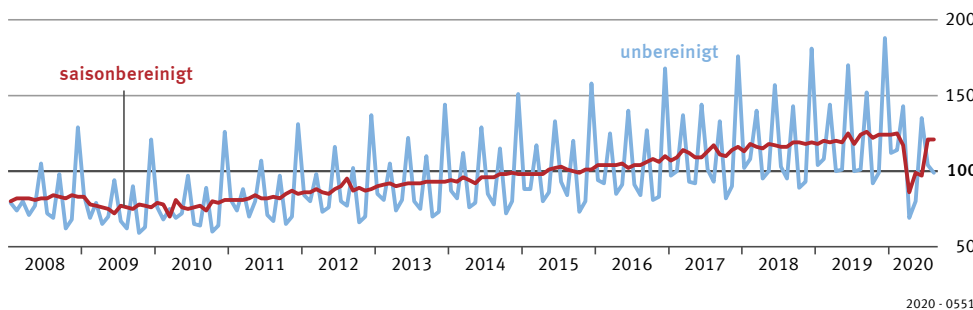
Die für die Saisonbereinigung ausgewählten Steuerarten sind in insgesamt 21 Zeitreihen untergliedert. Dabei ergeben sich folgende Steuerarten aus einer indirekten Saisonbereinigung: Gemeinschaftsteuern, Kapitalertragsteuer, Steuern vom Umsatz und vom Einkommen. Das bedeutet, dass sie keiner eigenen Bereinigung unterliegen, sondern sie lassen sich aus der Summe der einzelnen Steuerarten berechnen (siehe Tabelle 1).

Durch die Coronakrise treten extreme Werte in den unbereinigten kassenmäßigen Steuereinnahmen auf. Für die Berechnung der Kalender- und Saisonfaktoren wurde daher zunächst der Zeitraum Januar 1999 bis Februar 2020, dem Monat vor Ausbruch der Corona-Pandemie, verwendet. Dieser Lösungsansatz im Umgang mit der Corona-Pandemie schließt Einflüsse der Krise auf das Saisonmuster aus. Der Einfluss der Krise wird so vollständig in den saisonbereinigten Ergebnissen abgebildet.

Die Gemeinschaftsteuern machten im Jahr 2019 gut 80 % der gesamten monatlichen kassenmäßigen Steuer-

**Grafik 7**

Entwicklung der kassenmäßigen Einnahmen der Gemeinschaftsteuern  
2015 = 100



einnahmen (ohne Gemeindesteuern) aus. Die Grundlage der Berechnung der Gemeinschaftsteuern sind Umsatz und Einkommen der gesamten Wirtschaft; damit steht ihre Entwicklung in engem Zusammenhang mit den wirtschaftlichen Rahmenbedingungen. Sie sind somit ein wichtiger Konjunkturindikator.

↳ Grafik 7 zeigt die Entwicklung der saisonbereinigten kassenmäßigen Einnahmen aus Gemeinschaftsteuern für den Zeitraum Januar 2008 bis August 2020. In Blau ist die unbereinigte Zeitreihe dargestellt, sie weist ein deutliches Saisonmuster mit ausgeprägten Ausschlägen im Dezember auf. Die rote Kurve gibt den Indikator für die saisonbereinigten Gemeinschaftsteuern an. Die saisonbereinigte Zeitreihe weist eine deutlich geringere Volatilität als die unbereinigte Zeitreihe auf. So ist zum Beispiel der Beginn der Auswirkungen der Coronakrise auf das Aufkommen aus diesen Steuern besser zu erkennen. Von Februar 2020, dem Monat vor dem Beginn der Einschränkungen durch die Corona-Pandemie in Deutschland, bis April 2020 ist dieser Indikator um 31 % gesunken. Der Einbruch beruht überwiegend auf gewährten Stundungen der Umsatz-, veranlagten Einkommen- und Körperschaftsteuer aufgrund der eingeschränkten Zahlungsfähigkeit der Steuerzahler. Bis August 2020 stieg der Indikator auf 3% unter Vorkrisenniveau, insbesondere im Juli 2020 hat er sich um 25 % gegenüber dem Vormonat erholt. Dies lag vor allem daran, dass die zuvor für den April gewährten Stundungen der Steuerzahlungen in diesem Monat fällig wurden.

Weiterhin zeigt die Zeitreihe die negative Auswirkung der Finanzkrise ab Januar 2009 auf das Aufkommen aus Gemeinschaftsteuern, das sich erst im Mai 2011 wieder vollständig erholt hatte.

Die Gemeinschaftsteuern setzen sich aus den Steuern vom Einkommen und vom Umsatz zusammen.

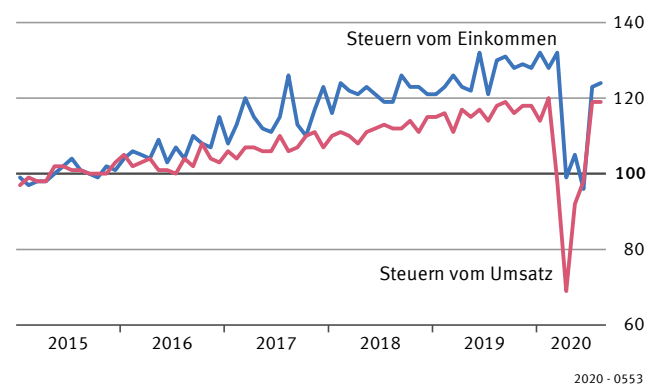
Die Steuern vom Umsatz brachen im März 2020 um 18% und im April 2020 um 29% ein, jeweils im Vergleich zum Vormonat. Die Steuern vom Einkommen gingen erst im April 2020 zurück, und zwar um 23% im Vergleich zum Februar 2020. Insbesondere im Juli 2020 zeigten beide Steuerarten eine Erholung im Vergleich zum Vormonat: Die Steuern vom Umsatz stiegen um 22%, die Steuern vom Einkommen um 28%. Im August 2020 lagen die Steuern vom Einkommen noch um 1% und die Steuern vom Umsatz um 4% unter dem Vorkrisenniveau vom Februar 2020.

↳ Grafik 8

**Grafik 8**

**Grafik 8**

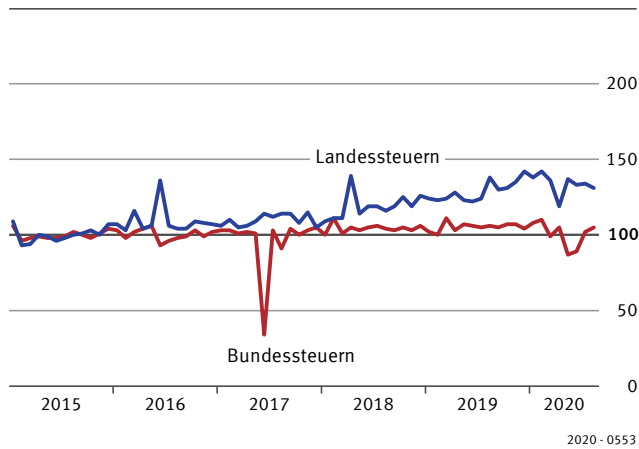
Saisonbereinigte Einkommensteuer und Steuern vom Umsatz  
2015 = 100



Die Einnahmen aus Bundes- und Landesteuern lagen im August 2020 rund 5 und 8% niedriger im Vergleich zu Februar 2020. Die Bundessteuern sind im Mai 2020 um 21% und die Landessteuern im April 2020 um 16% gegenüber Februar 2020 eingebrochen. Der starke Rückgang bei den Bundessteuern im Juni 2017 ist auf eine einmalige Rückzahlung zurückzuführen, da zuvor die Kernbrennstoffsteuer als verfassungswidrig eingestuft worden war. ↳ Grafik 9

**Grafik 9**

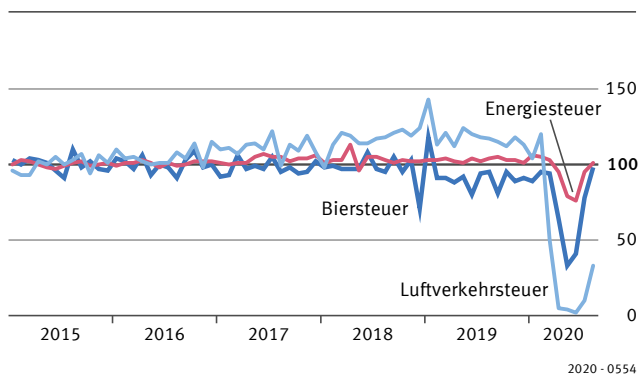
Einnahmen aus Bundes- und Landessteuern (saisonbereinigt)  
2015 = 100



Zur Verdeutlichung, wie sich die Maßnahmen zur Eindämmung der Corona-Pandemie auswirken, betrachtet [Grafik 10](#) die saisonbereinigte Entwicklung ausgewählter Bundes- und Landessteuern.

**Grafik 10**

Einnahmen ausgewählter Bundes- und Landessteuern (saisonbereinigt)  
2015 = 100



Die Luftfahrt ist von der Coronakrise und den Maßnahmen zur Eindämmung der Pandemie besonders betroffen; dies spiegelt sich deutlich in den Einnahmen aus der Luftverkehrsteuer wider: Im März und April 2020 sind sie insgesamt um 96 % eingebrochen und zeigen erst im Juli 2020 eine leichte Erholung. Des Weiteren sind die Bier- und Energiesteuer dargestellt, deren Tiefpunkte im Mai und Juni 2020 bei 56 % und 28 % im Vergleich zu Februar 2020 lagen. Hier zeigte sich allerdings eine Erholung bis August 2020 auf etwa das Vorkrisenniveau.

## 6

### Fazit

Seit August 2020 werden die Statistiken zur Entwicklung der monatlichen kassenmäßigen Steuereinnahmen zusätzlich auch in kalender- und saisonbereinigter Form zur Verfügung gestellt. Die Ergebnisse liegen im Durchschnitt 26 Tage nach Ende eines Berichtsmonats vor. Sie sind ein zusätzlicher Bestandteil des Veröffentlichungsprogramms der Statistiken zum Steueraufkommen.

Die saisonbereinigten kassenmäßigen Steuereinnahmen des Bundes und der Länder können einen wichtigen Beitrag zur Beurteilung der Finanzlage des Staates leisten. Da insbesondere die Gemeinschaftsteuern die wirtschaftliche Aktivität widerspiegeln, können diese darüber hinaus als Konjunkturindikator in Betracht gezogen werden. Im Verlauf der Corona-Pandemie sind die Einflüsse der Krise auf die Entwicklungstendenzen in den kassenmäßigen Steuereinnahmen besser darstellbar.

Eine besondere Herausforderung für die Saisonbereinigung der kassenmäßigen Steuereinnahmen besteht darin, dass sich die Steuerschuld oft zeitverzögert in den kassenmäßigen Steuereinnahmen niederschlägt und sich die Wirkung von Kalendereffekten gegenüber der Entstehung verzögern kann. Daher werden zeitlich verschobene Varianten von Kalenderregressoren bei der Umsatzsteuer, der Einfuhrumsatzsteuer, der Tabaksteuer, der Alkoholsteuer, der Energiesteuer und der Biersteuer mit einer Verschiebung von ein und zwei Monaten verwendet.

Darüber hinaus wirken sich die häufig vorkommenden Änderungen des Steuerrechts zum Teil um mehrere Jahre verzögert auf die Entwicklung der kassenmäßigen Steuereinnahmen aus. Steuerrechtsänderungen können das Saisonmuster in der Entwicklung der kassenmäßigen Steuereinnahmen grundlegend ändern, was über die Zeit stark veränderte Saisoneffekte erzeugt und somit eine präzise Abbildung erschwert. Ein Lösungsansatz zur Modellierung des Saisonmusters ist hier eine in verschiedene Zeiträume getrennte Saisonbereinigung. [\[1\]](#)

## LITERATURVERZEICHNIS

---

Bundesministerium der Finanzen. *Steuern von A bis Z. Ausgabe 2019*. [Zugriff am 16. November 2020]. Verfügbar unter: [www.bundesfinanzministerium.de](http://www.bundesfinanzministerium.de)

Deutsche Bundesbank. *Kalendarische Einflüsse auf das Wirtschaftsgeschehen*. In: Monatsbericht Dezember 2012, Seite 53 ff.

Eurostat. *ESS guidelines on seasonal adjustment*. Luxemburg 2015. [doi: 10.2785/317290]

Linz, Stefan/Fries, Claudia/Völker, Julia. *Saisonbereinigung der Konjunkturstatistiken mit X-12-ARIMA und mit X13 in JDemetra+*. In: WISTA Wirtschaft und Statistik. Ausgabe 4/2018, Seite 59 ff.

Statistisches Bundesamt. *Qualitätsbericht: Statistik über den Steuerhaushalt 2020*. [Zugriff am 16. November 2020]. Verfügbar unter: [www.destatis.de](http://www.destatis.de)

# KORREKTUR DES TÄTIGKEITS- SCHLÜSSELS DER BUNDESAGENTUR FÜR ARBEIT MITHILFE MASCHINELLER LERNVERFAHREN

Einsatz der Verdienststrukturerhebung

Dr. Elena Schmidt

↳ **Schlüsselwörter:** Maschinelles Lernen – Boosting – Tätigkeitsschlüssel – Verdienststrukturerhebung

## ZUSAMMENFASSUNG

Ziel des vorgestellten Projekts war, das Merkmal Vollzeit/Teilzeit des Tätigkeitsschlüssels in den Integrierten Erwerbsbiografien der Bundesagentur für Arbeit zu korrigieren. Dies sollte mithilfe des auch in der Verdienststrukturerhebung vorhandenen, aber manuell korrigierten Schlüssels erfolgen. Unter Berücksichtigung verschiedener projektspezifischer Anforderungen wurde ein überwachtes maschinelles Lernverfahren eingesetzt, welches das entsprechende Merkmal eines Beschäftigten anhand vorliegender Betriebs- und Mitarbeitermerkmale schätzen kann. Es zeigte sich, dass sich mit diesem Modell der Fehler in der Signierung des Tätigkeitsschlüssels bei einer aus dem Datensatz der Verdienststrukturerhebung erzeugten Testmenge um etwa 40% reduzieren lässt.

↳ **Keywords:** machine learning – boosting – occupational code number – structure of earnings survey

## ABSTRACT

*The aim of this project was to correct the “part-time/full-time” variable component of the occupational code number in the Integrated Labour Market Biographies of the Federal Employment Agency. The basic idea was to use the code number which is also employed in the structure of earnings survey but corrected manually. Against the background of various project-specific requirements, a supervised machine learning method was applied which can estimate the relevant variable for an employee based on the variables available for the local unit and the employee. A test of the model on a data subset of the structure of earnings survey showed that it can reduce the coding error in the occupational code number by about 40%.*



**Dr. Elena Schmidt**

ist Bioinformatikerin und als wissenschaftliche Mitarbeiterin im Referat „Maschinelles Lernen und Imputationsverfahren“ des Statistischen Bundesamtes tätig. Ihre Aufgabenschwerpunkte sind die Weiterentwicklung von maschinellen Lernverfahren und ihre Anwendung auf Daten der amtlichen Statistik.



## 1

---

### Einleitung

---

In den Integrierten Erwerbsbiografien (IEB) der Bundesagentur für Arbeit (BA) liegt die Information über die Arbeitszeit eines Beschäftigten als kategoriales Merkmal (Vollzeit/Teilzeit) innerhalb des Tätigkeitsschlüssels vor. Eine Änderung der statistischen Erfassung von Teilzeitarbeit (Bertat und andere, 2013) im Jahr 2011 machte deutlich, dass eine Untererfassung von Teilzeitbeschäftigten in den Jahren vor 2011 vorlag. Dies ist problematisch, da die Integrierten Erwerbsbiografien oft die Grundlage für Arbeitsmarktanalysen bilden (Fitzenberger/Seidlitz, 2020) und die Untererfassung zu fehlerhaften Auswertungen wie der Überschätzung des Niedriglohnanteils führt. Aus diesem Grund wurden bereits verschiedene Ansätze zur Korrektur entwickelt (Möller, 2016; Fitzenberger/Seidlitz, 2019).

In der Verdienststrukturerhebung (VSE) 2014 (Statistisches Bundesamt, 2016) hingegen liegen neben dem BA-Tätigkeitsschlüssel auch die bezahlten Arbeitsstunden als Erhebungsmerkmal vor. Diese konnten bei einer manuellen Korrektur im Rahmen der Plausibilitätsprüfung des Tätigkeitsschlüssels in der Verdienststrukturerhebung 2014 mitberücksichtigt werden. Somit ist von einer hohen Qualität der VSE-Daten auszugehen. Eine direkte Übertragung des korrigierten Schlüssels der VSE-Daten zur entsprechenden Korrektur der IEB-Daten ist allerdings nicht möglich, da es sich bei der Verdienststrukturerhebung um eine Stichprobe handelt. Aus Geheimhaltungsgründen wäre es aber auch rechtlich untersagt.

Das Ziel des Projekts war, den manuell korrigierten Tätigkeitsschlüssel des VSE-Datensatzes für eine Korrektur des entsprechenden Schlüssels der Integrierten Erwerbsbiografien zu nutzen. Die korrigierten VSE-Daten wurden als Trainingsdaten für ein überwacht maschinelles Lernverfahren (ML-Verfahren) verwendet, welches das Merkmal Vollzeit/Teilzeit einer oder eines Beschäftigten anhand vorliegender Unternehmens- und Mitarbeitermerkmale schätzen kann. Dieses ML-Modell soll dem Institut für Arbeitsmarkt- und Berufsforschung der Bundesagentur für Arbeit zur Verfügung gestellt und dort auf die IEB-Daten angewendet werden, um falsche Kodierungen im Tätigkeitsschlüssel aufzudecken.

Kapitel 2 beschreibt zunächst den für die Berechnung des ML-Modells verwendeten Datensatz und erläutert die Besonderheiten und sich daraus ergebenden Anforderungen des vorliegenden Problems. Kapitel 3 geht näher auf die Grundlagen der verwendeten Methoden, das maschinelle Lernverfahren Tree Boosting und die zur Modellbewertung eingesetzte Metrik ein. Kapitel 4 stellt das Vorgehen im Hinblick auf die in Kapitel 2 genannten Anforderungen genauer dar. Die Kapitel 5 und 6 zeigen auf, mit welchem Erfolg der Einsatz des verwendeten maschinellen Lernverfahrens möglich ist, und weisen auf weitere Optionen bei dessen Einsatz hin.

## 2

---

### Daten und Problemstellung

---

Das Training erfolgte auf dem VSE-Datensatz und das Modell soll anschließend auf den IEB-Daten angewendet werden. Daher waren nur die Merkmale, die in beiden Datensätzen enthalten sind, für das Training nutzbar. Diese sind die Betriebsmerkmale „Anzahl der Mitarbeiter eines Betriebes“, „Bundesland“, „Landkreis“ und „Wirtschaftszweig“ und die Mitarbeitermerkmale „Geschlecht“, „Alter“, „Dauer der bisherigen Beschäftigung“, „Bruttomonatsverdienst“, „Personengruppenschlüssel“ und „Tätigkeitsschlüssel“. Letzterer liefert Informationen über den Beruf, die Spezialisierung, das Arbeitsniveau, die Schulbildung, die Berufsausbildung und das Arbeitsverhältnis (einschließlich der fehlerhaften Vollzeit/Teilzeit-Kodierung).

Eine festgestellte Untererfassung der Teilzeitbeschäftigten wirkt sich einschränkend auf die Niedriglohnberechnung aus. Daher ist für die Auswertung der IEB-Daten für das Institut für Arbeitsmarkt- und Berufsforschung der Fall der Teilzeitbeschäftigten, die fälschlicherweise als Vollzeitbeschäftigte geführt werden, von besonderer Bedeutung. Die hier gezeigten Rechnungen beziehen sich ausschließlich auf diesen Fall, es wurden daher nur in der Verdienststrukturerhebung geführte Beschäftigte in den Datensatz aufgenommen, die vor der manuellen Korrektur als Vollzeitkräfte geführt wurden.

Um eine Überanpassung eines ML-Modells an den Trainingsdatensatz zu erkennen, ist es üblich, den für die Erstellung eines Modells zur Verfügung stehenden Datensatz in einen Trainings- und einen Testdatensatz

aufzuteilen. Der Trainingsdatensatz wird für die Berechnung des Modells verwendet, mit dem Testdatensatz wird die Klassifikationsgüte des Modells überprüft. Dies führte bei einer 80%/20%-Teilung des Datensatzes zu einem Trainingsdatensatz mit 389 298 Beschäftigten und einem Testdatensatz mit 97 324 Beschäftigten.

Aus der beschriebenen Situation und der Beschaffenheit der Daten ergaben sich eine Reihe von Anforderungen für die Umsetzung des Projekts:

1. **Geheimhaltung.** Da das Ziel des Projekts die Berechnung eines ML-Modells war, das vom Institut für Arbeitsmarkt- und Berufsforschung eingesetzt werden kann, musste gewährleistet sein, dass mit dem Modell keine Einzeldaten aus der Verdienststrukturerhebung weitergegeben werden.
2. Die Struktur eines Trainingsdatensatzes sollte möglichst ähnlich zu den Daten sein, auf die ein entsprechendes Modell angewendet werden soll. Daher war zu berücksichtigen, dass es sich bei den IEB-Daten um eine Vollerhebung handelt, bei der Verdienststrukturerhebung dagegen um eine Stichprobe.
3. Die Teilzeitbeschäftigten machen nur etwa 5 % der Beschäftigten im Datensatz aus. Wenn eine vorherzusagende Klasse im Datensatz sehr viel stärker vertreten ist als die andere, spricht man von einem „Imbalanced Data Problem“. Beim Trainieren eines ML-Modells kann in einem solchen Fall ein Bias in Richtung der Mehrheitsklasse auftreten. Diesem Effekt sollte, falls nötig, entgegengewirkt werden.
4. Da das Ziel die Korrektur des Tätigkeitsschlüssels im IEB-Datensatz war, war zu berücksichtigen, dass der Gesamtanteil der fehlerhaft kodierten Tätigkeitsschlüssel nach der Korrektur mithilfe des ML-Modells geringer sein sollte als vorher.

## 3

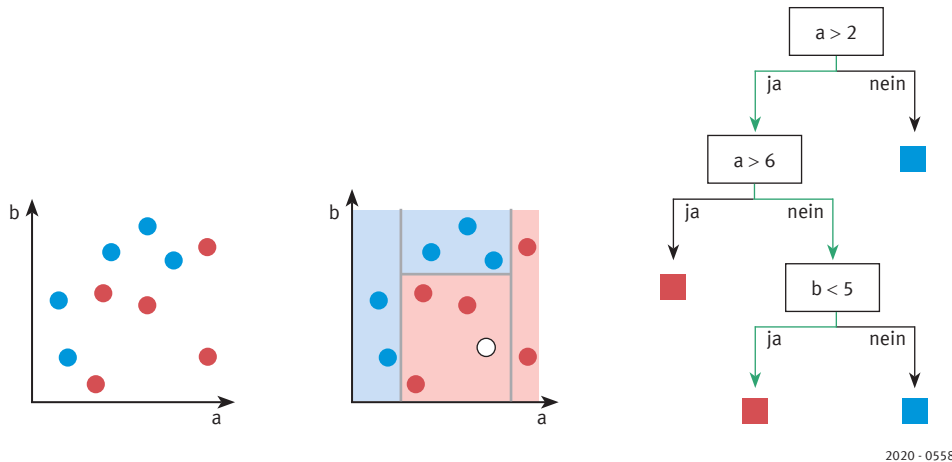
### Methodik

Die Anforderung, dass keine Einzeldaten mit dem berechneten ML-Modell weitergegeben werden durften, schränkte die Auswahl der einsetzbaren ML-Verfahren bereits ein. Support Vector Machines (SVMs) können diese Anforderung nicht erfüllen, da einzelne Datenpunkte Teil der Modelldefinition und damit bei einer Weitergabe des Modells aus diesem auslesbar sind. Für den Vergleich mit den anderen Verfahren wurden SVMs aber zunächst mitberücksichtigt. Für die Modellerstellung wurden letztlich Testrechnungen mit den vier verschiedenen (ML-)Verfahren Logistische Regression, SVMs, Random Forests und Gradient Boosting durchgeführt. Gemessen an der zu optimierenden Zielgröße konnte mit Gradient Boosting das beste Modell berechnet werden, sodass die weiteren Untersuchungen mit diesem Verfahren durchgeführt wurden.

#### 3.1 Gradient Boosting

Boosting (Schapire, 1990; Freund/Schapire, 1996; Breiman, 1998; Schapire/Freund, 2012) ist ein maschinelles Lernverfahren, bei dem eine Menge von schwachen Klassifikatoren (Base Classifiers) zu einem neuen starken Klassifikator kombiniert wird. Es gehört zu den überwachten Lernverfahren. Populär und in vielen ML-Wettbewerben erfolgreich ist in den letzten Jahren das Tree Boosting geworden (Chen/Guestrin, 2016), bei dem Entscheidungsbaume (Breiman und andere, 1984) als Base Classifiers eingesetzt werden. Die Grundidee eines Entscheidungsbaums ist, ein Regelwerk zur Entscheidungsfindung in Form eines Binärbaums zur Verfügung zu stellen, der auf einem Trainingsdatensatz basierend erstellt wird. Der Entscheidungsbaum kann anschließend dazu genutzt werden, neue unbekannte Datenpunkte zu klassifizieren. Das Beispiel in [↘ Grafik 1](#) zeigt einen Datensatz, für dessen Datenpunkte zwei Merkmale, a und b, gegeben sind und die zwei unterschiedlichen Klassen angehören, der positiven Klasse (blau) und der negativen Klasse (rot). Der aus dem Datensatz erzeugte Entscheidungsbaum enthält sogenannte Knoten. An diesen werden bei der Klassifikation eines neuen Datenpunkts

Grafik 1  
Entscheidungsbaum



(weiß) bei der Wurzel<sup>1</sup> beginnend dessen Merkmalsausprägungen abgefragt, bis ein Blatt<sup>2</sup> des Baumes erreicht wird, wo schließlich eine der beiden Klassen, blau oder rot, dem Datenpunkt zugewiesen wird.

Wie mit Entscheidungsbäumen können mit Tree Boosting Modelle für Klassifikations- und Regressionsprobleme berechnet werden. Im Gegensatz zu Random Forests (Breiman, 2001), die ebenfalls auf Entscheidungsbäumen basieren, werden die einzelnen Klassifikatoren beim Tree Boosting sukzessive trainiert, wobei jeder neu trainierte Klassifikator vom vorherigen abhängt. Dies wird so umgesetzt, dass allen Datenpunkten für das Training des ersten Klassifikators zunächst das gleiche Gewicht, welches für den Algorithmus der aktuellen Wichtigkeit einer korrekten Zuordnung dieses Datenpunktes entspricht, zugeordnet wird. Für das Training der folgenden Klassifikatoren werden die Gewichte jeweils so angepasst, dass die Gewichte der vom vorherigen Klassifikator falsch zugeordneten Datenpunkte erhöht werden. Beim Tree Boosting werden Entscheidungsbäume geringer Höhe<sup>3</sup> eingesetzt. Um die vom Gesamtmodell vorhergesagte Klasse für einen neuen Datenpunkt zu erhalten, wird unter Einbeziehung der Klassifikationsergebnisse aller berechneten Bäume abgestimmt. Die Stimmen der einzelnen Bäume gehen mit einer baumspezifischen Gewichtung in das Ergebnis

ein. Die Gewichtung berechnet sich aus ihrem individuellen Fehler auf den Trainingsdaten.

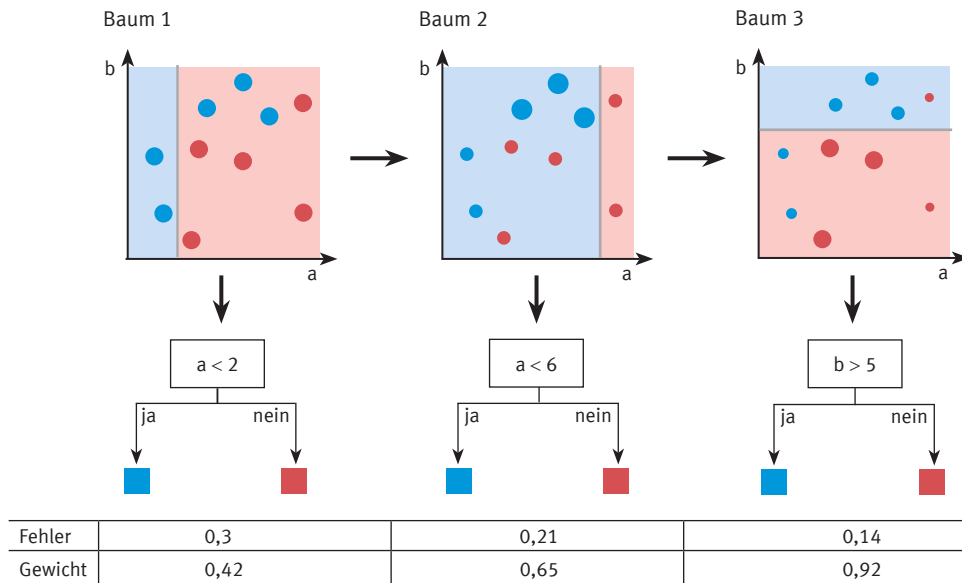
Im vorliegenden Beispiel (übernommen aus Schapire/Freund, 2012) soll nun das Boosting-Modell lernen, die richtige Klasse aus dem vorigen Beispiel anhand der gegebenen Merkmale vorherzusagen. Im ersten Trainingszyklus erhalten alle Datenpunkte des Trainingsdatensatzes das gleiche Gewicht (in [Grafik 2](#) durch die Größe der jeweiligen Datenpunkte dargestellt). Der erste Entscheidungsbaum wird unter Berücksichtigung dieser Gewichte erstellt und anschließend der Fehler für diesen Baum und darauf basierend sein Gewicht berechnet. Für das Training des zweiten Entscheidungsbaums werden die Gewichte der Datenpunkte, die vom ersten Baum falsch klassifiziert werden, erhöht (blaue Datenpunkte, die bei Baum 1 in den roten Klassifikationsbereich fallen) und die Gewichte der richtig klassifizierten Datenpunkte verringert. Nach demselben Prinzip werden die Gewichte der Datenpunkte für das Training des dritten Entscheidungsbaums basierend auf der Klassifikation des zweiten angepasst (Erhöhung der Gewichte der drei roten Datenpunkte, die in den blauen Klassifikationsbereich von Baum 2 fallen, Verringerung der übrigen Gewichte).

Soll das trainierte Boosting-Modell dazu eingesetzt werden, einen neuen Datenpunkt (weiß in [Grafik 3](#)) mit den Merkmalen  $a = 5$  und  $b = 2$  zu klassifizieren, werden alle trainierten Entscheidungsbäume einzeln ausgewertet, sodass jeder Baum unabhängig eine Klasse für den unbekanntem Datenpunkt vorhersagt.

1 Die Wurzel ist der Knoten, in den keine Kante hineinführt.  
 2 Ein Blatt ist ein Knoten, aus dem keine Kante hinausführt.  
 3 Die Höhe eines Entscheidungsbaums ist der längste Weg von der Wurzel zu einem Blatt.

**Grafik 2**

Erstellung eines Baum-Ensembles beim Boosting



2020 - 0559

Unter Berücksichtigung der jeweiligen Baumgewichte werden die Stimmen der drei Entscheidungsbäume für die positive (blau) und negative (rot) Klasse verrechnet und so die Klasse des neuen Datenpunkts bestimmt:

$$- 0,42 + 0,65 - 0,92 = - 0,69$$

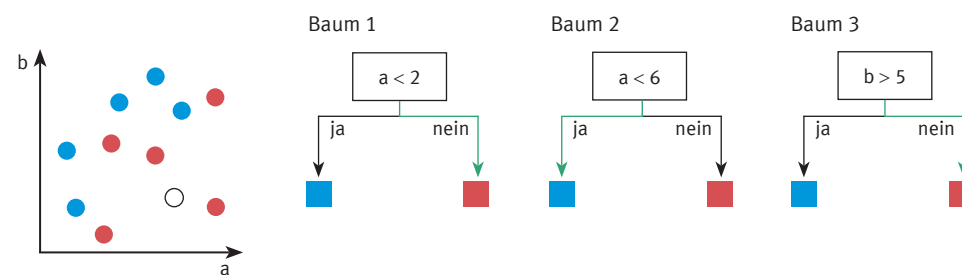
Der neue Datenpunkt wird somit der negativen Klasse (rot) zugeordnet.

Für die Berechnungen der für das vorliegende Projekt erzeugten Modelle wurde das in der Programmier-

sprache R verfügbare XGBoost-Paket (Chen/Guestrin, 2016) verwendet. Dieses bietet die Möglichkeit, über eine Vielzahl von Parametern Einfluss auf die Modellberechnung zu nehmen (Readthedocs Webseite von XGBoost, 2020). Um das beste Modell für den verwendeten Datensatz zu erhalten, wurde ein Teil dieser Parameter über eine Gittersuche variiert und so die optimalen Parameterwerte gesucht. Empfehlungen für zu variierende Parameter und deren Wertebereiche sind verschiedenen Quellen im Internet zu entnehmen (beispielsweise Hackerearth, 2020; Jain, 2016).

**Grafik 3**

Klassifizierung eines neuen Datenpunkts beim Boosting



2020 - 0560

### 3.2 Bewertung

Für die für den Vergleich notwendige Bewertung von ML-Modellen zur Klassifikation gibt es zahlreiche Möglichkeiten. Die Sensitivity gibt den Anteil der richtig als positiv klassifizierten Datenpunkte an allen tatsächlich positiven an, die Precision den Anteil der richtig als positiv klassifizierten an allen, die als positiv vorhergesagt wurden. Die Specificity gibt den Anteil der richtig als negativ klassifizierten an allen tatsächlich negativen an. Das Standardmaß ist die Accuracy (acc), die dem Anteil der richtig zugeordneten Datenpunkte an allen zugeordneten Datenpunkten entspricht. Gerade im Fall von starken Größenunterschieden der Klassen im Trainingsmaterial können aber bei der Accuracy unerwünschte Effekte auftreten. Geht man wie im vorliegenden Fall beispielsweise von einem 95 %-Anteil der Mehrheitsklasse im Gegensatz zu einem 5 %-Anteil der Minderheitsklasse aus, so würde ein Klassifikator, der grundsätzlich immer der Mehrheitsklasse zugeordnet wird, also völlig wertlos ist, eine Accuracy von 0,95 erreichen und damit eine gute Klassifikationsgüte suggerieren. Daher sollte zur Accuracy die „No Information Rate“ hinzugezogen werden, die angibt, welche Accuracy erreicht wird, wenn immer die Mehrheitsklasse vorhergesagt wird. Die Accuracy des Modells sollte diesen Wert übersteigen.

Grafik 4

Metriken zur Modellbewertung

		Tatsächliche Klasse	
		Positiv (P)	Negativ (N)
Vorhergesagte Klasse	Positiv	Richtig positiv (TP)	Falsch positiv (FP)
	Negativ	Falsch negativ (FN)	Richtig negativ (TN)

$$acc = \frac{TP + TN}{P + N}$$

$$sensitivity = \frac{TP}{TP + FN} = \frac{TP}{P}$$

$$specificity = \frac{TN}{TN + FP} = \frac{TN}{N}$$

$$precision = \frac{TP}{TP + FP}$$

$$bacc = \frac{sensitivity + specificity}{2} = \left( \frac{TP}{P} + \frac{TN}{N} \right) / 2$$

$$F1score = 2 \cdot \frac{sensitivity \cdot precision}{sensitivity + precision} = \frac{2TP}{2TP + FP + FN}$$

2020 - 0561

Andere Maße wie die Balanced Accuracy (bacc) oder der F1score bewerten ebenfalls in einem Wertebereich zwischen 0 und 1 das beste Modell mit 1. Sie gewichten die „Falsch Negativen“ nicht in dieser Form und liefern für das genannte Beispiel die Werte 0,5 beziehungsweise 0. Sie werden daher oft im Fall eines Imbalanced Data Szenarios eingesetzt. [↪ Grafik 4](#)

## 4

### Vorgehen

#### 4.1 Geheimhaltungsaspekt

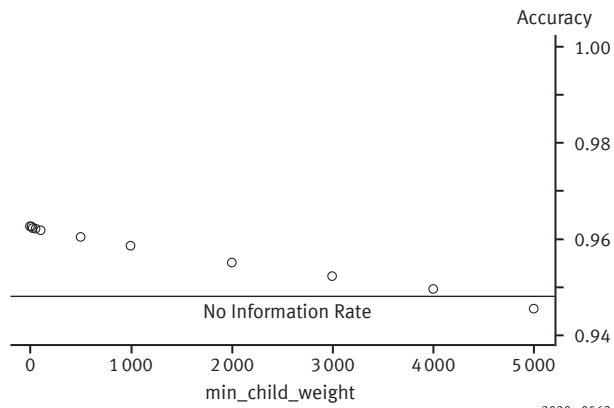
Auch beim Einsatz von Verfahren, die auf Entscheidungsbäumen basieren, ist es theoretisch möglich, anhand des berechneten Modells Rückschlüsse auf Einzeldaten zu ziehen, sofern ein Baum nur einen einzelnen Datenpunkt in einem Blatt enthält. Um das zu verhindern, ist sicherzustellen, dass sich eine Mindestzahl von Datenpunkten in jedem Blatt des Entscheidungsbaums befindet. Dies kann über einen Parameter beim Training des Modells gesteuert werden. In Absprache mit der für die Verdienststrukturerhebung zuständigen Fachabteilung wurden für die vorliegenden Rechnungen 25 Datenpunkte je Blatt als Minimum gewählt.

Dieser Parameter wird normalerweise für die Regularisierung<sup>4</sup> des Modells eingesetzt. Um seinen Einfluss auf die Accuracy der Ergebnisse einschätzen zu können, wurde eine Testreihe erstellt, bei der der betreffende Parameter zwischen 1 und 5 000 variiert wurde. Die Ergebnisse zeigen, dass die Verschlechterung der Accuracy im Bereich zwischen 1 und 25 zu vernachlässigen ist. Erst bei einem Parameterwert von über 4 000 erreicht die Accuracy den Wert der hier vorliegenden „No Information Rate“, bei dem ein Modell keinen Mehrwert bietet im Vergleich zu einem Schätzer, der immer die Mehrheitsklasse vorhersagt. [↪ Grafik 5](#)

4 Dieser Begriff bezeichnet die Verringerung der Komplexität eines Modells, um eine Überanpassung an die Daten zu verhindern.

**Grafik 5**

Einfluss des XGBoost-Parameters `min_child_weight` auf die Accuracy des Modells



2020 - 0562

## 4.2 Rückschluss von der Stichprobe auf den Gesamtdatensatz

Wie bereits erwähnt, handelt es sich bei der Verdienststrukturhebung im Gegensatz zu den Daten der Bundesagentur für Arbeit um eine geschichtete Stichprobe. Dieser Stichprobe entsprechend sind den Datenpunkten Hochrechnungsfaktoren zugeordnet, mit denen auf die Gesamtdaten geschlossen werden kann, wie sie auch von der Bundesagentur für Arbeit erhoben werden. Da das entwickelte ML-Modell auf die BA-Daten angewendet werden soll, ist es sinnvoll, die Hochrechnungsfaktoren bei der Modellberechnung mit einzubeziehen. Hierfür boten sich zwei Möglichkeiten: Der erste Fall berücksichtigt die Hochrechnungsfaktoren bei der Berechnung der verwendeten Gütemaße, indem er die einzelnen Datensätze mit den Faktoren gewichtet. Dies erfolgt, um beim Testen eine Verteilung mit den VSE-Daten abzubilden, die möglichst ähnlich zu der von den bei der Bundesagentur für Arbeit vorliegenden Daten ist. So vorgegangen wurde sowohl bei der Kreuzvalidierung im Rahmen des Parameter-Tunings wie auch bei der finalen Bewertung des Modells mithilfe des Testdatensatzes. Eine zweite Möglichkeit bietet der verwendete Gradient-Boosting-Algorithmus XGBoost, indem er die Einbeziehung von datenpunktspezifischen Gewichten bei der Modellerstellung ermöglicht. Diese Variante wurde getestet, hatte aber keinen statistisch signifikanten Einfluss auf die Ergebnisse und wurde für die Berechnung der finalen Modelle nicht eingesetzt.

## 4.3 Imbalanced-Data-Problem

Es gibt verschiedene Möglichkeiten, dem Effekt des Imbalanced-Data-Problems entgegenzuwirken. Für das vorliegende Projekt wurden drei Methoden getestet. Die erste Methode ist die Variation verschiedener Parameter, die die Behandlung der verschiedenen Klassen bei der Modellberechnung mit XGBoost beeinflussen. Die zweite Methode ist das sogenannte Downsampling der entsprechenden Mehrheitsklasse, bei dem das Verhältnis der Klassen im Trainingsdatensatz zueinander durch Reduktion der Datenpunkte der Mehrheitsklasse angeglichen wird. Die dritte Methode ist die Wahl einer entsprechenden Metrik für die Modellbewertung.

Die Variation der beiden von XGBoost bereitgestellten Parameter „`scale_pos_weight`“ und „`max_delta_step`“ (Readthedocs Webseite von XGBoost, 2020) brachte keine Verbesserung der Ergebnisse im Vergleich zur Standardrechnung. Daher wurden für die finale Modellberechnung die Standardwerte dieser Parameter verwendet. Um den Einfluss von Downsampling zu untersuchen, wurde ein neuer Trainingsdatensatz erstellt, der sich aus den Teilzeitbeschäftigten und 50% der Vollzeitbeschäftigten zusammensetzte. Das resultierende Modell wurde mit einem entsprechenden Modell, das auf dem Gesamttrainingsdatensatz trainiert wurde, verglichen.

Für die Modellbewertung wurden im Zuge der vorliegenden Testrechnungen die Metriken Balanced Accuracy, F1score und auch die Accuracy bestimmt. Um die Auswirkungen des Downsampling zu illustrieren und die berechneten Metriken zu vergleichen, sind im Folgenden die Werte für beide Modelle, mit und ohne Downsampling, dargestellt. [↘ Tabelle 1](#)

**Tabelle 1**

Bewertung der ML-Modelle mit und ohne Downsampling

Metrik	Standard	Mit Downsampling
Sensitivity	0,5203	0,6060
Precision	0,8219	0,7209
Accuracy	0,9650	0,9629
Balanced Accuracy	0,7566	0,7956
F1score	0,6373	0,6585

Gemessen an Balanced Accuracy und F1score führt Downsampling im Vergleich zum Standardvorgehen zu besseren Ergebnissen. Das ist darauf zurückzuführen, dass Downsampling auf eine Steigerung der Sensitivität des Modells abzielt, was heißt, dass möglichst viele Teilzeitbeschäftigte unter den als Vollzeitbeschäftigte deklarierten Beschäftigten identifiziert werden. Wenn der Anteil der Teilzeitbeschäftigten im Trainingsmaterial also höher ist, klassifiziert das Modell einen Beschäftigten auch eher als Teilzeitbeschäftigten. Anhand von Sensitivity und Precision ist zu erkennen, dass im Fall des Downsamplings zwar wesentlich mehr Teilzeitbeschäftigte als solche erkannt werden, aber zugleich auch viel mehr tatsächlich Vollzeitbeschäftigte als teilzeitbeschäftigt klassifiziert werden, also die Falsch-Positiv-Rate steigt.

Dieser Effekt ist bei der Bewertung zu berücksichtigen, da bei der vorliegenden Fragestellung nicht im Vordergrund stand, möglichst viele Teilzeitbeschäftigte zu identifizieren, sondern die Anzahl der fehlerhaften Tätigkeitsschlüssel zu minimieren. Anders formuliert bestand das Ziel darin, möglichst viele Beschäftigte, ob Teilzeit- oder Vollzeitbeschäftigte, richtig zu klassifizieren, was gleichbedeutend mit der Maximierung der Accuracy ist. Folglich wurde die Accuracy für die Bewertung der berechneten Modelle eingesetzt. Für die Modellrechnungen, bei denen die Auswirkung des Einsatzes von Downsampling getestet wurde, ergibt sich somit, dass die Klassifikationsgüte des mit Downsampling berechneten Modells minimal schlechter ist als die des Modells, das basierend auf dem Gesamttrainingsdatensatz berechnet wurde. Folglich wurde Downsampling bei der Berechnung des finalen Modells nicht eingesetzt.

## 5

### Ergebnisse

---

Das erstellte ML-Modell soll dazu eingesetzt werden, den fehlerhaften IEB-Datensatz zu korrigieren. Eine solche Korrektur ist dann erfolgreich, wenn der Fehler des Modells niedriger ist als der ursprüngliche Fehler im Datensatz oder – umgekehrt formuliert – wenn die Accuracy des Modells höher ist als die der Ausgangsdaten. In diesem Fall entspricht dies auch der „No Information Rate“. Die folgende Auswertung bezieht sich auf den

vom Modelltraining ausgeschlossenen Testdatensatz mit 97 324 Beschäftigten, der nach Berücksichtigung der Hochrechnungsfaktoren 3963994 Personen entspricht. Die Accuracy der Daten liegt vor der Korrektur mit dem ML-Modell bei 0,9402; das ist der Anteil der richtig als Vollzeitbeschäftigte geführten Beschäftigten. Mit dem berechneten XGBoost-Modell kann eine Accuracy von 0,965 erzielt werden, was einer Verbesserung um 0,0248 Prozentpunkte entspricht. In absoluten Zahlen unter Berücksichtigung der Hochrechnungsfaktoren entsteht diese Verbesserung dadurch, dass bei 234 254 fälschlich als Vollzeitbeschäftigte geführten Teilzeitbeschäftigten 121 246 richtig als Teilzeitbeschäftigte klassifiziert werden, aber auch 25 753 Vollzeitbeschäftigte falsch als Teilzeitbeschäftigte klassifiziert werden. Dadurch verringert sich insgesamt die Zahl der falsch signierten Beschäftigten um 95 493 (siehe auch Tabelle 2).

### Ausschluss kritischer Subgruppen

---

Als alternativer Ansatz wurde untersucht, ob es möglich ist, durch den Ausschluss von Teilgruppen, für die eine Vorhersage des Tätigkeitsschlüssels nur schwer möglich ist, das Ergebnis für die verbleibenden Datensätze zu verbessern. Hierzu erfolgten verschiedene Modellberechnungen, bei denen Beschäftigte beispielsweise abhängig von ihrem Bruttomonatsverdienst oder der Anzahl der Beschäftigten im Betrieb mit unterschiedlichen Schwellenwerten von der Berechnung ausgeschlossen wurden.

Um systematisch Gruppen zu identifizieren, für die die Klassifikationsgüte vergleichsweise schlecht ist, wurde letztlich ein Entscheidungsbaum eingesetzt, der auf dem Trainingsdatensatz mit dem Fehler des XGBoost-Modells als Zielvariable trainiert wurde. Die Idee war, mithilfe des Baums zu bestimmen, welche bekannten Merkmale innerhalb welcher Wertebereiche möglicherweise einen Hinweis darauf geben, in welchen Fällen der Vollzeit/Teilzeit-Schlüssel zuverlässig vorhergesagt werden kann und für welche Fälle dies nicht möglich ist, sodass letztere von der Berechnung ausgeschlossen werden können. [↪ Tabelle 2](#)

Es wurden verschiedene Varianten getestet, die sich durch die Anzahl der berücksichtigten Knoten des Entscheidungsbaums unterscheiden. Die getesteten Vari-

anten führen für die Teilgruppe im Vergleich zur Gesamtgruppe zu einer deutlichen Erhöhung der Sensitivität bei fast gleichbleibender Precision. Ein Beispiel, bei dem nur ein Blatt des Baums, das einer Auswahl von Beschäftigten mit einem maximalen Bruttomonatsverdienst von 1 899 Euro entspricht, berücksichtigt wurde, wird im Folgenden diskutiert.

**Tabelle 2**  
Bewertung der ML-Modelle

Gütemaß	Alle korrigiert	Nur Bruttomonatsverdienst unter 1 900 EUR korrigiert	
	Standard	Auswertung für Subgruppe	Auswertung für alle
Accuracy	0,9650	0,9127	0,9635
Accuracy vor der Korrektur	0,9402	0,8191	0,9402
Sensitivity	0,5176	0,6640	0,4915
Precision	0,8248	0,8192	0,8192
Echte Teilzeitbeschäftigte	234 254	173 395	234 254
Anzahl korrigierter Teilzeitbeschäftigter	95 493	89 724	89 724

Für die betrachtete Teilmenge, die sich durch den Ausschluss von Beschäftigten mit einem Bruttomonatsverdienst von mehr als 1 899 Euro ergibt, ist die Accuracy vor der Korrektur mit 0,8191 wesentlich geringer als für die Gesamtmenge (0,9402). Durch die Korrektur des ML-Modells wird die Accuracy für diese Teilmenge um etwa neun Prozentpunkte auf 0,9127 erhöht. Berechnet man analog zur Darstellung für die Gesamtmenge mithilfe der Hochrechnungsfaktoren die entsprechenden absoluten Zahlen, so würden bezogen auf die Testmenge die insgesamt 234 254 falsch klassifizierten Beschäftigten mit der Korrektur des ML-Verfahrens um 89 724 verringert. Für die Teilmenge ist das Potenzial zur Verbesserung also deutlich höher und es wird auch ein größerer Anteil von Beschäftigten korrigiert, die absolute Anzahl an korrigierten Datenpunkten bezogen auf den Gesamtdatensatz ist jedoch geringer.

Daher stellte sich die Frage, ob das auf der Teilmenge trainierte Modell besser an diese angepasst ist oder ob die beiden Modelle auf der Teilmenge ein vergleichbares Ergebnis liefern. Es zeigte sich, dass sich die Modelle selbst kaum unterscheiden. Die unterschiedliche Klassifikationsgüte, die auf der Teilmenge und der Gesamtmenge beobachtet wurde, lässt sich hauptsächlich durch die Unterschiede zwischen den Mengen, also die Merkmale der entsprechenden Beschäftigten, begründen.

Gemessen an der Accuracy bezogen auf den Gesamtdatensatz ist es möglich, ohne eine Subgruppenauswahl mit 0,965 im Vergleich zu 0,9635 (mit Ausschluss von Beschäftigten mit einem Bruttomonatsverdienst von mindestens 1 900 Euro) ein etwas besseres Ergebnis zu erzielen. In diesem Fall kann die Anzahl der falsch klassifizierten Teilzeitbeschäftigten insgesamt etwas weiter gesenkt werden. Es werden aber auch bewusst Korrekturen für eine Menge von Beschäftigten vorgenommen, für die das Modell nur sehr unzuverlässige Vorhersagen macht (Bruttomonatsverdienst von mindestens 1 900 Euro). Es kann nicht getestet werden, inwieweit sich die Güte der Vorhersagen für den VSE-Datensatz auf die Vorhersagen mit den Integrierten Erwerbsbiografien übertragen lässt, insbesondere da die Qualität anderer gemeinsamer Merkmale, auf denen die Vorhersage beruht, durch unterschiedliche Bearbeitung der beiden Datensätze leicht voneinander abweichen können. Daher besteht bei einer Subgruppe, für die die Vorhersage bei unserem Datenmaterial eher unzuverlässig ist, auch wenn sie den Fehleranteil minimal senkt, die Gefahr, dass bei Anwendung auf die Integrierten Erwerbsbiografien der Fehleranteil gleich bleibt oder sich sogar erhöht. Im letzteren Fall ist natürlich von einer Korrektur innerhalb dieser Subgruppe abzusehen, aber auch bei gleichbleibendem Fehleranteil ist die Originalkodierung vorzuziehen.

Zudem warf die Fehlsignierung der Teilzeitkräfte insbesondere Probleme bei Analysen des Niedriglohnbereichs auf. Im Jahr 2014 lag laut Verdienststrukturerhebung die Niedriglohnschwelle bei einem Bruttomonatsverdienst von 1 993 Euro. Somit liegt die Menge der Beschäftigten, für die nur eine eher unzuverlässige Vorhersage möglich ist, zu einem großen Teil außerhalb des für die Analyse der Daten besonders interessanten Bereichs. Damit könnten mit einer Einschränkung der zu korrigierenden Beschäftigten die genannten Probleme ohne starken Einfluss auf die Analyseergebnisse umgangen werden. Darüber hinaus ist anzumerken, dass zwar die von der Korrektur ausgeschlossene Menge 75 % der Beschäftigten ausmacht, aber nur 27 % der falsch klassifizierten Teilzeitbeschäftigten dazugehören. Für diese würde in der Folge des Ausschlusses keine Korrektur durch das Modell vorgenommen.

Zusammenfassend kann für die Anwendung auf die Integrierten Erwerbsbiografien festgehalten werden, dass das auf dem Gesamtdatensatz trainierte Modell eingesetzt werden sollte.




Sollen aber Korrekturen, die auf eher unzuverlässigen Vorhersagen beruhen, nicht vorgenommen werden, ist es trotzdem möglich, entsprechende Teilmengen des Datensatzes direkt bei der Korrektur auszuschließen.

## 6

---

### Fazit

---

Mit dem vorgestellten Projekt sollte ein Korrekturverfahren für das im Tätigkeitsschlüssel der Integrierten Erwerbsbiografien kodierte Merkmal „Vollzeit/Teilzeit“ entwickelt werden. Um eine hohe Datenqualität zu gewährleisten, findet für die Erstellung des VSE-Datensatzes des Statistischen Bundesamtes mit der Plausibilisierung ein aufwendiges, manuelles Korrekturverfahren statt. Im Zuge dessen wurde das auch in der Verdienststrukturerhebung vorhandene Merkmal „Vollzeit/Teilzeit“ für die Erhebung von 2014 überprüft und korrigiert. Zur Korrektur wurden auch Angaben zu den bezahlten Arbeitsstunden verwendet, die in der Verdienststrukturerhebung im Gegensatz zu den Integrierten Erwerbsbiografien vorhanden sind. Für das vorliegende Problem war dies nutzbar, indem ein maschinelles Lernverfahren auf den manuell korrigierten VSE-Daten zur Vorhersage der korrekten Angabe über die Vollzeit- oder Teilzeitbeschäftigung trainiert wurde. Unter Berücksichtigung verschiedener projektspezifischer Anforderungen wurde so ein ML-Modell erstellt, das in der Lage ist, den Fehler in der Signierung des Tätigkeitsschlüssels bei einer aus dem VSE-Datensatz erzeugten Testmenge zu reduzieren, bei 234 254 falsch kodierten Einträgen um 95 493. Mit dem entwickelten Verfahren besteht somit auf die Integrierten Erwerbsbiografien angewendet die Möglichkeit, die dort festgestellte Untererfassung der Teilzeitbeschäftigten teilweise zu korrigieren. 

## LITERATURVERZEICHNIS

---

- Bertat, Thomas/Dundler, Agnes/Grimm, Christopher/Kiewitt, Jochen/Schomaker, Christine/Schridde, Henning/Zemann, Christian. *Neue Erhebungsinhalte "Arbeitszeit", "ausgeübte Tätigkeit" sowie "Schul- und Berufsabschluss" in der Beschäftigungsstatistik*. Methodenbericht der Statistik der BA. 2013. [Zugriff am 12. November 2020]. Verfügbar unter: [statistik.arbeitsagentur.de](http://statistik.arbeitsagentur.de)
- Breiman, Leo. *Arcing classifier (with discussion and a rejoinder by the author)*. In: *Annals of Statistics*. Jahrgang 26. Ausgabe 3/1998, Seite 801 ff.
- Breiman, Leo/Friedman, Jerome H./Olshen, Richard A./Stone, Charles J. *CART: Classification and Regression Trees*. Boca Raton 1984.
- Breiman, Leo. *Random Forests*. In: *Machine Learning*. Ausgabe 45/2001, Seite 5 ff.
- Chen, Tianqi/Guestrin, Carlos. *XGBoost: A Scalable Tree Boosting System*. In: arXiv, 2016, arXiv:1603.02754
- Fitzenberger, Bernd/Seidlitz, Arnim. *The 2011 Break in the Part-Time Indicator and the Evolution of Wage Inequality in Germany*. ZEW Discussion Paper No. 19-029, Mannheim 2019. [Zugriff am 12. November 2020]. Verfügbar unter: [ftp.zew.de](http://ftp.zew.de)
- Fitzenberger, Bernd/Seidlitz, Arnim. *Die Lohnungleichheit von Vollzeitbeschäftigten in Deutschland: Rückblick und Überblick*. In: *AStA Wirtschafts- und Sozialstatistisches Archiv*. Jahrgang 14. Ausgabe 2/2020, Seite 125 ff. [Zugriff am 12. November 2020]. Verfügbar unter: [link.springer.com](http://link.springer.com)
- Freund, Yoav/Schapire, Robert E. *Experiments with a new boosting algorithm*. In: *Machine Learning: Proceedings of the Thirteenth International Conference*. 1996. [Zugriff am 12. November 2020]. Verfügbar unter: [cseweb.ucsd.edu](http://cseweb.ucsd.edu)
- HackerEarth. *Beginners Tutorial on XGBoost and Parameter Tuning in R*. [Zugriff am 12. November 2020]. Verfügbar unter: [www.hackerearth.com](http://www.hackerearth.com)
- Jain, Aarshay. *Complete Guide to Parameter Tuning in XGBoost with codes in Python*. 2016. [Zugriff am 12. November 2020]. Verfügbar unter: [www.analyticsvidhya.com](http://www.analyticsvidhya.com)
- Möller, Joachim. *Lohnungleichheit – Gibt es eine Trendwende?* In: *Wirtschaftsdienst*. Jahrgang 96. Ausgabe 1/2016, Seite 38 ff. [Zugriff am 12. November 2020]. Verfügbar unter: [www.wirtschaftsdienst.eu](http://www.wirtschaftsdienst.eu)
- Readthedocs Webseite von XGBoost. *XGBoost Parameters*. [Zugriff am 12. November 2020]. Verfügbar unter: [xgboost.readthedocs.io](http://xgboost.readthedocs.io)
- Schapire, Robert E./Freund, Yoav. *Boosting: Foundations and Algorithms*. 2012.
- Schapire, Robert E. *The Strength of Weak Learnability*. In: *Machine Learning*. Jahrgang 5. Ausgabe 2/1990, Seite 197 ff. [Zugriff am 12. November 2020]. Verfügbar unter: [rob.schapire.net](http://rob.schapire.net)
- Statistisches Bundesamt. *Qualitätsbericht Verdienststrukturerhebung 2014*. Wiesbaden 2016. Verfügbar unter: [www.destatis.de](http://www.destatis.de)



**Dr. Hanna Brenzel**

hat internationale Volkswirtschaft studiert und 2018 ihre Promotion an der Fakultät Sozial- und Wirtschaftswissenschaften in Bamberg abgeschlossen. Sie leitet das Referat „Methoden der Datenanalyse“ im Statistischen Bundesamt, welches aktuell schwerpunktmäßig die Themenfelder Georeferenzierung und Mikrosimulation umfasst.



**Kathrin Gebers**

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

## WERKSTATTBERICHT: GEOREFERENZIERUNG IM STATISTISCHEN VERBUND

Dr. Hanna Brenzel, Kathrin Gebers

↘ **Schlüsselwörter:** Datenintegration – Geoinformationen – Georeferenzierung – Statistischer Verbund – Infrastruktur

### ZUSAMMENFASSUNG

Die Integration statistischer und geografischer Informationen gewinnt zunehmend an Bedeutung. Der Statistische Verbund arbeitet aktiv an der Schaffung einer nachhaltigen Infrastruktur zur dauerhaften Implementierung der Integration von Geo- und Statistikdaten. Damit soll Politik, Wissenschaft und Gesellschaft eine neue, erweiterte Grundlage für evidenzbasierte Entscheidungen mit Regionalbezug, die zur Willensbildung und in Entscheidungsprozessen in einer demokratischen Gesellschaft notwendig sind, zur Verfügung stehen. Der Artikel gibt einen Überblick über die Rahmenbedingungen, die angestoßenen Aktivitäten und die Analyseprojekte auf der Basis integrierter statistischer und geografischer Informationen.

↘ **Keywords:** data integration – geospatial information – geo-referencing – German official statistics network – infrastructure

### ABSTRACT

*The integration of statistical and geographical information is becoming increasingly important. The German official statistics network is actively working on the creation of a sustainable infrastructure for the permanent implementation of the integration of geodata and statistical data. This will provide politics, science and the society with a new, extended basis for evidence-based decision-making with regional relevance. This improved basis of information is necessary for the development of informed opinions and for decision-making processes in a democratic society. The article gives an overview of the framework conditions, the activities initiated and the current analysis projects based on integrated statistical and geographical information.*

### 1

---

## Einleitung

---

Die Integration beziehungsweise Kombination von statistischen und geografischen Informationen gewinnt zunehmend an Bedeutung. Gesellschaften, Volkswirtschaften und unsere Umwelt haben sich schon immer verändert, aber die heutige globale Gemeinschaft muss sich mit Veränderungen von beispielloser Komplexität und Geschwindigkeit und mit potenziell verheerenden Auswirkungen auseinandersetzen, sollte auf die Herausforderungen nicht in der richtigen Weise reagiert werden. Regierungen und Verwaltungen können diese Aufgabe nur mit Informationen bewältigen, die schnell und mit einer hohen Granularität zur Verfügung stehen (UN-GGIM:Europe, 2019).

Die Integration von statistischen und geografischen Daten stellt dabei einen vielversprechenden Weg dar, um zeitnah zuverlässige, relevante und detaillierte Informationen bereitzustellen. Sie ermöglicht es, den gesamten politischen Lebenszyklus sowie Prognosen zu unterstützen und die verschiedenen Dimensionen der nachhaltigen Entwicklung (Umwelt, Wirtschaft, Gesellschaft) in ein kohärentes Bild zu bringen. Informationen können mit der jeweils benötigten räumlichen Auflösung von lokal bis global geliefert werden und ermöglichen so politische Interventionen auf der Ebene, auf der sie am wirksamsten sind. Durch die Integration von Geo- und Statistikdaten entsteht damit eine neue, erweiterte Grundlage für evidenzbasierte Entscheidungen mit Regionalbezug, die zur Willensbildung und in Entscheidungsprozessen in einer demokratischen Gesellschaft notwendig sind.

Die Erweiterung des Regionalbezugs besteht in der räumlichen Tiefe (Auflösung) und Flexibilität (Diversität) der analysierten Regionaleinheiten. Eine effiziente Datenintegration ermöglicht es, diese aktueller, zuverlässiger, relevanter und detaillierter als bisher bereitzustellen. Die Vorteile der Datenintegration werden von den einzelnen Interessengruppen (Wirtschaft, Gesellschaft, Wissenschaft) zunehmend erkannt; auch die Anbieter von Geodaten und statistischen Daten sind bestrebt, Regierungen und internationalen Institutionen sowie sonstigen Interessengruppen bestmögliche Informationen bereitzustellen.

Auf internationaler Ebene hat die Expertengruppe „Integration of Statistical and Geospatial Information“ der Vereinten Nationen das [Global Statistical Geospatial Framework](#) erarbeitet. Dieses umfassende Rahmenwerk beschreibt Wege zur Datenintegration sowie Mechanismen, um weltweit eine entsprechende Wirksamkeit der gewonnenen Informationen sicherzustellen.

Der folgende Beitrag gibt einen Überblick über bereits realisierte sowie noch ausstehende Anstrengungen und Bestrebungen der Statistischen Ämter des Bundes und der Länder (Statistischer Verbund) im Bereich der Georeferenzierung. Kapitel 2 skizziert die rechtliche nationale und europäische Ausgangslage sowie die organisatorische Struktur, in Kapitel 3 wird die strategische und technische Realisierung in Deutschland dargelegt. Kapitel 4 stellt bereits im Statistischen Verbund realisierte Projekte vor und erläutert das Potenzial, welches sich aus der Integration statistischer und geografischer Informationen ergibt. Ein Ausblick rundet den Bericht ab.

### 2

---

## Ausgangslage

---

### 2.1 Rechtliches

---

Um eine grenzübergreifende Nutzung von Geoinformationen in Europa zu erleichtern, haben auf europäischer Ebene bereits im Jahr 2007 das Europäische Parlament und der Europäische Rat die INSPIRE-Richtlinie 2007/2/EG verabschiedet. Diese Infrastructure for SPatial Information in Europe (INSPIRE) ist das Vorhaben für eine gemeinsame Geodateninfrastruktur in Europa. Sie fordert eine einheitliche Beschreibung der Geodaten und deren Bereitstellung im Internet mit Diensten für Suche, Visualisierung und Download. Fachliche und technische Einzelheiten regelt die Europäische Union (EU) mit Durchführungsbestimmungen, die für die Mitgliedstaaten direkt verbindlich sind. Am 15. Mai 2007 trat die INSPIRE-Richtlinie in Kraft und wurde mittlerweile von den Mitgliedstaaten in nationales Recht umgesetzt.<sup>1</sup>

---

1 Weiterführende Informationen hierzu bieten die folgenden Webseiten: [www.gdi-de.org](http://www.gdi-de.org) oder [inspire.ec.europa.eu](http://inspire.ec.europa.eu)

Zwar wird die Erstellung neuer Daten von INSPIRE nicht gefordert, dennoch hat diese europäische Initiative erheblich dazu beigetragen, dass in vielen Bereichen Maßnahmen eingeleitet worden sind, um künftig Informationen kleinräumiger als bisher zur Verfügung stellen zu können. Dies gilt auch für die amtliche Statistik. Gemäß Bundesstatistikgesetz war jahrzehntelang die Nutzung der Gemeinde und der Blockseite<sup>2</sup> als kleinste Einheit für die regionale Zuordnung der Erhebungsmerkmale zulässig. Erst mit einer Novellierung des Bundesstatistikgesetzes im August 2013 wurde die rechtliche Grundlage zur Geokodierung in der amtlichen Statistik in Deutschland geschaffen. Zwar ist nach § 10 Bundesstatistikgesetz die dauerhafte Speicherung des Raumbezugs statistischer Daten weiterhin nicht adressscharf erlaubt, jedoch mit Bezug auf ein geografisches Gitter mit einer Mindestgitterweite von 100 Metern. Eine Ausnahme stellen Register<sup>3</sup> dar: Seit der Verabschiedung des E-Government-Gesetzes im Jahr 2013 ist durch § 14 Georeferenzierung die Speicherung des Geokoordinatenpaars explizit vorgeschrieben. Diese Gesetzesänderungen haben den Startpunkt der Geokodierung im Statistischen Verbund gesetzt. Bei allen Statistiken, für die keine darüber hinausgehende Regelung besteht, sollten daher vor der rechtlich vorgeschriebenen Löschung der Adress- und Koordinatenangaben die Angaben zu den Gitterzellen, in welche die Koordinaten fallen, ergänzt werden. Dabei werden die Gitterzellen durch einen Identifikator, die Gitterzellen-ID, kodiert. Die im Statistischen Verbund genutzte Syntax der Gitterbezeichnung und der Gitterzellen-ID orientiert sich an den INSPIRE-Vorgaben, genauer an den Durchführungsbestimmungen zum INSPIRE-Thema „Statistische Einheiten“ (Details dazu enthält Kapitel 3).

2 Eine Blockseite ist nach § 10 Absatz 3 Bundesstatistikgesetz definiert als „innerhalb eines Gemeindegebiets die Seite mit gleicher Straßenbezeichnung von der durch Straßeneinmündungen oder vergleichbare Begrenzungen umschlossenen Fläche“. Blockseiten liegen nicht flächendeckend vor und werden im Wesentlichen von kommunalen Statistikstellen genutzt.

3 § 14 E-Government-Gesetz: Georeferenzierung  
(1) Wird ein elektronisches Register, welches Angaben mit Bezug zu inländischen Grundstücken enthält, neu aufgebaut oder überarbeitet, hat die Behörde in das Register eine bundesweit einheitlich festgelegte direkte Georeferenzierung (Koordinate) zu dem jeweiligen Flurstück, dem Gebäude oder zu einem in einer Rechtsvorschrift definierten Gebiet aufzunehmen, auf welches sich die Angaben beziehen.  
(2) Register im Sinne dieses Gesetzes sind solche, für die Daten auf Grund von Rechtsvorschriften des Bundes erhoben oder gespeichert werden; dies können öffentliche und nichtöffentliche Register sein.

#### ↳ Geokodierung gegenüber Georeferenzierung

Häufig werden die Begriffe Georeferenzierung und Geokodierung als Synonyme verwendet. Grundsätzlich unterscheiden sich allerdings Geokodierung – die Zuweisung eines Geokoordinatenpaares zu einer Anschrift – und Georeferenzierung – das in Bezug Setzen eines Datensatzes in einen räumlichen Kontext – voneinander. Dies entspricht der aktuellen Überarbeitung des Generic Statistical Business Process Model GSBPM (Version 5.1) durch die Wirtschaftskommission für Europa der Vereinten Nationen (UNECE). In der Version 5.1 wurden die Geokodierung und die Nutzung von georeferenzierten Daten in das Geschäftsprozessmodell eingearbeitet.

Aufgrund des Föderalismus in Deutschland und der entsprechenden Kooperation in der amtlichen Statistik gibt es für die Umsetzung der Geokodierung der statistischen Einheiten zwei notwendige Voraussetzungen: Neben der Novellierung des § 10 Bundesstatistikgesetz ist dies die Regelung der Nutzung der Geobasisdaten der Vermessungsverwaltungen der Länder durch Bundes- und Landeseinrichtungen. Zu den Geobasisdaten zählen unter anderem die Geometrien der Verwaltungsgebiete, Luftbilder, Landschaftsmodelle und georeferenzierte Adressdaten. In den letzten Jahren haben sich die Rahmenbedingungen für die Nutzung dieser Datensätze positiv entwickelt. Insbesondere die automatisierte Übermittlung der Daten für digitale Prozessketten steht dabei im Vordergrund. Manuelle oder analoge Lizenzierungsregelungen stehen dem noch im Wege. Gleiches gilt für restriktive Nutzungsregelungen, die eine Weitergabe an Dritte einschränken. Der Vertrag über die kontinuierliche Übermittlung amtlicher digitaler Geobasisdaten der Länder zur Nutzung im Bundesbereich (V GeoBund) ist die rechtliche Grundlage, über die das Bundesamt für Kartographie und Geodäsie (BKG) die Geobasisdaten der Länder Bundeseinrichtungen zur Verfügung stellt. Um den genannten Anforderungen gerecht zu werden, hat das BKG einen Neuabschluss des bestehenden Vertrags mit den Ländern forciert und umgesetzt. Der neue V GeoBund ist zum 1. Juli 2019 in Kraft getreten. Seit dem 1. September 2019 regelt der Vertrag über die gegenseitige Nutzung amtlicher digitaler Geobasisdaten der Länder (V GeoLänder) die Nutzung zwischen den Landesämtern bei Vorhaben ohne oder mit Beteiligung des Bundes.

Der V GeoLänder beschreibt unter anderem, dass und zu welchen Bedingungen ein Bundesland mit den Geobasisdaten eines anderen Bundeslands arbeiten darf. Dies ist für die Nutzung bundesländerübergreifender Datenbestände durch ein statistisches Amt nach dem Patenlandprinzip notwendig. Ein Beispiel hierzu stellt der Unfallatlas dar (siehe Kapitel 4).

### 2.2 Organisatorisches

---

Nachdem das Bundesstatistikgesetz die gesetzlichen Voraussetzungen zur Geokodierung statistischer Einheiten geschaffen hatte, übernahm die Steuerungsgruppe Analyse und Darstellung georeferenzierter Daten (SG Geo) als Bund-Länder-Gremium die übergreifende Koordination dieses neuen Handlungsfelds der amtlichen Statistik. Ziel war ein gemeinschaftliches flächendeckendes kleinräumiges Datenangebot der Bundesstatistik. Zum konkreten Aufgabenprofil zählte unter anderem:

- › ein gemeinsames Grundverständnis zur Nutzung und zu Analysepotenzialen integrierter statistischer und geografischer Daten zu erarbeiten,
- › weitere rechtliche Voraussetzungen bei einzelnen Fachstatistiken zu schaffen,
- › Methodenkompetenzen auszubauen, sowie
- › das Thema Geoinformationen und kleinräumige Analysen in den statistischen Ämtern weiter zu etablieren.

Die SG Geo wurde mit Beschluss der Amtsleiterkonferenz<sup>4</sup> vom Juni 2019 aufgelöst, nachdem sie ihre initialen Steuerungsaufgaben erfüllt hatte. Ihr Nachfolgegremium ist der Arbeitskreis Geo. Er ist für die Koordination der Verankerung der Arbeit mit geokodierten Statistiken im Statistikerstellungsprozess, für die Weiterentwicklung der Teilprozesse und für die Klärung methodischer Fragen im Statistischen Verbund zuständig.

---

4 Die Amtsleiterkonferenz ist das höchste Beratungs- und Entscheidungsgremium der Leiterinnen und Leiter der Statistischen Ämter des Bundes und der Länder.

## 3

---

### Umsetzung der Geokodierung

---

Ein Beschluss der Amtsleiterkonferenz vom November 2017 bedeutete den Startschuss für die praktische Umsetzung der Geokodierung in der deutschen amtlichen Statistik. Ambitioniertes Ziel war die Geokodierung aller geokodierbaren Statistiken innerhalb von zwei Jahren. Bis Ende 2018 sollten bereits 80% der Statistiken geokodiert im Statistischen Verbund vorliegen. Um dieses Ziel zu erreichen, war es notwendig, eine einheitliche Definition von geokodierbaren Statistiken zu entwickeln. Statistiken werden demnach als geokodierbar gekennzeichnet, wenn für die auskunftgebende/meldende Stelle oder die statistische Einheit ein regionaler Bezug unterhalb der Gemeindeebene vorhanden ist (Klumpe und andere, 2020). Derzeit gilt eine Statistik als geokodiert, wenn die Geokoordinaten oder Gitterzellen-IDs der statistischen Einheiten im Statistischen Verbund zum jeweils letzten verfügbaren Berichtszeitpunkt (möglicherweise auch als redundanter Datensatz) vorliegen. Neben Registern und Verzeichnissen zählen also Statistiken, bei denen entweder die statistische Einheit selbst oder die auskunftgebende Stelle geokodiert werden kann, zu den geokodierbaren Statistiken. Hierbei ist wichtig, dass der vorliegende Raumbezug der auskunftgebenden Stelle nicht zwischen den Bundesländern variiert, sondern deutschlandweit homogen<sup>5</sup> ist. In einem ersten Schritt wurden 197 geokodierbare Statistiken identifiziert. Darin enthalten sind 15 Statistiken, die aufgrund rechtlicher, technischer oder organisatorischer Hindernisse erst nach 2019 geokodiert werden können. Zum Stand 31. Dezember 2019 lagen 176 der 197 geokodierbaren Statistiken geokodiert im Statistischen Verbund vor. Bei sechs Statistiken sind die Voraussetzungen geschaffen worden; die Statistiken werden geokodiert, sobald dies möglich ist, beispielsweise sobald die Anschriften vorliegen.

Um die fachlich zuständigen Abteilungen bei der Umsetzung der Aufgabe technisch zu unterstützen und um die Prozesse zu standardisieren, wurden Standardprozesse

---

5 Beispielsweise sind in den Sozialstatistiken die Träger der Sozialleistungen auskunftspflichtig. Diese melden jedoch nicht immer einzeln an den Statistischen Verbund, sondern teilweise sammelt eine zentrale Meldestelle von allen auskunftspflichtigen Trägern die Meldungen und erteilt für alle Auskunftseinheiten zusammen Auskunft.

für die Geokodierung von Anschriften sowie Anleitungen zum Vorgehen erarbeitet. Sie stehen im Statistischen Verbund zur Verfügung. Zur Zuspiegelung der Koordinaten an die Anschriften wird standardmäßig der [Geokodierungsdienst des BKG](#) genutzt. Anschriften zählen zu den Einzelangaben im Sinne des § 16 Absatz 1 Bundesstatistikgesetz, daher fallen sie unter die statistische Geheimhaltung. Dementsprechend müssen die statistischen Ämter als anonyme Nutzer auf den Geokodierungsdienst zugreifen. Dies wird für die Statistischen Ämter des Bundes und der Länder über eine anonyme Vergabe der Zugangskennungen gewährleistet.

Zur dauerhaften Speicherung des Raumbezugs über Gitterzellen sowie bei Veröffentlichungen mittels Rasterkarten wurde das metrische Koordinatenreferenzsystem ETRS89/LAEA festgelegt. Es handelt sich dabei um die flächentreue azimutale Projektion nach Lambert zum Datum ETRS89 beziehungsweise um das Referenzsystem mit dem EPSG-Code<sup>6</sup> 3035. Als Gitterbezeichnung ergibt sich durch die genannten Vorgaben für das (100 m x 100 m)-Gitter die Zeichenkette CRS3035RES100m<sup>7</sup>. Zur Bildung der Gitterzellen-ID sind im Wesentlichen die Koordinatenangaben der südwestlichen Ecken der Gitterzellen anzufügen: CRS3035RES100mN<y-Koordinate der SW-Ecke>E<x-Koordinate der SW-Ecke><sup>8</sup>.

Ebenso wurden Qualitätskennzeichen für Geokoordinaten und Gitterzellen hinsichtlich der räumlichen Genauigkeit zugeordneter statistischer Einzelinformationen erarbeitet, die aus den Qualitätsparametern der Benutzeroberflächen des Geokodierungsdienstes abgeleitet werden. Diese Qualitätskennzeichen bündeln die Einzelinformationen der Qualitätsparameter des Geokodierungsdienstes und ermöglichen es Personen, die nicht

in dem Thema Geokodierung geschult sind, die Qualität der Geokodierung zu beurteilen. Die Qualitätskennzeichen werden in Registern beziehungsweise fachstatistischen Datenbeständen mitgeführt.

In der Aktualisierung des Geschäftsprozessmodells Amtliche Statistik (GMAS) hat die Integration statistischer und geografischer Daten ebenso Eingang gefunden. Dieses generische Modell beschreibt den Statistikerstellungsprozess in acht einzelnen Phasen mit jeweiligen Teilprozessen. Diese werden nicht stets in identischer Reihenfolge durchlaufen und nicht jeder Prozess enthält alle Teilprozesse. Sollte aber ein Teilprozess im Rahmen der Statistikproduktion durchlaufen werden, so ist er idealerweise wie im GMAS beschrieben zu organisieren (Gehle/Lüüs, 2017).

Der Prozess der Geokodierung wird durch die Einbindung des vom BKG zur Verfügung gestellten Geokodierungsdienstes sukzessive in die einzelnen fachstatistischen Produktionsprozesse implementiert.

---

## 4

---

### Realisierte Projekte

---

Die Umsetzung der Geokodierung im Statistischen Verbund hat in den vergangenen Jahren eine ausgezeichnete Datengrundlage geschaffen. Wie im Grundverständnis des Statistischen Verbunds formuliert, ermöglicht es diese Datengrundlage, statistische Informationen adäquat, zeitgemäß und entsprechend den internationalen Standards bereitstellen und verbreiten zu können. Der Aufbau eines entsprechenden neuen Datenangebots erfordert jedoch auch, neue Analyseideen zu entwickeln und zu etablieren. Dies sollte im Kontext der Verschneidung von statistikübergreifend thematisch vielschichtigen räumlichen Informationen erfolgen. Zu den prominentesten Produkten, die bereits frühzeitig durch die Integration statistischer und geografischer Daten veröffentlicht wurden, zählen der [Agraratlas](#), der [Zensusatlas](#), der [Krankenhausatlas](#) sowie der [Unfallatlas](#). Mit dem Agraratlas und dem Zensusatlas werden zwei interaktive Kartenanwendungen auf der Grundlage von Rasterkarten angeboten. Der Agraratlas umfasst insgesamt 16 Karten zum Thema Landwirtschaft. Die Datengrundlage dazu bilden georeferenzierte Daten aus der

---

6 Zum EPSG-Code siehe zum Beispiel [https://de.wikipedia.org/wiki/European\\_Petroleum\\_Survey\\_Group\\_Geodesy](https://de.wikipedia.org/wiki/European_Petroleum_Survey_Group_Geodesy). Da neben dem geodätischen Datum ETRS89 und der Projektion LAEA weitere Parameter zur genauen Festlegung des Referenzsystems notwendig sind, ist der EPSG-Code im Vergleich zu „ETRS89/LAEA“ die exaktere Angabe (siehe dazu <https://epsg.io/3035>). Allgemein wird unter der Bezeichnung ETRS89/LAEA bereits das Koordinatenreferenzsystem zum EPSG-Code 3035 verstanden. Dies wird auch in diesem Dokument so gehandhabt.

7 Die Kürzel „CRS“ beziehungsweise „RES“ stehen für „Coordinate Reference System“ beziehungsweise „Resolution“.

8 In blauer Schrift sind die Platzhalter für die jeweiligen Koordinaten hervorgehoben. Die Koordinaten der Gitterecken sind aufgrund der INSPIRE-Vorgaben stets ganzzahlig und werden hier als ganze Zahlen ohne Nachkommastellen, also als reine Ziffernfolge, angegeben. Man beachte, dass die Angabe der y-Koordinate vor der Angabe der x-Koordinate erfolgt. Die Buchstaben „N“ und „E“ stehen entsprechend für „North“ und „East“.

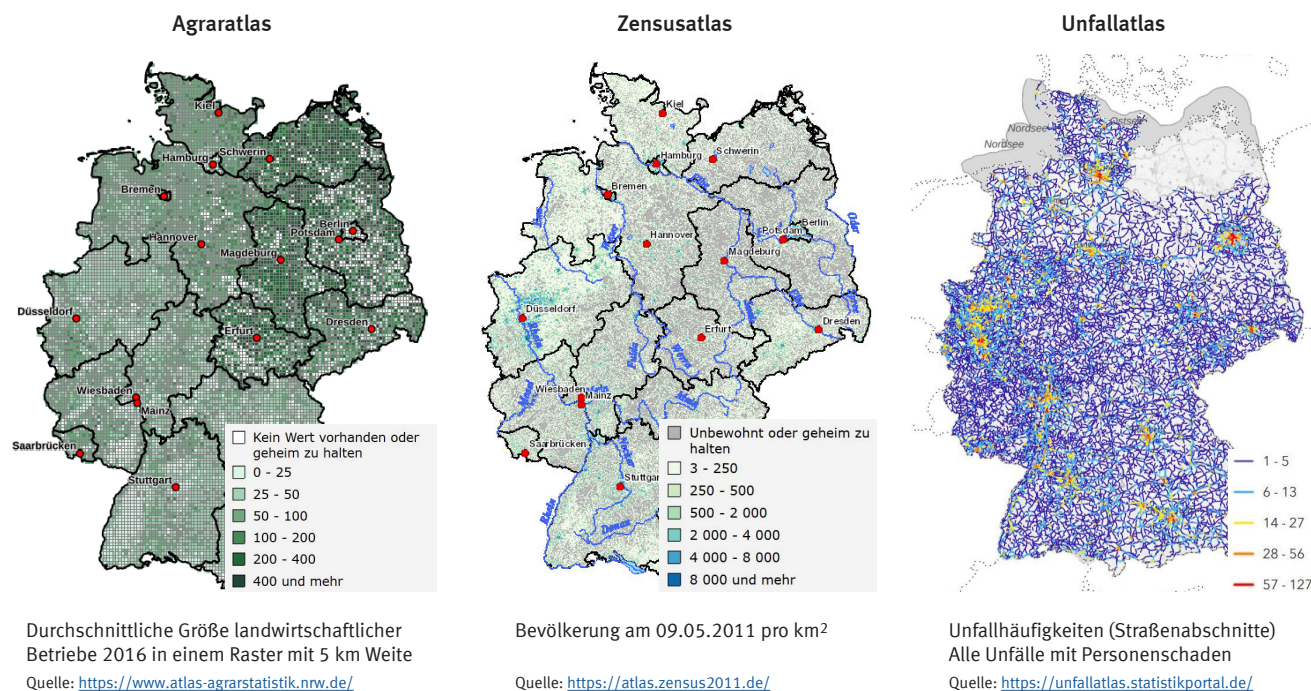
Landwirtschaftszählung 2010<sup>9</sup>. Der Zensusatlas bietet Daten zur Bevölkerung und zum Thema Wohnen aus dem Zensus 2011 auf der räumlichen Basis eines 1-Kilometer-Gitters. Zudem ist es möglich, die Einwohnerzahl für beliebige Flächen zu berechnen („Einwohnerrechner“). Beispiele dafür sind ein Kreis mit beliebigem Radius um ein Einkaufszentrum, ein Gebiet für einen Lieferservice oder eine Lärmzone um einen Flughafen. Der Unfallatlas zeigt georeferenzierte Unfallorte der Straßenverkehrsunfallstatistik und ermöglicht die Betrachtung der Statistik unter anderem anhand hochauflösender Karten. Somit können Fragen beantwortet werden wie „Wo passieren in meiner Stadt die meisten Unfälle?“, „Wo bin ich auf meinem Weg zur Arbeit besonders gefährdet?“, „Welche Kreuzungen oder Autobahnabschnitte sind besonders unfallträchtig?“, „Wo gab es Unfälle mit Verkehrstoten?“ oder „Waren Pkw, Motorräder, Fahrräder oder Fußgängerinnen und Fußgänger am Unfall beteiligt?“ ➔ Grafik 1.

Neben den Beispielen klassischer kartografischer Darstellungen lassen sich darüber hinaus durch die Integration statistischer und geografischer Daten weitere Produkte generieren, wie die Ergebnisse sogenannter Erreichbarkeitsanalysen. Ein aktuelles Beispiel aus der amtlichen Statistik dafür ist der Krankenhausatlas. Neben den Standorten von Krankenhäusern mit ausgewählten Fachabteilungen zeigt dieser in Isochronen-Karten die Fahrzeit von jedem Punkt Deutschlands zum nächstgelegenen Krankenhaus. Dabei ist es möglich, zwischen sechs verschiedenen Leistungsangeboten von Krankenhäusern zu unterscheiden. So lassen sich zum Beispiel deutlich Regionen erkennen, in denen mit besonders langen Fahrzeiten zu Geburtskliniken oder zu Krankenhäusern mit psychiatrischen Fachabteilungen zu rechnen ist. Beim Krankenhausatlas können die Isochronen-Karten mit den hochauflösenden Bevölkerungsdaten des Zensus 2011 sowie den Stadt-Land-Gliederungen des Bundesinstituts für Bau-, Stadt- und Raumforschung kombiniert werden. Dies ermöglicht zusätzliche Auswertungen in Form von Diagrammen, die die Fahrzeiten zu Krankenhäusern für unterschiedliche Bevölkerungsgruppen zeigen. Dadurch lässt sich zum

9 Aufgrund einer gesetzlichen Vorgabe auf europäischer Ebene war die Umsetzung des Agraratlas, welche die Geokodierung der landwirtschaftlichen Betriebsätze erforderte, nicht von der Anpassung des Bundesstatistikgesetzes abhängig (siehe Abschnitt 2.1).

## Grafik 1

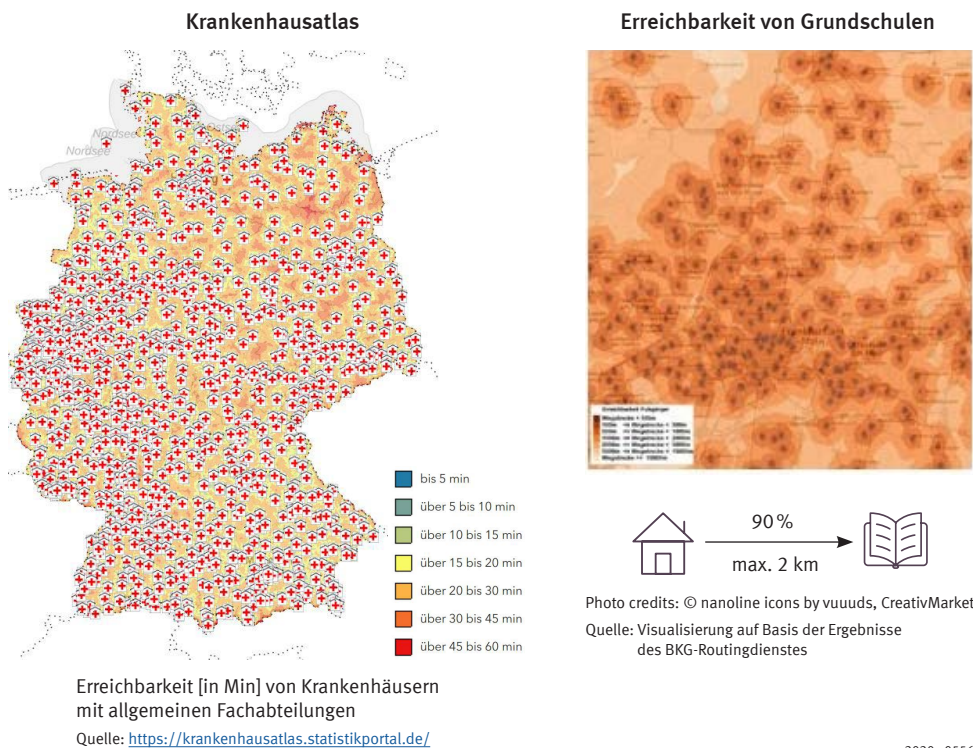
### Beispiele für kartografische Darstellungen





**Grafik 2**

Beispiele für Erreichbarkeitsanalysen



Beispiel erkennen, dass 90% der städtischen Bevölkerung in maximal 15 Minuten das nächstgelegene Krankenhaus mit allgemeinen Fachabteilungen erreichen, die durchschnittliche Fahrzeit desselben Bevölkerungsanteils in ländlichen Regionen aber rund zehn Minuten länger ist. Ein weiteres Beispiel für Erreichbarkeitsanalysen stellt das Projekt „Erreichbarkeiten von Grundschulen“ dar. Ein Team aus Vertreterinnen und Vertretern des Statistischen Bundesamtes, des Hessischen Statistischen Landesamtes sowie des Bundesamts für Kartographie und Geodäsie hat dieses Projekt im Rahmen eines Proof of Concepts (PoC) realisiert. Dieses Beispiel zeigt zudem, dass allein durch das Anspielen von Informationen über den Raumbezug in Verbindung mit dessen Nutzung in der anschließenden Auswertung ein relevanter Mehrwert generiert werden kann. In der konkreten Fragestellung wurde untersucht, wie durch die Kenntnis der Adressen von Grundschulen die Stichprobenerhebung des Mikrozensus mit zusätzlichen Informationen – das war im konkreten Projekt die Distanz zur nächsten Grundschule – angereichert werden kann. Hierzu wurden die Grundschulen und die Haushalte des Mikrozensus geokodiert. Mithilfe des BKG-Routingdienstes

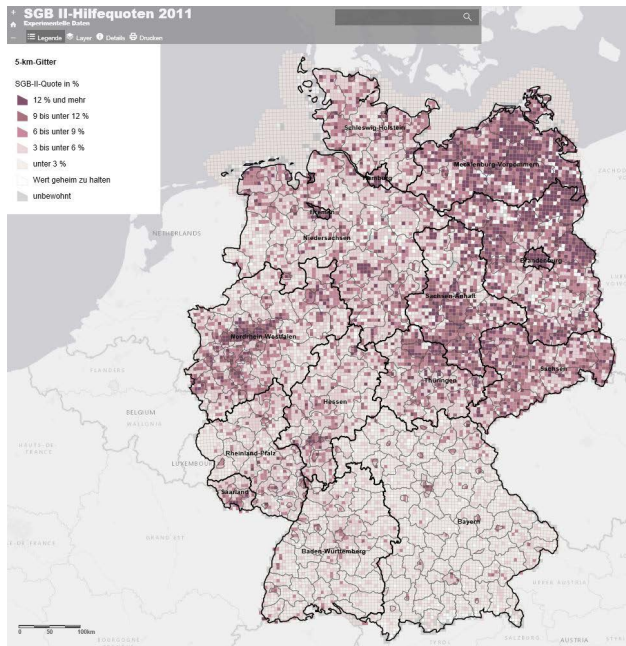
wurden die Entfernungen der Mikrozensus-Haushalte zur nächstgelegenen Grundschule für Familien in städtischen und ländlichen Region ermittelt (Gebers/Gräze, 2019). Die ermittelten Werte können anschließend wie die „normalen“ Mikrozensusmerkmale allein oder kombiniert mit diesen hochgerechnet werden. ➤ Grafik 2

Ein weiteres aktuelles Beispiel stellt der [SGB-II-Atlas](#) dar, den das Statistische Bundesamt und der Bereich Statistik der Bundesagentur für Arbeit gemeinsam entwickelt haben. Die SGB-II-Hilfequoten werden dabei kartografisch in Form von Rasterkarten visualisiert. ➤ Grafik 3 Dabei wird für jede Rasterzelle die berechnete Quote ausgewiesen, sofern es die statistische Geheimhaltung und Anforderungen an die Datenqualität erlauben. Bei diesem Projekt sind noch qualitative Verbesserungen notwendig.<sup>10</sup> Daher ist dieses Projekt derzeit noch unter der [Rubrik experimentelle Daten](#) angesiedelt. Dennoch zeigen die georeferenzierten SGB-II-Hilfequoten für Deutschland das hohe Potenzial

10 Zum Beispiel bei der Adressqualität der Leistungsempfänger oder der noch ausstehenden Bevölkerungsfortschreibung auf Gitterzellen-ebene.

## Grafik 3

### Georeferenzierte SGB-II-Hilfequoten



Quelle: <https://www.destatis.de/DE/Service/EXDAT/Datensaetze/sgb-ii-hilfequoten.html>

2020 - 0557

georeferenzierter Daten. Mit der kleinräumigen Darstellung kann eine Datenbasis geschaffen werden, die Aussagen über teilräumliche Gebietsgliederungen bis auf Gitterzellenebene ermöglicht. Dies ist beispielsweise für die örtliche Sozialplanung unter dem Aspekt der Entwicklung gleichwertiger Lebensverhältnisse notwendig.

## 5

### Ausblick

Die in Kapitel 4 beschriebenen Beispiele belegen, dass ein Wandel in der amtlichen Statistik stattfindet und die Integration statistischer und geografischer Daten Einzug hält. Um deren Nutzung durch geeignete Infrastrukturelemente zu ermöglichen und zu vereinfachen, wird aktuell an verschiedenen, nachfolgend erläuterten Projekten gearbeitet.


Hierbei gilt es, die Herausforderungen der Phase der Georeferenzierung zu meistern: Der Prozess der Findung neuer Auswertungs- und Analyseideen, die sich in tatsächlich realisierbare, neue Produkte und Verfah-

ren überführen lassen, erfordert, das Spezialwissen über die jeweilige Fachstatistik mit dem Know-how von Geoexpertinnen und -experten zusammenzubringen. Zugleich sind die dabei zu entwickelnden Aufbereitungsverfahren technisch so umzusetzen, dass bereits bestehende Prozessabläufe in der Datengewinnung um Geofunktionalitäten angereichert werden, ohne dass hierfür ressourcenintensive Datenexporte in Geoinformationssysteme und anschließende Re-Importe in die Fachverfahren erforderlich werden. Für dieses Ziel arbeitet der Statistische Verbund im Arbeitskreis Geo eng zusammen. Um den Prozess der Georeferenzierung bestmöglich zu unterstützen, wird derzeit die bestehende GIS-IT-Infrastruktur im Statistischen Bundesamt zu einer integrierten Geo-Plattform ausgebaut. Diese soll es erlauben, die bestehenden professionellen Geofunktionalitäten des GIS-Arbeitsplatzes in Form standardisierter Geoprocessing-, Daten- und Kartendienste für Produktionsverfahren verfügbar zu machen. Zugleich kann die Geo-Plattform als Austauschplattform für Analyse- und Auswertungsideen genutzt werden.

Ein weiteres Projekt im Kontext der Erweiterung der Auswertungsinfrastruktur ist der Aufbau einer sogenannten Gitterzellendatenbank in Kooperation mit dem [Bundesamt für Kartographie und Geodäsie](#) und dem [Johann Heinrich von Thünen-Institut](#). Die Datenbank soll standardisiert Informationen bereitstellen, mit denen wissenschaftliche statistische Analysen in einen räumlichen Kontext eingebettet werden können, der die Auswertung neuer Fragestellungen ermöglicht und somit den Nutzen der Statistiken erhöht. Ein Beispiel für die Nutzung der Gitterzellendatenbank liefert der oben beschriebene PoC zu den Erreichbarkeiten von Grundschulen: Die Erhebungsmerkmale des Mikrozensus dürfen gemäß Bundesstatistikgesetz räumlich nicht gebäude-, sondern lediglich gitterzellenscharf gespeichert werden. Daher werden die Erreichbarkeiten den Gitterzellen als deren Raumeigenschaften zugeordnet.

Die Gitterzellendatenbank soll externe sowie statistikinterne raumbezogene Daten (auf Basis geografischer Koordinaten, Gitterzellen beziehungsweise eines georeferenzierbaren Adressbestands) zentral und datenbankbasiert zur Verfügung stellen. Die Bezugsgröße der Datenbank ist hierbei die (100 m x 100 m-)INSPIRE-Gitterzelle, deren eindeutige ID als Schlüssel die Verschneidung der Daten untereinander und mit entsprechend geokodierten Statistiken erlaubt. Zu den raumbe-

zogenen Merkmalen der Datenbank zählen Punktdaten (Points of Interest, POI), beispielsweise Informationen zu Infrastruktureinrichtungen, und Flächendaten (Structures & Spheres of Interest, SOI), zum Beispiel Lärm- oder Wetterdaten. Diese werden den Gitterzellen zugeordnet (POI) oder nach einem definierten Konzept auf diese verschnitten (SOI). Sie dienen auch als Basis für die Ableitung weiterer raumbezogener Daten, für POI sind dies typischerweise Erreichbarkeitsinformationen basierend auf einem Wegenetz. Die Datenhaltung und -verarbeitung sowie der Nutzerkreis unterscheiden sich je nach Quelle der Inhalte. Auf statistikinternen Informationen basierende Daten können auch nur statistikintern verarbeitet und genutzt werden. Inhalte, deren Ursprung nicht die amtliche Statistik ist, können außerhalb des Statistischen Verbunds verarbeitet und bereitgestellt werden und – unter Berücksichtigung (lizenz-) rechtlicher Fragen – einem breiteren Nutzerkreis (zum Beispiel anderen Bundesbehörden, Forschung und Wissenschaft und so weiter) zur Verfügung gestellt werden.

Durch die Umsetzung der Geokodierung im Statistischen Verbund wurde in den vergangenen Jahren eine ausgezeichnete Datengrundlage geschaffen. Sie ermöglicht es, statistische Informationen adäquat, zeitgemäß und entsprechend den internationalen Standards bereitzustellen und zu verbreiten. Die Nutzung dieser Datenquellen, kombiniert mit dem weiteren Aufbau der Infrastruktur und von Know-how im Statistischen Verbund zur Analyse integrierter statistischer und geografischer Informationen, bietet großes Potenzial für die Umsetzung weiterer konkreter Produkte. 

### LITERATURVERZEICHNIS

---

Gebers, Kathrin/Graze, Philip. *Statistische Datengewinnung durch die Nutzung geografischer Informationen*. In: WISTA Wirtschaft und Statistik. Ausgabe 4/2019, Seite 11 ff.

Gehle, Christian/Lüüs, Hans-Peter. *Prozessmanagement im Statistischen Bundesamt*. In: WISTA Wirtschaft und Statistik. Ausgabe 5/2017, Seite 46 ff.

Klumpe, Bettina/Schröder, Jette/Zwick, Markus. *Qualität bei zusammengeführten Daten. Befragungsdaten, administrative Daten, neue digitale Daten: miteinander besser?* Wiesbaden 2020.

UN-GGIM:Europe. *The integration of statistical and geospatial information – a call for political action in Europe*. Ausgabe 2019. [Zugriff am 9. November 2020]. Verfügbar unter: [ec.europa.eu](https://ec.europa.eu)

### 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 6 des Gesetzes vom 10. Juli 2020 (BGBl. I Seite 1648) geändert worden ist.

Gesetz zur Förderung der elektronischen Verwaltung (E-Government-Gesetz – EGovG) vom 25. Juli 2013 (BGBl. I Seite 2749), das zuletzt durch Artikel 15 des Gesetzes vom 20. November 2019 (BGBl. I Seite 1626) geändert worden ist.

Richtlinie 2007/2/EG des Europäischen Parlaments und des Rates vom 14. März 2007 zur Schaffung einer Geodateninfrastruktur in der Europäischen Gemeinschaft (INSPIRE) (Amtsblatt der EU Nr. L 108, Seite 1).



Sandra Schymura

hat Sozialwissenschaft an der Ruhr-Universität Bochum studiert und arbeitet seit 2019 als wissenschaftliche Mitarbeiterin im Referat „Verdienststrukturerhebung, Arbeitskostenerhebung, Tarifstatistiken“ des Statistischen Bundesamtes. Ihr Aufgabenbereich umfasst hauptsächlich die Verdiensterhebung nach § 7 Absatz 1 Bundesstatistikgesetz.

# BESCHÄFTIGTE UND IHRE VERDIENSTE NACH DER ZWEITEN ERHÖHUNG DES MINDESTLOHNS

Sandra Schymura

↘ **Schlüsselwörter:** Mindestlohn – Verdienststrukturen – Bruttostundenverdienste – Betriebsbefragung – Mindestlohnerhöhung

## ZUSAMMENFASSUNG

Die Statistischen Ämter des Bundes und der Länder führten 2019 zum vierten Mal eine freiwillige Verdiensterhebung durch. Im Auftrag des Bundesministeriums für Arbeit und Soziales sollte festgestellt werden, wie sich die zweite Erhöhung des Mindestlohns auf die Verdienststrukturen ausgewirkt hat. Die Variable „Tarifbindung des Betriebs“ wird in den Verdiensterhebungen nicht erhoben, könnte aber ein wichtiges Klassifikationsmerkmal bei Beschäftigungsentwicklung und Verteilung der Bruttostundenverdienste sein. Diese Annahme führte zu der Frage, ob sich dieses Merkmal nachträglich ergänzen ließe. Der Beitrag berichtet über die Vorüberlegungen zur Imputation, deren Durchführung und die Prüfung der Qualität. Anschließend folgt eine Analyse der Verdienste mithilfe der Variablen.

↘ **Keywords:** *minimum wage – structure of earnings – gross hourly earnings – survey of local units – increase of minimum wage*

## ABSTRACT

*In 2019, the statistical offices of the Federation and the Länder carried out a fourth voluntary earnings survey. The survey was commissioned by the Federal Ministry of Labour and Social Affairs in order to examine how the second increase in the minimum wage has affected earnings structures. The variable “local unit covered by collective bargaining” is not collected in the earnings surveys but could be an important classification feature with regard to the development of employment and the distribution of gross hourly earnings. This led to the question whether the variable could be added subsequently. This contribution reports on the preliminary considerations concerning the imputation, its implementation and the quality tests. Then the earnings are analysed with the help of this variable.*

## 1

---

### Einleitung

---

Am 16. August 2014 ist das Gesetz zur Regelung eines allgemeinen Mindestlohns (Mindestlohngesetz) in Kraft getreten. Ab dem 1. Januar 2015 betrug der Mindestlohn 8,50 Euro brutto je Arbeitsstunde, die erste Anpassung des Mindestlohns erfolgte zum 1. Januar 2017 und sah eine Erhöhung des Mindestlohns auf 8,84 Euro vor. Mit der zweiten Anpassungsentscheidung stieg am 1. Januar 2019 der Mindestlohn auf 9,19 Euro und zum 1. Januar 2020 auf 9,35 Euro brutto je Arbeitsstunde. Die vierjährlich, zuletzt für 2018 durchgeführte Verdiensterhebung liefert Angaben zur Verdienstsituation der Beschäftigten, jedoch keine Aussagen zu Zeitpunkten dazwischen. Daher hat das Bundesministerium für Arbeit und Soziales die Statistischen Ämter des Bundes und der Länder wiederholt beauftragt, Erhebungen nach § 7 Absatz 1 Bundesstatistikgesetz durchzuführen. Diese Rechtsgrundlage ermöglicht einer obersten Bundesbehörde, kurzfristig eine Bundesstatistik ohne Auskunftspflicht in Auftrag zu geben. Insgesamt wurden vier Verdiensterhebungen für die Berichtsjahre 2015, 2016, 2017 und 2019 durchgeführt, um die Datenlücken für die jeweiligen Jahre zu schließen.

Die Beteiligung an diesen Erhebungen ist freiwillig und die Belastung der Auskunftgebenden soll niedrig gehalten werden – auch, um eine höhere Teilnahmequote zu erzielen. Daher wird in den Verdiensterhebungen im Vergleich zu den Verdienststrukturerhebungen eine geringere Anzahl an Merkmalen erhoben. Die Tarifbindung des Betriebs ist eine der aus diesen Überlegungen heraus nicht erhobenen Variablen. Sie interessiert als Klassifizierungsmerkmal für Analysen des Verdienstes, da Unterschiede in der Lohnentwicklung bei Beschäftig-

#### ↳ Exkurs

Als tarifgebunden wird ein Betrieb gezählt, wenn er seine Beschäftigten nach einem Branchen- oder Firmentarifvertrag vergütet. Ein Beschäftigungsverhältnis wird entsprechend als tarifgebunden klassifiziert, wenn die innehabende Person in einem Betrieb arbeitet, in welchem nach einem Branchen- oder Firmentarifvertrag vergütet wird. Dies bedeutet jedoch nicht, dass alle Arbeitnehmerinnen und Arbeitnehmer eines tarifgebundenen Betriebs tatsächlich einem Branchen- oder Firmentarifvertrag unterliegen.

ten in tarifgebundenen Betrieben gegenüber nicht tarifgebundenen Betrieben zu erwarten sind. Zur Verdiensterhebung 2015 meldeten nur Betriebe, welche bereits an der Verdienststrukturerhebung 2014 teilgenommen hatten; so war es möglich, Angaben über die Betriebe wie die Tarifbindung zu übernehmen (Frentzen/Günther, 2017, hier: Seite 26). Ein großer Teil der Stichprobe der Verdiensterhebung 2019 umfasst Betriebe, die bereits für die Verdienststrukturerhebung 2018 befragt wurden. Daraus entstand die Idee, dies zu nutzen, um die Variable Tarifbindung des Betriebs in der Verdiensterhebung 2019 nachträglich zu ergänzen. Kapitel 2 stellt zunächst kurz die Konzeption der Verdiensterhebung sowie die Datengrundlage im Berichtsjahr 2019 vor. In Kapitel 3 erfolgt ein Zeitvergleich ohne Tarifbindung. Anschließend erläutert Kapitel 4 die Vorüberlegungen, das Vorgehen bei der Imputation der Variable Tarifbindung und die Prüfung der Qualität. Der Aufsatz beantwortet somit auch die Frage, ob die Tarifbindung von Betrieben einen Unterschied bei den Verdiensten der Beschäftigten nach der zweiten Erhöhung des Mindestlohns macht.

## 2

---

### Datenbasis

---

Mit der Verdiensterhebung 2019 wurden wie bei den vorherigen Verdiensterhebungen personenbezogene Daten über Bruttoverdienste, Arbeitszeiten und verdiensterklärende Merkmale für den Berichtsmont April erhoben. Analog zu den regulären Verdienststatistiken wurden diese arbeitnehmerbezogenen Angaben bei den Arbeitgebern (Betrieben) erfragt, die diese Daten der betrieblichen Entgeltabrechnung beziehungsweise der Personalverwaltung entnehmen. Im Fokus der Verdiensterhebung als Sondererhebung nach § 7 Bundesstatistikgesetz stand die Zahl der Beschäftigungsverhältnisse (Jobs), in denen der gesetzliche Mindestlohn gezahlt wurde.

Für die Verdiensterhebung 2019 wurden insgesamt 83 150 Betriebe angeschrieben. Die Stichprobe bestand dabei aus zwei Teilen. Einen Teil bildeten die 53 150 Betriebe, die bis zum 15. August 2019 zur Verdienststrukturerhebung 2018 gemeldet hatten. Diese kannten entsprechend den Ablauf der Erhebung bereits in ähnlicher Form und sollten somit leicht zu der freiwilligen Teilnahme zu motivieren sein. Der zweite Teil umfasste

30 000 Betriebe einer Ergänzungsstichprobe, die nicht bei der Verdienststrukturerhebung 2018 ausgewählt worden waren.

Insgesamt beteiligten sich 7 206 Betriebe auf freiwilliger Basis an der Erhebung, das waren 8,7 % der angeschriebenen Betriebe. Jedoch lieferten 207 Betriebe aus unterschiedlichen Gründen keine verwertbaren Angaben.<sup>1</sup> Wie in den Verdienststatistiken üblich, wurden ergänzend weitere Angaben aus anderen Datenquellen genutzt. Bei 2 000 Betrieben handelte es sich um eine Stichprobe von Betrieben mit ausschließlich geringfügig entlohnten Beschäftigten, die aus dem Verwaltungsdatenspeicher<sup>2</sup> gezogen wurde. Die Angaben zu den Arbeitnehmerinnen und Arbeitnehmern dieser Betriebe wurden auf Basis des Nearest-Neighbor-Verfahrens imputiert. Als Spender dienten hierbei Angaben zu geringfügig Beschäftigten<sup>3</sup> aus der Verdiensterhebung. Die Angaben zu den Beschäftigten des öffentlichen Dienstes<sup>4</sup> wurden ebenfalls nicht erhoben, sondern aus der Personalstandstatistik abgeleitet. Hier wurden die Merkmale, die nicht in der Personalstandstatistik verfügbar sind, ergänzend zugeschätzt, annahmebasiert abgeleitet beziehungsweise imputiert. Die Verdiensterhebung 2019 umfasst daher etwa 99 000 Arbeitneh-

merdatensätze, hochgerechnet auf rund 41 Millionen Beschäftigungsverhältnisse.<sup>5</sup> [↘ Tabelle 1](#)

### 3

## Ergebnisse im Zeitvergleich

Thema dieses Kapitels ist der Einfluss der zweiten Mindestlohnerhöhung auf die Verdienststrukturen. Dazu werden die Ergebnisse der Verdiensterhebung 2019 mit denen der Verdienststrukturerhebung 2018 verglichen. Für die Zeitreihe erfolgt zudem eine Analyse mit Daten aus der Verdienststrukturerhebung 2014 sowie den Verdiensterhebungen 2015 bis 2017.

### 3.1 Entwicklung der Beschäftigung

Im Fokus der Verdiensterhebungen befinden sich die Beschäftigungsverhältnisse im Geltungsbereich des Mindestlohngesetzes. Mit den in den Verdienststatistiken zur Verfügung stehenden Instrumenten lässt sich der Geltungsbereich jedoch nicht eindeutig bestimmen. Dies liegt an Ausnahmeregelungen, die eine Bezahlung unter Mindestlohn zulassen. Zu den Ausnahmen vom Mindestlohn zählen Auszubildende, Langzeitarbeitslose, unter bestimmten Voraussetzungen Praktikantinnen und Praktikanten sowie Personen unter 18 Jahren.<sup>6</sup> Die folgenden Auswertungen schließen diese Ausnahmen weitestgehend aus und analysieren

- 1 Die Gründe sind im Einzelnen aufgeschlüsselt auf Seite 8 des Methodenberichts der Verdiensterhebung 2019 (Statistisches Bundesamt, 2020).
- 2 Die Bundesagentur für Arbeit liefert auf Basis des Verwaltungsdatengesetzes aus ihrer Beschäftigtenstatistik monatlich betriebsbezogene Beschäftigtendaten an die statistischen Ämter. Diese Angaben umfassen unter anderem die Anzahl der ausschließlich geringfügig entlohnten Beschäftigten. Eine detaillierte Erläuterung findet sich bei Lorenz/Opfermann (2017, hier: Seite 54 f.).
- 3 Dies ist methodisch plausibel, da durch die gesetzlichen Vorgaben die Personengruppe geringfügig Beschäftigter in beiden Datenquellen eine ähnliche Struktur aufweist.
- 4 Wirtschaftsabschnitte O „Öffentliche Verwaltung, Verteidigung; Sozialversicherung“ und P „Erziehung und Unterricht“ (ohne private Einrichtungen) der Klassifikation der Wirtschaftszweige, Ausgabe 2008 (WZ 2008).

- 5 Weitere Informationen zu Methoden und Eckdaten finden sich im Methodenbericht der Verdiensterhebung 2019 (Statistisches Bundesamt, 2020).
- 6 Nähere Erläuterungen dazu enthalten § 22 Mindestlohngesetz sowie die Methodenberichte der Verdiensterhebungen (Statistisches Bundesamt, 2017a, 2017b, 2018 und 2020).

**Tabelle 1**

Datensätze der Verdiensterhebung 2019 nach Herkunft der Daten

	Betriebe		Jobs	
	Fallzahl	Hochrechnung	Fallzahl	Hochrechnung
	Anzahl	Mill.	Anzahl	Mill.
Insgesamt	9 749	2 453	99 134	40 885
Erhebung	6 999	2 087	70 821	35 178
Imputation (Betriebe ohne sozialversicherungspflichtig Beschäftigte)	2 000	365	4 674	945
Berechnung (Personalstandstatistik) <sup>1</sup>	X	X	23 639	4 726

1 Die Personalstandstatistik enthält keine Angaben über Betriebe.

## Beschäftigte und ihre Verdienste nach der zweiten Erhöhung des Mindestlohns

nur die vom Mindestlohngesetz erfassten Beschäftigungsverhältnisse.

Im Jahr 2018 lagen 37,9 Millionen abhängige Beschäftigungsverhältnisse im Geltungsbereich des Mindestlohngesetzes, im Jahr 2019 betraf dies 39,4 Millionen Jobs; somit war ein Anstieg um 1,5 Millionen Jobs zu verzeichnen. Zwischen 2017 und 2019 betrug der Anstieg 2,2 Millionen Jobs. [↘ Tabelle 2](#)

Da der Bruttostundenverdienst nicht direkt erfragt wird, sind bei dessen Bestimmung Unschärfen möglich. Der Bruttostundenverdienst wird mithilfe des Bruttomonatsverdienstes und der bezahlten Arbeitszeit berechnet. Die Erfassung der Arbeitszeit ist dabei mit größeren Unsicherheiten verbunden als die Erfassung der Verdienste. Um dieser Unschärfe Rechnung zu tragen, legt das Statistische Bundesamt für Auswertungen ein 10-Cent-Intervall um den Mindestlohn. Für die Jobs im Mindestlohnbereich in den Jahren 2015 und 2016 umfasste das Intervall 8,45 bis 8,54 Euro, in den Jahren 2017 und 2018 lag es bei 8,79 bis 8,88 Euro und im Jahr 2019 bei 9,14 bis 9,23 Euro. Für das Jahr 2014 wurde für Vergleichszwecke ein fiktiver Mindestlohn in Höhe von 8,50 Euro unterstellt und ein Intervall von 8,45 bis 8,54 Euro um die Jobs im Mindestlohnbereich gelegt.

Während die Verdiensterhebungen 2017 und 2019 rund 1,4 Millionen Jobs im Mindestlohnbereich nachwiesen, waren es bei der Verdienststrukturerhebung 2018 lediglich 0,9 Millionen Beschäftigungsverhältnisse. Bei der Betrachtung ist jedoch zu berücksichtigen, dass zum

1. Januar 2017 der Mindestlohn auf 8,84 Euro erhöht worden war, während er im Jahr 2018 konstant blieb, bevor dann zum 1. Januar 2019 die Erhöhung auf 9,19 Euro folgte.

Als „unter Mindestlohn“ bezahlt galten – entsprechend der oben festgelegten Intervalle – in den Erhebungen für die Jahre 2014, 2015 und 2016 Jobs mit einer Entlohnung von weniger als 8,45 Euro brutto die Stunde. In den Erhebungen für die Jahre 2017 und 2018 lagen Jobs, bei denen der Bruttostundenverdienst weniger als 8,79 Euro betrug, unter der Mindestlohngrenze. Gleiches galt in der Verdiensterhebung 2019 für die Jobs, in denen weniger als 9,14 Euro brutto pro Stunde gezahlt wurde. Sowohl in tarifgebundenen als auch in nicht tarifgebundenen Betrieben könnte ein Teil der Jobs, in denen weniger als der Mindestlohn gezahlt wurde, auf die oben erwähnten Unschärfen zurückzuführen sein. Verstöße gegen das Mindestlohngesetz sind jedoch nicht auszuschließen. Unschärfen entstanden in den früheren Erhebungen auch dadurch, dass es bis Ende 2017 branchenspezifische Übergangsregelungen gab, welche beim Vorhandensein eines allgemeinverbindlichen Tarifvertrags eine Bezahlung unter Mindestlohn erlaubten. Durch den Wegfall der Übergangsregelungen konnten diese Unschärfen in den Jahren 2018 und 2019 nicht mehr auftreten, dies könnte zumindest teilweise die gesunkene Zahl an Jobs unter Mindestlohn erklären. Während es hiervon im Jahr 2017 noch 832 000 gab, wurden im Jahr 2018 in 483 000 Jobs und im Jahr 2019 in 527 000 Beschäftigungsverhältnissen Stundenverdienste unter Mindestlohn bezahlt.

**Tabelle 2**

### Entwicklung der Jobs im Zeitverlauf

	Verdienststruktur- erhebung 2014	Verdienst- erhebung 2015	Verdienst- erhebung 2016	Verdienst- erhebung 2017	Verdienststruktur- erhebung 2018	Verdienst- erhebung 2019
	1 000					
Jobs, für die das Mindestlohngesetz gilt	35 613	36 477	36 444	37 111	37 883	39 352
Mit Tarifbindung	16 019	15 847	X	X	16 412	X
Ohne Tarifbindung	19 594	20 630	X	X	21 471	X
Darunter:						
Jobs, die im Mindestlohnbereich liegen	395	1 907	1 754	1 371	926	1 421
Mit Tarifbindung	113	165	X	X	94	X
Ohne Tarifbindung	282	1 742	X	X	832	X
Jobs unter der Mindestlohngrenze	3 849	1 014	751	832	483	527
Mit Tarifbindung	672	236	X	X	85	X
Ohne Tarifbindung	3 177	778	X	X	398	X

Zur Schaffung einer differenzierteren Vergleichsbasis wurde für 2014 ein fiktiver Mindestlohn in Höhe von 8,50 Euro unterstellt. In den Methodenberichten wurde lediglich der Bereich unter 8,50 Euro abgegrenzt. Die Darstellung in diesem Artikel ist daher abweichend.



Werden die Jobs im Mindestlohnbereich und unter der Mindestlohngrenze zusammen vom Jahr der Mindestlohneinführung 2015 an betrachtet, wird zudem deutlich, dass ihre Zahl stark abgenommen hat. Während im Jahr 2015 noch 2,9 Millionen Beschäftigungsverhältnisse in diese Bereiche fielen, waren es im Jahr 2019 nur noch 1,9 Millionen. Anders ausgedrückt: Betrug der Anteil der Jobs im und unter dem Mindestlohn im Jahr 2015 noch 8,0%, waren es im Jahr 2019 nur noch 5,0%. In Verbindung mit dem Anstieg der Jobs im Geltungsbereich des Mindestlohngesetzes deutet das darauf hin, dass für eine große Zahl von Jobs mittlerweile mehr als Mindestlohn bezahlt wird beziehungsweise neue Jobs oberhalb des Mindestlohnbereichs entstanden sind.

### 3.2 Verteilung der Jobs nach Stundenverdiensten

Die zweite Erhöhung des allgemeinen gesetzlichen Mindestlohns im Januar 2019 auf 9,19 Euro wirkte sich auf die Verteilung der Jobs nach Bruttostundenverdiensten in Form einer Verschiebung in Richtung höhere Bruttostundenverdienste aus. Die Höhe des Mindestlohns in den Jahren 2017 und 2018 war gleich, trotzdem ergaben sich auch zwischen diesen beiden Jahren unterschiedliche Verteilungen. Im Jahr 2018 lag lediglich bei rund

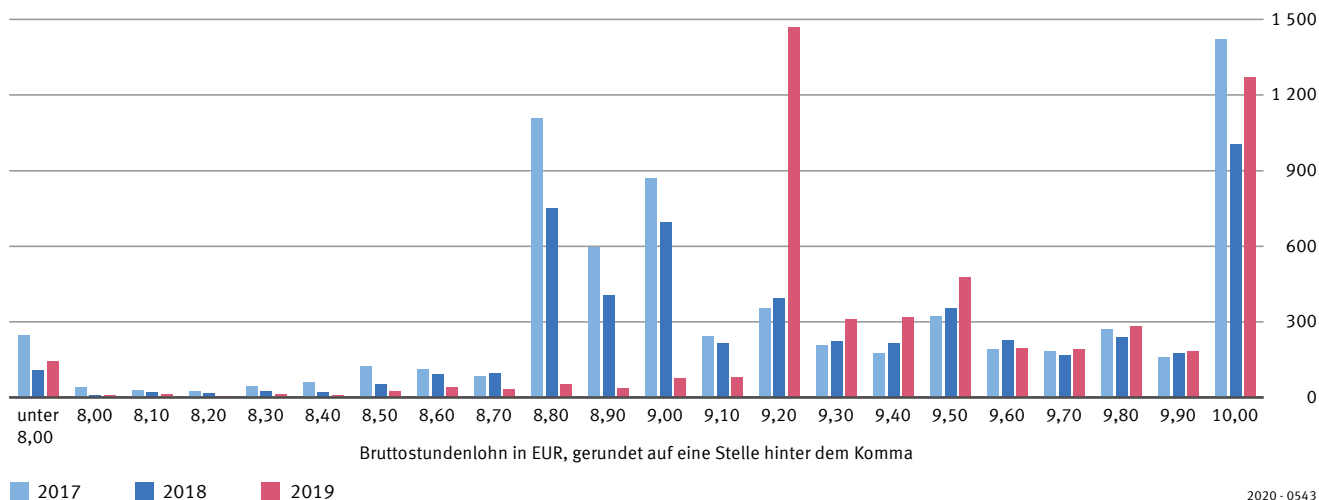
10 Euro ein Peak mit einer Million betroffenen Beschäftigungsverhältnissen. Im Gegensatz dazu gab es im Jahr 2017 mit 1,1 Millionen Jobs bei rund 8,80 Euro und 1,4 Millionen Jobs bei rund 10 Euro Höhepunkte. Im Jahr 2019 lagen die Spitzenwerte mit 1,5 Millionen Jobs bei rund 9,20 Euro und knapp 1,3 Millionen Jobs bei rund 10 Euro. [↪ Grafik 1](#)

Bei der Betrachtung der Grafik ist zu berücksichtigen, dass die Zahl der Beschäftigungsverhältnisse in den letzten Jahren kontinuierlich gestiegen, die Zahl der Jobs im Bereich bis zu 10 Euro jedoch gesunken ist. Fielen im Jahr 2017 noch 6,8 Millionen Jobs in diesen Bereich, waren es im Jahr 2018 nur noch 5,4 Millionen Jobs und im Jahr 2019 lediglich 5,1 Millionen Jobs. Dies deutet darauf hin, dass in vielen Beschäftigungsverhältnissen bereits im Jahr 2018, möglicherweise aufgrund von regulären Lohnerhöhungen, Bruttostundenverdienste über dem Mindestlohn bezahlt wurden.

Eine Analyse der Jobs im Bereich von 10 bis 12 Euro bestätigt das. Obwohl sich die Zahl der Beschäftigungsverhältnisse in diesem Bereich zwischen den Jahren 2017 mit 5,6 Millionen Jobs und 2018 mit 5,8 Millionen Jobs kaum veränderte, verschob sich die Verteilung in höhere Stundenverdienste. Der Anstieg in diesem Bereich auf 6,2 Millionen Jobs im Jahr 2019 ist vermutlich zu großen Teilen auf die Entstehung von neuen Beschäftigungsverhältnissen zurückzuführen.

#### Grafik 1

Beschäftigungsverhältnisse nach der Höhe des Bruttostundenlohns im Bereich von bis unter 8 Euro bis 10 Euro  
Verdienststrukturerhebung 2018 sowie Verdiensterhebungen 2017 und 2019, in 1 000



### 4

## Imputation und Analyse der Tarifbindung

Der Mindestlohn wurde unter anderem eingeführt, um der abnehmenden Tarifbindung entgegenzuwirken und eine Lohnuntergrenze für Jobs in nicht tarifgebundenen Betrieben zu schaffen (Oschmiansky, 2014). Dementsprechend sollten vor allem Jobs in nicht tarifgebundenen Betrieben von Erhöhungen des Mindestlohns profitieren. Um zu erkennen, wie sich die Erhöhung des Mindestlohns je nach Tarifbindung auswirkt, interessiert sowohl bei der Entwicklung der Beschäftigung als auch bei der Verteilung der Stundenverdienste eine Auswertung nach Tarifbindung für 2019. Auch wären Vergleiche mit den Jahren 2014, 2015 und 2018 möglich. Da jedoch bei der freiwilligen Verdiensterhebung im Gegensatz zur Verdienststrukturerhebung nur ein reduziertes Spektrum an Merkmalen erhoben wird, liegen Informationen zur betrieblichen Tarifbindung originär nicht vor. Um dennoch Aussagen zur Tarifbindung zu ermöglichen und Analysen durchzuführen, erfolgte die nachfolgend beschriebene Imputation der Variable.

### 4.1 Vorgehen bei der Imputation

Wie bereits in Kapitel 2 geschildert, bestand die Stichprobe der Verdiensterhebung 2019 zu einem großen Teil aus Betrieben, die bereits an der Verdienststrukturerhebung 2018 teilgenommen hatten. Für diese Betriebe lässt sich mit großer Wahrscheinlichkeit vermuten, dass die Ausprägung der Tarifbindung des Betriebs 2019 dem Stand 2018 entspricht.<sup>7</sup> Dies trifft auf 5 104 Betriebe (70,8%) zu, für welche die Ausprägung des Merkmals der Tarifbindung daher aus der Verdienststrukturerhebung 2018 übertragen werden konnte. Nicht an der Verdienststrukturerhebung 2018, jedoch an der Verdienststrukturerhebung 2014 teilgenommen hatten 277 Betriebe (3,8%). In diesen Fällen wurde die Ausprägung der Tarifbindung der Betriebe von 2014 in die Verdiensterhebung 2019 übernommen unter der Annahme, dass die Tarifbindung in den vergangenen Jahren zwar abgenommen hat, aber dennoch ein relativ konstanter

Wert zu unterstellen ist (Ellguth/Kohaut, 2019, hier: Tabelle 9). Die Betriebe aus der Personalstandstatistik gehören dem öffentlichen Dienst an und sind daher tarifgebunden.

Für die übrigen 3 825 Betriebe aus der Verdiensterhebung 2019, für die keine Vorinformationen vorlagen, wurde eine Imputation durchgeführt. Hierzu zählen auch die Betriebe mit ausschließlich geringfügig entlohnten Beschäftigten, die deshalb in den Verdiensterhebungen nicht direkt erhoben werden. Im Zuge der Imputation wurde für diese Betriebe ohne Vorinformationen auf die Daten ähnlicher Betriebe zurückgegriffen, für die das Merkmal der Tarifbindung vorliegt. Zur Anwendung kam das Programm CANCEIS (CANadian Census Edit and Imputation System) mit dem Nearest-Neighbour-Verfahren.<sup>8</sup> Methodisch handelt es sich dabei um eine Single-Imputation, bei welcher für jeden fehlenden Wert ein dem Empfängerdatensatz möglichst ähnlicher Spenderdatensatz gesucht wird und der zu imputierende Wert dann aus diesem übernommen wird. Bei der Suche nach geeigneten Spendern wurden der amtliche Gemeindegemeinschaftsschlüssel, der Regionalschlüssel, die Unternehmensgröße (Anzahl der Mitarbeiterinnen und Mitarbeiter) und der Wirtschaftszweig mit unterschiedlichen Distanzfunktionen und Gewichtungen berücksichtigt. Dahinter steht die Annahme, dass sich geografisch nahe, im selben oder verwandten Wirtschaftszweig tätige und von der Größe vergleichbare Betriebe hinsichtlich der Tarifbindung besonders ähnlich sind. Als Spenderdaten wurden die Fälle aus der Verdienststrukturerhebung 2018 herangezogen, welche erhoben und nicht imputiert<sup>9</sup> worden sind. Somit handelt es sich um eine Cold-Deck-Imputation, da Werte imputiert werden, welche nicht aus derselben Erhebung stammen. Insgesamt gab es so 56 140 Betriebe als potenzielle Spender.

Um die Qualität der Imputation zu prüfen, sind zwei Verfahren zum Einsatz gekommen. Im ersten Verfahren wurde bei einem Drittel der Fälle der Wert für die Tarifbindung aus der Verdienststrukturerhebung 2018 zufällig gelöscht, um diese gelöschten Werte mit CANCEIS zu imputieren. Dabei zeigte sich, dass die Differenz zwischen der tatsächlichen und der imputierten Verteilung minimal ist. In allen Simulationen betrug die Verände-

<sup>7</sup> Das belegen zum Beispiel Tabelle 1 in Ellguth/Kohaut (2019) und Tabelle 1 in Ellguth/Kohaut (2020).

<sup>8</sup> Das Programm und das Vorgehen werden ausführlich in Frentzen/Günther (2017) auf Seite 36 ff. beschrieben.

<sup>9</sup> Imputiert wurden Betriebe mit ausschließlich geringfügig Beschäftigten und Betriebe aus der Personalstandstatistik.

rung bei den Anteilen zwischen tarifgebunden und nicht tarifgebunden weniger als einen halben Prozentpunkt. Im zweiten Verfahren wurden die Verteilungen aus den Vorjahren übernommen und mit den imputierten Werten in der Verdiensterhebung 2019 verglichen. Die Differenz zwischen der aus den Vorjahren übernommenen und der imputierten Verteilung beträgt hier rund zwei Prozentpunkte und weicht somit schon deutlicher ab. Es fällt auf, dass die imputierten Betriebe seltener tarifgebunden sind, sodass tendenziell der Anteil der tarifgebundenen Betriebe unterschätzt wird.

## 4.2 Analyse der Tarifbindung

Im Vergleich der Jahre 2018 und 2019 zeigt sich, dass der Anteil der Jobs in tarifgebundenen Betrieben insgesamt um 2,5 Prozentpunkte von 43,3 auf 45,8% zugenommen hat. Von 2014 bis 2015 nahm im Vergleich dazu der Anteil der Jobs in tarifgebundenen Betrieben von 45,0 auf 43,4% um 1,6 Prozentpunkte ab. Effekte lassen sich auch für die Jobs im Mindestlohnbereich beobachten, der Anteil in tarifgebundenen Betrieben nahm von 2018 auf 2019 um 4,3 Prozentpunkte zu. Von 2014 zu 2015 erhöhte sich infolge der Einführung des gesetzlichen Mindestlohns die Zahl der Jobs im Mindestlohnbereich in tarifgebundenen Betrieben um fast 1,5 Millionen. In tarifgebundenen Betrieben belief sich die Zunahme nur auf rund 50 000 Jobs. Bei den Jobs unterhalb der Mindestlohngrenze nahm der Anteil in tarifgebundenen Betrieben zwischen 2018 und 2019

um 6,2 Prozentpunkte zu, zwischen 2014 und 2015 hatte er sich um 5,8 Prozentpunkte erhöht. [↘ Tabelle 3](#)

Zum Vergleich wurde das IAB-Betriebspanel herangezogen, welches die Tarifbindung von Beschäftigten erhebt und bezüglich der Tarifbindung als belastbarer als die Verdiensterhebungen angesehen werden kann. Während sich die Anzahl der Jobs in tarifgebundenen Betrieben nach der Verdiensterhebung 2019 erhöht hat, haben sich diese nach dem IAB-Betriebspanel verringert. Jedoch hat sich der Abstand 2019 zu den prozentualen Anteilen des IAB-Betriebspanels gegenüber 2018 reduziert.<sup>10</sup>

Tieferegehende Auswertungen der Anzahl der Beschäftigungsverhältnisse nach Tarifbindung sind nicht möglich, da fast alle Werte für 2019 aufgrund zu großer statistischer Unsicherheit gesperrt werden müssen. Hier zeigt sich die Problematik eines potenziell zu niedrigen Rücklaufs bei freiwilligen Erhebungen, welche bei einer hohen Varianz zu geringe Fallzahlen für tiefer gegliederte Auswertungen aufweisen.

[↘ Grafik 2](#) zeigt die unterschiedliche Verteilung der Bruttostundenverdienste nach Tarifgebundenheit. Bei den Jobs in nicht an Tarifverträge gebundenen Betrieben liegt der Höhepunkt für 2019 im Intervall von 9 bis 10 Euro, der Anteil der betroffenen Beschäftigungsverhältnisse ist mit 14,0% ziemlich hoch. Das Intervall von

<sup>10</sup> Das belegt die jeweilige Tabelle 1 in Ellguth/Kohaut (2019) und in Ellguth/Kohaut (2020).

**Tabelle 3**

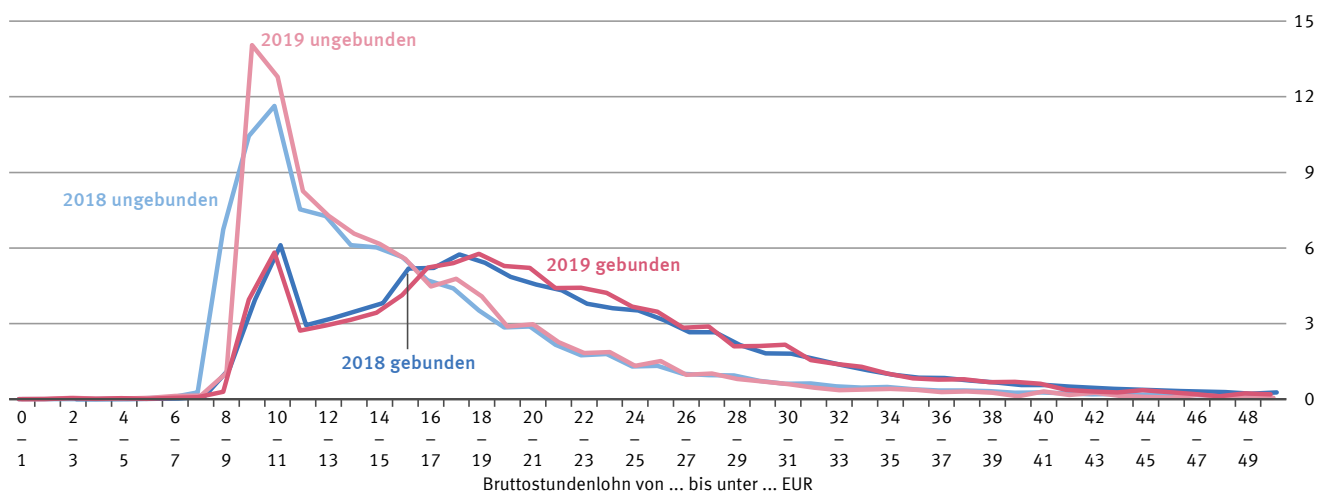
### Entwicklung der Jobs im Zeitverlauf zur Analyse der Tarifbindung

	Verdienststrukturhebung 2014		Verdiensterhebung 2015		Verdienststrukturhebung 2018		Verdiensterhebung 2019	
	1 000	%	1 000	%	1 000	%	1 000	%
Jobs, für die das Mindestlohngesetz gilt	35 613	X	36 477	X	37 883	X	39 352	X
Mit Tarifbindung	16 019	45,0	15 847	43,4	16 412	43,3	18 006	45,8
Ohne Tarifbindung	19 594	55,0	20 630	56,6	21 471	56,7	21 346	54,2
Darunter:								
Jobs, die im Mindestlohnbereich liegen	395	X	1 907	X	926	X	(1 421)	X
Mit Tarifbindung	113	28,6	165	8,7	94	10,2	(206)	14,5
Ohne Tarifbindung	282	71,4	1 742	91,3	832	89,8	(1 215)	85,5
Jobs unter der Mindestlohngrenze	3 849	X	1 014	X	483	X	527	X
Mit Tarifbindung	672	17,5	236	23,3	85	17,6	(125)	23,8
Ohne Tarifbindung	3 177	82,5	778	76,7	398	82,4	(401)	76,2

Zur Schaffung einer differenzierteren Vergleichsbasis wurde für 2014 ein fiktiver Mindestlohn in Höhe von 8,50 Euro unterstellt. In den Methodenberichten wurde lediglich der Bereich unter 8,50 Euro abgegrenzt. Die Darstellung in diesem Artikel ist daher abweichend.

Grafik 2

Verteilung der Bruttostundenverdienste nach Tarifgebundenheit  
Verdienststrukturerhebung 2018 und Verdiensterhebung 2019, in %



2020 - 0544

10 bis 11 Euro ist mit 12,8% Anteil bei den Jobs in nicht tarifgebundenen Betrieben ebenfalls stark besetzt. Ein Vergleich mit dem Vorjahr zeigt, dass der Peak 2018 bei den Jobs in nicht tarifgebundenen Betrieben bei 10 bis 11 Euro lag, jedoch war der Anteil mit 11,6% kleiner. Im Intervall 9 bis 10 Euro betrug der Anteil 10,4%. Dies zeigt, dass bei den Jobs in nicht tarifgebundenen Betrieben durch die Erhöhung des Mindestlohns auf 9,19 Euro im Jahr 2019 von links eine Stauchung der Verteilung stattfand. Anders ausgedrückt: Die Jobs wurden aus den niedrigeren Bereichen in einen höheren, mit dem erhöhten Mindestlohn konform gehenden Bereich „geschoben“.

Bei den Beschäftigungsverhältnissen in tarifgebundenen Betrieben gibt es im Gegensatz dazu zwei Spitzen in der Verteilung. Das ist auf das höhere Lohnniveau der Jobs zurückzuführen. Ein erster Peak liegt im Jahr 2019 bei 10 bis 11 Euro, hiervon sind allerdings nur rund 5,8% aller Jobs in tarifgebundenen Betrieben betroffen. Einen zweiten Höhepunkt gibt es bei 18 bis 19 Euro mit einem Anteil von ebenfalls 5,8%. Jedoch ähnelt dieser mehr einem Plateau, da die Jobs in den Intervallen von 16 bis 21 Euro alle jeweils mehr als 5% Anteil an den Beschäftigungsverhältnissen bei tarifgebundenen Betrieben erreichen. Im Jahr 2018 sah die Verteilung sehr ähnlich aus, ein erster Höhepunkt befand sich mit einem Anteil von 6,1% im Bereich 10 bis 11 Euro. Ebenso lässt sich

ein Hochplateau mit einem Anteil von je mehr als 5% beobachten, dieses befindet sich jedoch im Bereich 15 bis 19 Euro und hat einen Peak bei 17 bis 18 Euro mit einem Anteil von 5,8%. Insgesamt lässt sich eine Verschiebung der Stundenlöhne in höhere Lohnintervalle von 2018 nach 2019 beobachten.

Diese Auswertungen zeigen, dass es bei der Entwicklung der Verdienste sinnvoll ist, die Jobs nach solchen in tarifgebundenen Betrieben und Jobs in nicht tarifgebundenen Betrieben zu klassifizieren. Denn lediglich bei den Jobs in nicht tarifgebundenen Betrieben lässt sich eine unmittelbare Auswirkung des Mindestlohns erkennen. Die Verteilung wurde von links gestaucht, die Jobs mit Bruttostundenlöhnen niedriger als der ab 2019 geltende Mindestlohn von 9,19 Euro in höhere Stundenlöhne verschoben. Bei den Jobs in tarifgebundenen Betrieben verschiebt sich im mittleren Bereich die Verteilung deutlich in Richtung höhere Stundenlöhne, dies trifft bei den niedrigen Stundenlöhnen nur in geringem Umfang zu. Die Verschiebung der Verteilung könnte auf in Tarifverträgen vorgesehene Lohnsteigerungen zurückzuführen sein, die vor der Festlegung des neuen Mindestlohns geschlossen wurden beziehungsweise diesen nicht berücksichtigten.

## 5

### Zusammenfassung und Ausblick

Für das Jahr 2019 führte das Statistische Bundesamt zum vierten Mal eine Sondererhebung der Verdienste gemäß § 7 Absatz 1 Bundesstatistikgesetz durch. Mittels gebundener Hochrechnung konnten trotz einer niedrigen Rücklaufquote aussagekräftige Ergebnisse auf Bundesebene ausgewiesen werden. Auf Basis der Verdienststrukturhebung lagen Daten für das Jahr 2018 vor. Daher war es möglich, die unterschiedlichen Verteilungen vor und nach der zweiten Erhöhung des Mindestlohns zu untersuchen. Die Betrachtung der Ergebnisse im Zeitverlauf zeigte, dass der gesetzliche Mindestlohn auf den unteren Bereich der Verdienste wirkte und die Verdienstverteilung sich in Richtung höherer Stundenlöhne verschoben hat. Es fanden sich außerdem Hinweise, dass der Mindestlohn über sogenannte Spillover-Effekte auch in benachbarten höheren Verdienstbereichen wirkte.

Das Anspielen des Merkmals Tarifbindung des Betriebs gelang durch die Übernahme der Angaben aus den Vorjahren kombiniert mit der Imputation der Werte für Betriebe, die nicht an den Verdienststrukturhebungen 2018 und 2014 teilgenommen hatten. Simulationen und Vergleiche der Verteilungen belegen, dass die Imputation belastbare Ergebnisse liefert. Die Auswertung der Anzahl von Jobs nach Tarifbindung des Betriebs erwies sich als schwierig, da die Hochrechnung bei tiefer gegliederten Auswertungen keine ausreichend belastbaren Werte liefern kann. Jedoch belegen die Verteilungen der Bruttostundenverdienste, dass sich diese bedeutsam bei Jobs in tarifgebundenen Betrieben von den Jobs in nicht tarifgebundenen Betrieben unterscheiden. Während sich die Verteilung der Bruttostundenlöhne bei den Jobs in nicht tarifgebundenen Betrieben entsprechend des veränderten Mindestlohns in höhere Bereiche verschob, ist ein solcher Effekt bei den Jobs in tarifgebundenen Betrieben nur schwach ausgeprägt. Der höhere Mindestlohn bewirkt somit unmittelbar positive Lohn-effekte bei Jobs in nicht tarifgebundenen Betrieben, für Jobs in tarifgebundenen Betrieben lässt sich ein solcher Effekt nicht direkt belegen.

Am 30. Juni 2020 sprach die Mindestlohnkommission Empfehlungen für künftige Erhöhungen des Mindestlohns aus. So soll dieser zum 1. Januar 2021 auf 9,50

Euro, zum 1. Juli 2021 auf 9,60 Euro, zum 1. Januar 2022 auf 9,82 Euro und zum 1. Juli 2022 auf 10,45 Euro brutto je Arbeitsstunde steigen. Mit der am 1. Januar 2021 in Kraft tretenden Änderung des Verdienststatistikgesetzes werden die Verdienststrukturhebung und die Viertel-jährliche Verdiensterhebung durch die neue digitale Verdiensterhebung abgelöst. Diese monatliche Erhebung wird eine regelmäßige Ergebnisdarstellung der Effekte der Mindestlohnanpassung ab dem Berichtsjahr 2022 ermöglichen. Mit einem Stichprobenumfang von rund 58 000 Betrieben, dem Charakter einer Statistik mit Auskunftspflicht und einer damit verbundenen, gegenüber den (freiwilligen) Verdiensterhebungen deutlich geringeren Unit-Non-Responsequote werden dann wesentlich dezidiertere Auswertungen möglich sein – auch im Hinblick auf das Klassifikationsmerkmal Tarifbindung von Betrieben. Aus den monatlichen Arbeitnehmerdatensätzen wird zudem ein Beschäftigtenpanel gebildet, das eine Abgrenzung der vom Mindestlohn betroffenen von einer Kontrollgruppe der vom Mindestlohn nicht betroffenen Beschäftigten für die Mindestlohnforschung zulässt. So ist es künftig möglich, regelmäßig die Auswirkungen von Mindestloohnerhöhungen unter anderem auf die Verdienstverteilung oder die Verdienstentwicklungen genau zu analysieren und auszuwerten. [u](#)

### LITERATURVERZEICHNIS

---

Ellguth, Peter/Kohaut, Susanne. *[Tarifbindung und betriebliche Interessenvertretung: Ergebnisse aus dem IAB-Betriebspanel 2018](#)*. In: WSI Mitteilungen. Ausgabe 4/2019, Seite 290 ff.

Ellguth, Peter/Kohaut, Susanne. *[Tarifbindung und betriebliche Interessenvertretung: Ergebnisse aus dem IAB-Betriebspanel 2019](#)*. In: WSI Mitteilungen. Ausgabe 4/2020, Seite 278 ff.

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

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

Oschmiansky, Frank. *Mindestlöhne: Hintergründe und Positionen*. Berlin 2014. [Zugriff am 8. September 2019]. Verfügbar unter: [www.bpb.de](http://www.bpb.de)

Statistisches Bundesamt. *Verdiensterhebung 2015*. Wiesbaden 2017a. [Zugriff am 8. September 2019]. Verfügbar unter: [www.destatis.de](http://www.destatis.de)

Statistisches Bundesamt. *Verdiensterhebung 2016*. Wiesbaden 2017b. [Zugriff am 8. September 2019]. Verfügbar unter: [www.destatis.de](http://www.destatis.de)

Statistisches Bundesamt. *Verdiensterhebung 2017*. Wiesbaden 2018. [Zugriff am 8. September 2019]. Verfügbar unter: [www.destatis.de](http://www.destatis.de)

Statistisches Bundesamt. *Verdiensterhebung 2019*. Wiesbaden 2020. [Zugriff am 8. September 2019]. Verfügbar unter: [www.destatis.de](http://www.destatis.de)

### RECHTSGRUNDLAGEN

---

Gesetz zur Regelung eines allgemeinen Mindestlohns (Mindestlohngesetz – MiLoG) vom 11. August 2014 (BGBl. I Seite 1348), das zuletzt durch Artikel 2 Absatz 1 des Gesetzes vom 10. Juli 2020 (BGBl. I Seite 1657) geändert worden ist.

Gesetz über die Statistik für Bundeszwecke (Bundesstatistikgesetz – BStatG) in der Fassung der Bekanntmachung vom 20. Oktober 2016 (BGBl. I Seite 2394), das zuletzt durch Artikel 6 des Gesetzes vom 10. Juli 2020 (BGBl. I Seite 1648) geändert worden ist.



**Dr. Sandra Jung**

ist Volkswirtin und leitet das Referat „Analyse der Unternehmensstrukturen“ des Statistischen Bundesamtes. Arbeitsschwerpunkte des Referats sind die Erstellung der Statistik für kleine und mittlere Unternehmen und für auslandskontrollierte Unternehmen sowie Globalisierungsthemen.



**Dr. Wolfhard Kaus**

ist Diplom-Volkswirt und promovierte 2012 am Max-Planck-Institut für Ökonomik. Seit 2014 ist er im Statistischen Bundesamt tätig, seit 2016 als Referent im Referat „Analyse der Unternehmensstrukturen“.

# UNTERNEHMENSSTRUKTURSTATISTIKEN UND STATISTIK FÜR KLEINE UND MITTLERE UNTERNEHMEN NACH DEM EU-UNTERNEHMENSBEGRIFF

## Ergebnisse für 2018

Dr. Sandra Jung, Dr. Wolfhard Kaus

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

### ZUSAMMENFASSUNG

Mit dem Berichtsjahr 2018 ist in den Unternehmensstrukturstatistiken der EU-Unternehmensbegriff neu eingeführt worden. Aufbereitung, Veröffentlichung und Darstellung der Unternehmensstrukturstatistiken haben sich dadurch verändert. Der Beitrag beschreibt die Ergebnisse auf der Grundlage der neuen Unternehmensdefinition und vergleicht sie mit den Ergebnissen auf der Grundlage der bisherigen Unternehmensdefinition (rechtliche Einheiten). Darüber hinaus werden Ergebnisse für kleine und mittlere Unternehmen (KMU) präsentiert. Die Umstellung der Unternehmensdefinition führt zu Qualitätsverbesserungen der KMU-Statistik, weil unechte KMU nun zu großen Unternehmen zusammengefasst werden. Abschließend werden Auswirkungen des Wechsels der Unternehmensdefinition auf Produktivitätskennzahlen aufgezeigt.

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

### ABSTRACT

*As of reference year 2018, the new EU enterprise concept was introduced in structural business statistics, which led to changes in their processing, publication and presentation. This article describes the results compiled in accordance with the new enterprise definition and compares them with the results of the statistics produced on the basis of the previous enterprise definition (legal units). In addition, results for small and medium-sized enterprises (SMEs) are presented. The change in the enterprise definition leads to quality improvements in SME statistics, because falsely as SMEs considered enterprises are now grouped together into large enterprises. Finally, the article discusses the effects of the change in the enterprise definition on productivity indicators.*

## 1

### Einleitung

Die Einführung der Unternehmensdefinition der Europäischen Union (EU) zählt zu den größten methodischen Veränderungen in der Unternehmensstrukturstatistik der letzten Jahrzehnte. Sie war notwendig geworden, um europäische Vorgaben zu erfüllen. Die zentrale, wirtschaftsbereichsübergreifende Umsetzung hat den Aufbereitungsprozess sowie die Ergebnisermittlung in der Unternehmensstrukturstatistik verändert. Die Erhebungsseite bleibt jedoch unverändert, das heißt erhoben werden die für die Unternehmensstrukturstatistiken erforderlichen Angaben weiterhin bei den rechtlichen Einheiten.

Die Änderungen und Anpassungen, die bei der Erstellung von Ergebnissen für Unternehmen nach EU-Definition notwendig sind, wurden im Einzelnen in mehreren Beiträgen in dieser Zeitschrift ausführlich erläutert:

- › die methodische Einordnung der Einheitendiskussion (Sturm/Redecker, 2016),
- › die notwendige Einführung des Profilings zur Identifizierung der Unternehmen nach EU-Definition (Redecker/Sturm, 2017),
- › die Erläuterung des Aktionsplans und des Grobkonzepts zur Umsetzung in den Unternehmensstrukturstatistiken (Opfermann/Beck, 2018),
- › die Einführung der neuen Arbeitsschritte Imputation und Konsolidierung (Baumgärtner und andere, 2018) sowie kürzlich
- › die Darstellung von Konzept, Umsetzung und Auswirkungen (Beck und andere, 2020a; Beck und andere, 2020b).

Zwischenzeitlich liegen Ergebnisse für das Berichtsjahr 2018 nach der neuen Unternehmensdefinition vor und wurden fristgerecht an das Statistische Amt der Europäischen Union (Eurostat) übermittelt. Der Aktionsplan zur Erfüllung der europäischen Vorgaben wurde damit erfolgreich umgesetzt<sup>1</sup>.

1 Eurostat hat die betroffenen Mitgliedstaaten aufgefordert, Aktionspläne vorzulegen, um den EU-Einheitenbegriff entsprechend der europäischen Einheitenverordnung Nr. 696/93 umzusetzen. Die Leitungen der Statistischen Ämter in Deutschland haben im November 2015 einen solchen Aktionsplan beschlossen. Inhalt und Zeitplan wurden von Eurostat akzeptiert (Opfermann/Beck, 2018).

Der folgende Beitrag stellt wirtschaftsbereichsübergreifend die Ergebnisse der Unternehmensstrukturstatistiken nach der EU-Unternehmensdefinition für das Berichtsjahr 2018 vor. Dabei steht der Begriff Unternehmen verkürzend für Unternehmen nach der EU-Definition. Neben der Darstellung für Unternehmen werden in Kapitel 2 auch ausgewählte Ergebnisse für die bisher geltende Unternehmensdefinition (rechtliche Einheiten) präsentiert. Nutzerinnen und Nutzer erhalten so einen Eindruck, wie sich die Ergebnisse zwischen beiden Unternehmensdefinitionen unterscheiden.

Kapitel 3 schildert, wie sich die Einführung der neuen Unternehmensdefinition auf die Ergebnisse der Statistik für kleine und mittlere Unternehmen auswirkt, Kapitel 4 beschreibt den Einfluss auf ausgewählte Produktivitätskennziffern. Der Artikel gibt in Kapitel 5 einen Überblick über die Veröffentlichungsformate der bereichsübergreifenden Unternehmensstrukturstatistiken und fasst abschließend in Kapitel 6 die Änderungen nach der neuen Unternehmensdefinition kurz zusammen.

## 2

### Unternehmensstrukturstatistiken

Die Einführung der Unternehmensdefinition der EU hat für die Unternehmensstrukturstatistik einige Vorteile. Ein Ziel der Strukturstatistiken ist es, die Wirtschaftsstruktur möglichst realitätsnah abzubilden. So weisen die Strukturergebnisse in den vergangenen Jahren einen Zuwachs der wirtschaftlichen Bedeutung des Dienstleistungsbereichs zu Lasten des Produzierenden Gewerbes aus. Diese Strukturverschiebung war jedoch nur teilweise Ausdruck der ökonomischen Realität. Teilweise resultiert die Verschiebung hin zum Dienstleistungssektor aus (künstlichen) Auslagerungen von Hilfstätigkeiten in eigenständige rechtliche Einheiten, vor allem aus unternehmensinternen Gründen. Diese neuen Einheiten bilden weiterhin mit der auslagernden Stelle eine wirtschaftliche Einheit und sind von ihr abhängig. Die Umstellung der Unternehmensdefinition ermöglicht es, diese künstlichen Verlagerungen angemessener als bisher zu berücksichtigen. Rechtliche Einheiten werden mittels Profiling zu Unternehmen zusammengeführt und unternehmensinterne Transaktionen ohne Marktbezug werden mithilfe der Konsolidierung aus dem statis-



tischen Ergebnis herausgerechnet. Weitere Vorteile, die mit der Umstellung einhergehen, sind die kohärente Darstellung der Strukturstatistiken über alle Wirtschaftsbereiche sowie die europäische Vergleichbarkeit der Strukturstatistiken.

Grundsätzlich gehen in die Ergebnisermittlung für Unternehmen zwei „Unternehmenstypen“ ein. Diese sind zum einen die Unternehmen, die aus genau einer rechtlichen Einheit bestehen (einfache Unternehmen), und zum anderen die komplexen, aus mehreren rechtlichen Einheiten bestehenden Unternehmen. Vorangegangene Beiträge erläuterten diese Unternehmenstypen sowie deren Besonderheiten ausführlich (Beck und andere, 2020a). Wichtige Erkenntnis ist, dass für die komplexen Unternehmen die Arbeitsschritte Imputation und Konsolidierung neu eingeführt wurden (Baumgärtner und andere, 2018). Denn nur durch die Imputation ist es möglich, für die nicht erhobenen rechtlichen Einheiten in komplexen Unternehmen Merkmalswerte zu gewinnen, und nur mittels der Konsolidierung können interne Transaktionen zwischen rechtlichen Einheiten bereinigt werden. Grundlegende Voraussetzung für die Identifikation komplexer Unternehmen sind entsprechende Angaben aus dem Profiling (Redecker/Sturm, 2017). Erhobene einfache Unternehmen (rechtliche Einheiten) werden wie bisher behandelt, das heißt soweit diese Einheiten in Stichprobenerhebungen befragt werden, werden ihre Angaben entsprechend hochgerechnet. Die bisherigen, bereichsspezifischen Hochrechnungen wurden wirtschaftsbereichsübergreifend harmonisiert. Zur Anwendung kommt eine einheitliche freie Hochrechnungsmethodik, die Untererfassungen (etwa durch

rechtliche Einheiten, die den Wirtschaftsbereich wechseln) kompensiert.

Weiterhin ist es für die Analyse der Ergebnisse wichtig zu verstehen, dass für jeden der beiden Unternehmenstypen unterschiedliche Berechnungseffekte auftreten. Vereinfacht gesagt bedeutet dies, dass bei den erhobenen einfachen Unternehmen ein Effekt durch die veränderte Hochrechnung auftritt, bei den komplexen Unternehmen hingegen Imputations-, Profiling- und Konsolidierungseffekte. Die genannten vier Effekte haben Auswirkungen auf das Gesamtergebnis in unterschiedlicher Richtung. Dies ist bedeutsam bei einem Vergleich von Ergebnissen der beiden Unternehmensdefinitionen miteinander. Die einzelnen Effekte wurden mit den Testergebnissen für das Berichtsjahr 2017 bereits ausführlich erläutert und deskriptiv dargestellt (Beck und andere, 2020b).

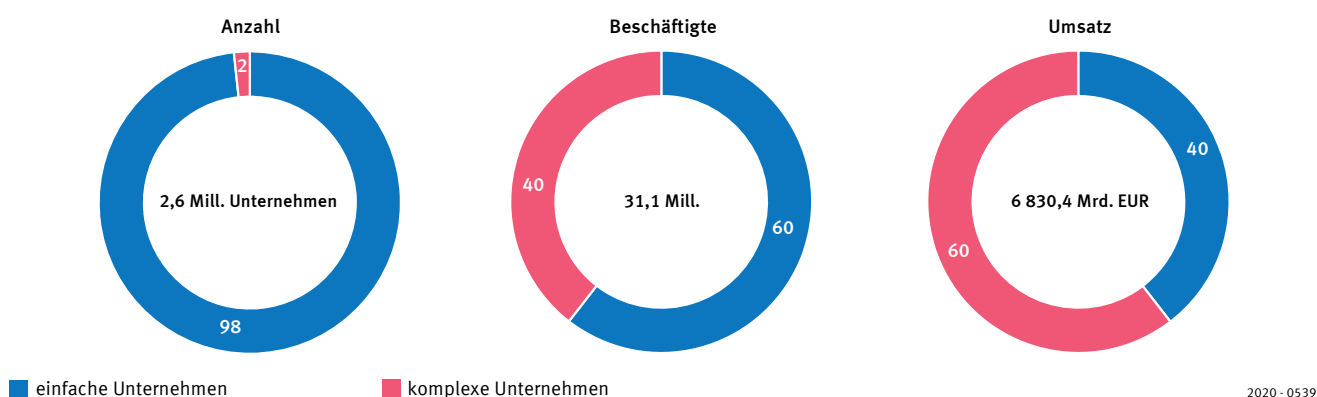
Durch die Einführung des EU-Unternehmensbegriffs kommt es bei der Ergebnisdarstellung ab Berichtsjahr 2018 zu einem Strukturbruch. Im Folgenden werden daher ausgewählte Ergebnisse der beiden Unternehmensdefinitionen gegenübergestellt.

## 2.1 Vergleich der Ergebnisse nach alter und neuer Unternehmensdefinition

Wie groß ist der Anteil komplexer Unternehmen am Gesamtergebnis? Das zeigt [Grafik 1](#) für die Anzahl der Unternehmen, die Beschäftigten und den Umsatz. Bezogen auf die Gesamtzahl der Unternehmen in den

**Grafik 1**

Anzahl, Beschäftigte und Umsatz einfacher und komplexer Unternehmen 2018  
in %



2020 - 0539

# Unternehmensstrukturstatistiken und Statistik für kleine und mittlere Unternehmen nach dem EU-Unternehmensbegriff

für die Unternehmensstrukturstatistiken relevanten Wirtschaftsbereichen<sup>2</sup> betrug der Anteil der komplexen Unternehmen im Berichtsjahr 2018 nur 1,7%. Jedoch generierten diese Unternehmen rund 60% des Gesamtumsatzes und setzten fast 40% aller Beschäftigten ein.

Für das Berichtsjahr 2018 wurden parallel Ergebnisse nach der alten Unternehmensdefinition (rechtliche Einheiten) und nach der neuen Unternehmensdefinition (Unternehmen) veröffentlicht. Bei der Interpretation der Unterschiede zwischen den Ergebnissen von alter und neuer Unternehmensdefinition sind einige wichtige Aspekte zu beachten, die im Folgenden (anhand der vier oben genannten Effekte) genauer erläutert werden.

Die vier Berechnungseffekte (Beck und andere, 2020b) entstehen durch die verschiedenen Arbeitsschritte von der alten Unternehmensdefinition (rechtliche Einheit) hin zum Ergebnis für Unternehmen nach neuer EU-Definition. Je nach Typ des Unternehmens (einfach oder komplex) und Art des Merkmals (additiv oder nicht additiv) wirken sich die vier Effekte aus oder nicht.

#### › Imputationseffekt:

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

<sup>2</sup> Darunter fallen die Wirtschaftszweige B bis N, S95 (ohne K) der Klassifikation der Wirtschaftszweige, Ausgabe 2008 (WZ 2008).

#### › Hochrechnungseffekt:

Dieser Effekt zeigt die Abweichung der Merkmalswerte, 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 in komplexen Unternehmen ändern: Für sich betrachtet bewirkt er eine Verringerung der Werte bei nicht additiven Merkmalen und gleichbleibende Werte bei additiven Merkmalen.

↘ **Tabelle 1** stellt wirtschaftsbereichsübergreifend die Effekte der Einführung der neuen Unternehmensdefinition nach Arbeitsschritten für ausgewählte Merkmale dar. Zusätzlich zu den Ergebnissen nach der alten Unternehmensdefinition (rechtliche Einheiten) in Spalte (1)

**Tabelle 1**

### Gesamtwirtschaftliche Effekte nach Arbeitsschritten für ausgewählte Merkmale 2018

	Rechtliche Einheiten		Ergebnis nach neuer EU-Definition Unternehmen	Imputations- und Hochrechnungseffekt	Profiling- und Konsolidierungseffekt	Gesamteffekt
	Ergebnis nach alter Definition	Zwischenergebnis nach Imputation und Anpassung der Hochrechnung				
	(1)	(2)	(3)	$[(2)-(1)] / (1) \cdot 100 = (4)$	$[(3)-(2)] / (2) \cdot 100 = (5)$	$[(3)-(1)] / (1) \cdot 100 = (6)$
	Anzahl			%		
Einheiten	2 633 965	2 700 562	2 600 925	+ 2,5	- 3,7	- 1,3
Beschäftigte	30 507 940	31 120 840	31 120 840	+ 2,0	± 0,0	+ 2,0
	Mill. EUR			%		
Umsatz	6 951 350	7 071 272	6 830 401	+ 1,7	- 3,4	- 1,7
Waren- und Dienstleistungskäufe insgesamt	5 146 297	5 230 895	4 990 023	+ 1,6	- 4,6	- 3,0
Produktionswert	4 730 129	4 802 271	4 701 556	+ 1,5	- 2,1	- 0,6
Bruttowertschöpfung zu Faktorkosten	1 820 970	1 857 661	1 857 661	+ 2,0	± 0,0	+ 2,0
Bruttobetriebsüberschuss	654 621	679 156	679 156	+ 3,7	± 0,0	+ 3,7
Bruttoinvestitionen in Sachanlagen	283 900	285 361	285 361	+ 0,5	± 0,0	+ 0,5

und Angaben nach der neuen Unternehmensdefinition in Spalte (3) enthält Spalte (2) ein Zwischenergebnis für rechtliche Einheiten nach angepasster Hochrechnung und Imputation. Damit ist es möglich, den Profiling- und Konsolidierungseffekt zu isolieren und zu berechnen, wie in Spalte (5) beschrieben. Darüber hinaus stellt Tabelle 1 den Imputations- und Hochrechnungseffekt in Spalte (4) und den Gesamteffekt in Spalte (6) dar. Der Gesamteffekt ist die prozentuale Veränderung zwischen Ergebnissen nach alter Definition, die Spalte (1) enthält, und den Ergebnissen nach neuer EU-Definition aus Spalte (3).

Auf den ersten Blick unterscheiden sich die absoluten Zahlen zwischen Unternehmen nach alter und neuer Definition nicht allzu sehr. Demnach haben die 31 Millionen Beschäftigten in den 2,6 Millionen Unternehmen der hier betrachteten Wirtschaftsbereiche einen Umsatz von 6,8 Billionen Euro erzielt. Verglichen mit den entsprechenden Werten für das Ergebnis nach alter Definition in Spalte (1) der Tabelle 1 sind diese Angaben recht ähnlich. Im Gesamtbild spiegelt dieses Ergebnis die Erwartungen, die mit der Umstellung des Unternehmensbegriffs einhergingen, nur zum Teil wider. Erwartungsgemäß sollten die Anzahl der Unternehmen sowie die nicht additiven Merkmalswerte (zum Beispiel Umsatz) im Vergleich zu den Ergebnissen nach alter Definition in der gesamtwirtschaftlichen Betrachtung rückläufig sein. Die additiven Merkmalswerte (zum Beispiel Beschäftigte) sollten keine Veränderungen im Vergleich mit den Ergebnissen nach alter Definition aufweisen. Die folgenden Absätze erläutern, wie sich die unterschiedlichen, zum Teil gegenläufigen Effekte auf die Ergebnisse auswirken und die Erwartungen im Gesamteffekt in Spalte (6) der Tabelle 1 teilweise konterkarieren.

Mit der Einführung des Unternehmens nach EU-Definition gingen notwendige Methoden Anpassungen einher, die es erst ermöglichen, die Ergebnisse für Unternehmen in der amtlichen Unternehmensstatistik in Deutschland zu erstellen (Beck und andere, 2020a). Durch die Einführung der bereichsübergreifenden freien Hochrechnung wächst die Anzahl rechtlicher Einheiten und deren Merkmalswerte zunächst, wodurch die bisherige bereichsspezifische Untererfassung, insbesondere kleiner und mittlerer Unternehmen, kompensiert wird.<sup>3</sup> Dies wird im

3 Der Imputations- und Hochrechnungseffekt betrifft kleine und mittlere Unternehmen (KMU) stärker, da große rechtliche Einheiten in den Auswahlansätzen der bereichsspezifischen Stichproben überrepräsentiert sind. Insofern gehen KMU strukturell mit größeren Hochrechnungsfaktoren in die Hochrechnung ein.

Zwischenergebnis nach Imputation und Anpassung der Hochrechnung nachgewiesen, das Spalte (2) in Tabelle 1 zeigt, und schlägt sich im Imputations- und Hochrechnungseffekt nieder, wie aus Spalte (4) in Tabelle 1 hervorgeht.<sup>4</sup>

Änderungen aufgrund der Einführung des neuen Unternehmensbegriffs stellen der Profiling- und der Konsolidierungseffekt dar. Bei additiven Merkmalen, wie der Anzahl der Beschäftigten oder dem Bruttobetriebsüberschuss, ergeben sich wirtschaftsbereichsübergreifend keine Veränderungen ( $\pm 0\%$ ) durch den Konsolidierungseffekt. Jedoch führt der Profilingeffekt bei diesen Merkmalen zu einer Verschiebung der Merkmalswerte zwischen einzelnen Wirtschaftszweigen (siehe den folgenden Abschnitt 2.2). Bei allen nicht additiven Merkmalen wie dem Umsatz reduziert sich das Ergebnis für Unternehmen durch den Konsolidierungseffekt, da hier die internen Transaktionen zwischen den rechtlichen Einheiten komplexer Unternehmen herausgerechnet werden.

Aufgrund des Profilingeffekts sinkt die Anzahl der Unternehmen im Berichtsjahr 2018 um 3,7%. Der Umsatz ist infolge der Konsolidierung der Merkmale rechtlicher Einheiten in komplexen Unternehmen um 3,4% niedriger. Eine Verringerung zeigt sich auch bei den Waren- und Dienstleistungskäufen insgesamt (-4,6%) sowie dem Produktionswert (-2,1%).

Zwar stellt der Profiling- und Konsolidierungseffekt die zu erwartende Änderung aufgrund der Einführung des neuen Unternehmensbegriffs dar. Jedoch unterscheidet sich dieser vom Gesamteffekt in Spalte (6) der Tabelle 1, da der Imputations- und Hochrechnungseffekt dem Profiling- und Konsolidierungseffekt entgegenwirkt. So steigt die Anzahl der Einheiten methodenbedingt insgesamt um 2,5%. Wenn dieser Effekt mit einbezogen wird, verringert sich die Anzahl der Unternehmen im Gesamteffekt nur noch um 1,3%.

Auch bei weiteren nicht additiven Merkmalen relativiert der Imputations- und Hochrechnungseffekt den Konsolidierungseffekt, sodass der Gesamteffekt geringer ausfällt als erwartet. So mindern sich beispielsweise der Umsatz und der Produktionswert im Gesamteffekt nur noch um 1,7 beziehungsweise 0,6%. Durch die Einbe-

4 Da bei der Berechnung der Effekte die Bezugsgrößen variieren, sind die drei Effekte nicht additiv verrechenbar.

ziehung des Imputations- und Hochrechnungseffekts verändern sich auch die additiven Merkmale: Die Anzahl der Beschäftigten steigt um 2,0% und die Investitionen in Sachanlagen erhöhen sich um 0,5%. Bei diesen Merkmalen entspricht der Imputations- und Hochrechnungseffekt dem Gesamteffekt.

## 2.2 Gibt es Strukturverschiebungen zwischen Produzierendem Gewerbe und Dienstleistungen?

Leichte Strukturverschiebungen lässt der Vergleich beider Unternehmensdefinitionen nach Wirtschaftsbereichen erkennen. Demnach haben sich die Anteile des Produzierenden Gewerbes an allen einbezogenen Wirtschaftsbereichen im Vergleich zu den Dienstleistungsbereichen mit der neuen Unternehmensdefinition leicht erhöht. Bezogen auf das Berichtsjahr 2018 stammten nach der neuen Definition 49,3% der Umsätze von Unternehmen des Produzierenden Gewerbes, nach der bisherigen Definition (rechtliche Einheiten) hätte der entsprechende Anteil nur bei 46,7% gelegen. Entsprechend verringerte sich der Anteil der erfassten Dienstleistungsbereiche am Umsatz von 53,3 auf 50,7%.

↘ Tabelle 2

**Tabelle 2**

Wirtschaftsbereiche in den Unternehmensstrukturstatistiken nach Unternehmensdefinition 2018

	Ergebnisse nach alter Definition (rechtliche Einheiten)	Ergebnis nach neuer EU-Definition Unternehmen
	Anteile in %	
Anzahl der Einheiten	100	100
Produzierendes Gewerbe <sup>1</sup>	23,9	24,7
Dienstleistungen <sup>2</sup>	76,1	75,3
Umsatz	100	100
Produzierendes Gewerbe	46,7	49,3
Dienstleistungen	53,3	50,7
Beschäftigte	100	100
Produzierendes Gewerbe	34,6	36,3
Dienstleistungen	65,4	63,7

1 Abschnitte B bis F der Klassifikation der Wirtschaftszweige, Ausgabe 2008 (WZ 2008).  
2 Abschnitte G bis J, L bis N, S95 der WZ 2008.

↘ Tabelle 3 stellt die wirtschaftsbereichsübergreifenden Effekte aus Tabelle 1 gesondert nach Wirtschaftsabschnitten dar, um die Verschiebungen zwischen den einzelnen Wirtschaftsabschnitten zu illustrieren. Im Sinne der Übersichtlichkeit der Darstellung sind in dieser Tabelle nur der Gesamt- und der Profiling- und Konsolidierungseffekt abgebildet.

Bei der Anzahl der Einheiten zeigt sich, dass mit dem Profiling erwartungsgemäß eine Reduzierung der Unterneh-

**Tabelle 3**

Effekte nach Arbeitsschritten und Wirtschaftsabschnitt für ausgewählte Merkmale 2018

	Verarbeitendes Gewerbe (WZ B, C)		Wasser- und Energieversorgung (WZ D, E)		Baugewerbe (WZ F)		Handel (WZ G)		Gastgewerbe (WZ I)		Dienstleistungen (WZ H, J, L, M, N, S95)	
	Gesamteffekt	Profiling- und Konsolidierungseffekt	Gesamteffekt	Profiling- und Konsolidierungseffekt	Gesamteffekt	Profiling- und Konsolidierungseffekt	Gesamteffekt	Profiling- und Konsolidierungseffekt	Gesamteffekt	Profiling- und Konsolidierungseffekt	Gesamteffekt	Profiling- und Konsolidierungseffekt
	%											
Anzahl der Einheiten	+ 14,9	- 4,2	- 8,4	- 8,4	- 1,6	- 1,0	+ 0,7	- 2,6	+ 1,0	- 1,1	- 4,5	- 5,1
Anzahl der Beschäftigten	+ 8,6	+ 3,4	+ 5,1	+ 5,1	+ 2,5	+ 0,9	+ 1,3	+ 0,9	+ 4,9	- 0,1	- 3,0	- 3,4
Umsatz	+ 5,2	+ 3,4	- 1,3	- 1,3	+ 4,7	+ 2,3	- 7,2	- 9,9	+ 4,6	+ 0,5	- 6,3	- 6,6
Waren- und Dienstleistungskäufe insgesamt	+ 4,0	+ 2,6	- 1,8	- 1,8	+ 5,4	+ 2,5	- 8,7	- 11,1	+ 0,4	- 1,5	- 7,0	- 7,7
Produktionswert	+ 2,5	+ 0,7	- 1,7	- 1,7	+ 3,6	+ 1,2	- 1,6	- 4,6	+ 4,2	+ 0,1	- 6,6	- 7,2
Bruttowertschöpfung zu Faktorkosten	+ 8,8	+ 5,9	+ 2,7	+ 2,7	+ 4,0	+ 1,9	+ 1,8	- 2,8	+ 8,9	+ 2,4	- 5,0	- 5,1
Bruttobetriebsüberschuss	+ 13,0	+ 10,4	+ 1,0	+ 1,0	+ 6,2	+ 2,6	+ 8,1	- 6,3	+ 18,6	+ 4,5	- 4,5	- 4,3
Bruttoinvestitionen in Sachanlagen	+ 9,4	+ 7,7	+ 2,8	+ 2,8	+ 15,7	+ 13,5	+ 2,6	+ 7,2	+ 17,1	+ 16,2	- 6,1	- 7,0

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

menszahl in allen Wirtschaftsbereichen einhergeht. Der Imputations- und Hochrechnungseffekt gleicht diesen Effekt jedoch in einzelnen Wirtschaftsabschnitten, etwa im Verarbeitenden Gewerbe, mehr als aus, sodass die Anzahl der Unternehmen im Gesamteffekt zum Teil steigt.

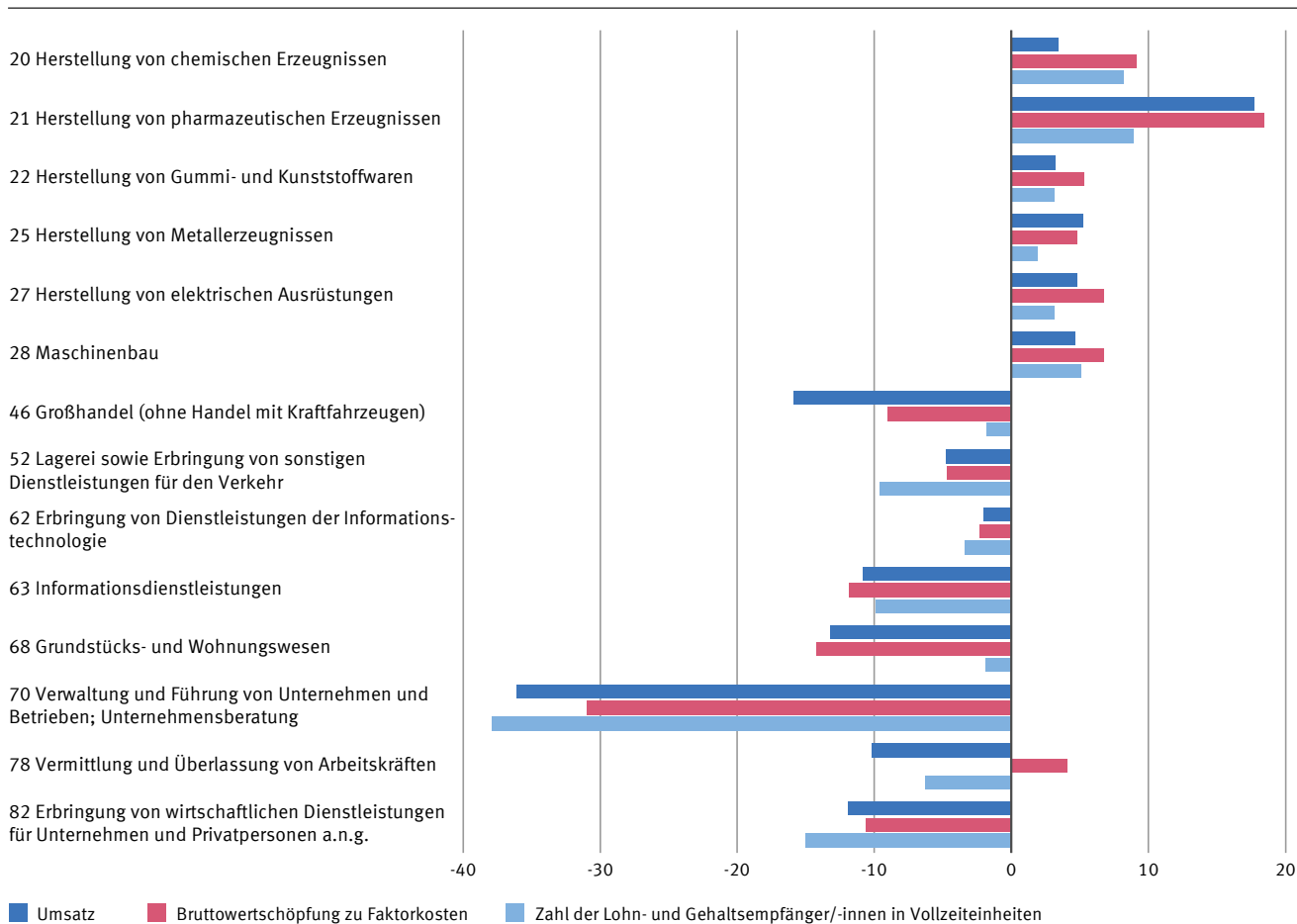
Bei den strukturellen Verschiebungen zwischen den Wirtschaftsbereichen profitieren vor allem das Verarbeitende Gewerbe und das Baugewerbe vom Profiling- und Konsolidierungseffekt. Durch die Zuordnung mehrerer rechtlicher Einheiten zu komplexen Unternehmen und die Konsolidierung der entsprechenden Merkmalswerte erhöhen sich in diesen Bereichen etwa der Umsatz, die Bruttowertschöpfung zu Faktorkosten und der Bruttobetriebsüberschuss. Dagegen verringern sich die Werte im Handels- und Dienstleistungsgewerbe.

Da es bei additiven Merkmalen keinen Konsolidierungseffekt gibt, schlägt sich bei diesen Merkmalen nur der Profilingeffekt nieder. Die Anzahl der Beschäftigten reduziert sich im Dienstleistungsbereich um etwa 380 000. Diese Beschäftigten werden nun überwiegend dem Verarbeitenden Gewerbe oder dem Handel zugeordnet, sodass sich dort die Anzahl der Beschäftigten um 3,4 beziehungsweise 0,9% erhöht.

↳ Grafik 2 veranschaulicht weitere strukturelle Verschiebungen durch die Einführung der neuen Unternehmensdefinition auf Ebene der WZ-Abteilungen (WZ-2-Steller nach der WZ 2008). Einige WZ-Abteilungen sind besonders stark vom Profiling- und Konsolidierungseffekt betroffen. Erwartungsgemäß zeigen sich deutlich negative Effekte bei Umsatz, Bruttowertschöpfung

**Grafik 2**

Profiling- und Konsolidierungseffekt ausgewählter Merkmale nach ausgewählten Wirtschaftszweigen<sup>1</sup> 2018 in %



1 Abteilungen der Klassifikation der Wirtschaftszweige, Ausgabe 2008 (WZ 2008).

sowie der Zahl der Lohn- und Gehaltsempfänger bei der Verwaltung und Führung von Unternehmen (WZ 70), der Übermittlung und Überlassung von Arbeitskräften (WZ 78), der Erbringung von wirtschaftlichen Dienstleistungen (WZ 82), IT-Dienstleistungen (WZ 62 und WZ 63), der Lagerei (WZ 52), im Grundstücks- und Wohnungswesen (WZ 68) und dem Großhandel ohne Handel mit Kraftfahrzeugen (WZ 46). In diesen WZ-Abteilungen ist die künstliche Auslagerung der Hilfstätigkeiten besonders stark ausgeprägt. Demgegenüber profitieren traditionelle Wirtschaftszweige im Verarbeitenden Gewerbe vom Profiling- und Konsolidierungseffekt.

## 3

### Kleine und mittlere Unternehmen

Die Statistik für kleine und mittlere Unternehmen (KMU-Statistik) stellt Informationen über die Struktur und Tätigkeit der Unternehmen in Deutschland nach Größenklassen dar. Dadurch ist es möglich, Aussagen über die Bedeutung und wirtschaftliche Leistungsfähigkeit der kleinen und mittleren Unternehmen zu treffen.

Die KMU-Statistik untergliedert die Ergebnisse der Unternehmensstrukturstatistiken in Anlehnung an die Empfehlung der Europäischen Kommission (2003/361/EG) nach Umsatz- sowie Beschäftigtengrößenklassen.

#### ↳ Übersicht 1

Bis zum Berichtsjahr 2017 wurde die KMU-Statistik nach der alten Unternehmensdefinition (rechtliche Einheit) veröffentlicht. Mit dem Berichtsjahr 2018 wurde auch die KMU-Statistik auf die neue Definition (Unternehmen) umgestellt.

Bei der KMU-Statistik zeigt sich der Vorteil der neuen Unternehmensdefinition besonders deutlich, da mit ihr gleich zwei methodischen Schwächen begegnet werden kann.

#### Übersicht 1

##### Größenklassen der KMU-Statistik

	Beschäftigte		Umsatz	
Kleinstunternehmen	bis 9	und	bis 2 Mill. Euro	
Kleine Unternehmen	bis 49	und	bis 10 Mill. Euro	und kein Kleinstunternehmen
Mittlere Unternehmen	bis 249	und	bis 50 Mill. Euro	und kein kleines Unternehmen
Großunternehmen	250 und mehr	oder	größer 50 Mill. Euro	

Eine berechtigte Kritik an der KMU-Statistik bestand in der Vergangenheit darin, dass auch solche rechtlichen Einheiten als KMU zählten, die Teil einer Unternehmensgruppe sind. Dies widerspricht aber der Empfehlung der Europäischen Kommission (2003/361/EG), wonach „echte“ KMU nicht Teil einer großen Unternehmensgruppe sein sollten.<sup>15</sup>

Die neue Definition trägt dieser Kritik Rechnung. So werden nun Beschäftigte aus allen rechtlichen Einheiten eines komplexen Unternehmens addiert. Darüber hinaus werden die Umsätze der zusammenwirkenden rechtlichen Einheiten konsolidiert und damit auch um interne Transaktionen zwischen den rechtlichen Einheiten bereinigt. Die Zuordnung der Größenklassen bei komplexen Unternehmen orientiert sich nun nicht mehr an den Angaben der rechtlichen Einheiten, sondern an der Gesamtbeschäftigung und dem konsolidierten Gesamtumsatz des komplexen Unternehmens. Insofern muss die neue Unternehmensdefinition die Bedeutung der KMU relativieren, da einige KMU unter Berücksichtigung aller rechtlicher Einheiten innerhalb eines komplexen Unternehmens nun zu den Großunternehmen zählen.

Darüber hinaus kann mit der gesamtwirtschaftlichen Hochrechnung bei der neuen Unternehmensdefinition die Anzahl der KMU realitätsnäher abgebildet werden, da die bisherige Praxis der wirtschaftsbereichsspezifischen Hochrechnung tendenziell eine Untererfassung der KMU aufweist. Dies verdeutlicht der positive Imputations- und Hochrechnungseffekt bei KMU in [↳ Grafik 3](#).

Beim Vergleich der Ergebnisse der KMU-Statistik nach alter Definition und neuer EU-Unternehmensdefinition ist zu berücksichtigen, dass beide Effekte gegenläufige Auswirkungen haben. Während die Umstellung auf

5 Mit der zusätzlichen Berücksichtigung von Gruppenzugehörigkeiten beschäftigte sich zwischen 2015 und 2020 auch eine Arbeitsgruppe der Europäischen Kommission, an der das Statistische Bundesamt beteiligt war (Task Force on Small and Medium Sized Enterprises Data, TF SMED). Nähere Informationen unter: <https://ec.europa.eu>

**Grafik 3**

Effekte nach Arbeitsschritten und KMU-Größenklassen für ausgewählte Merkmale 2018  
in %



die neue Definition (Profiling- und Konsolidierungseffekt) die Bedeutung der KMU relativiert, korrigieren die Methodenverbesserungen (Imputations- und Hochrechnungseffekt) die Anzahl der KMU nach oben. Insofern ist der Gesamteffekt bei der KMU-Statistik – wie auch zuvor für die bereichsübergreifende Unternehmensstatistik erläutert – nicht geeignet, um Aussagen über die Ver-

änderungen durch die Einführung der neuen Unternehmensdefinition abzuleiten.

↘ **Tabelle 4** zeigt die Bedeutung der KMU am Beispiel von drei Merkmalen sowohl absolut als auch relativ. Demnach verringert sich die Anzahl der Einheiten beim Übergang auf die neue Definition bei den kleinen und

# Unternehmensstrukturstatistiken und Statistik für kleine und mittlere Unternehmen nach dem EU-Unternehmensbegriff

**Tabelle 4**

**Ergebnisse der KMU-Statistik nach Arbeitsschritten und ausgewählten Merkmalen 2018**

	Rechtliche Einheiten		Ergebnis nach neuer EU-Definition Unternehmen	Rechtliche Einheiten		Ergebnis nach neuer EU-Definition Unternehmen
	Ergebnis nach alter Definition	Zwischenergebnis nach Imputation und Anpassung der Hochrechnung		Ergebnis nach alter Definition	Zwischenergebnis nach Imputation und Anpassung der Hochrechnung	
	Anzahl			%		
	Einheiten					
KMU	2 614 605	2 681 686	2 585 156	99,3	99,3	99,4
Kleinstunternehmen	2 135 009	2 186 416	2 126 937	81,1	81,0	81,8
Kleine Unternehmen	402 824	417 399	392 298	15,3	15,5	15,1
Mittlere Unternehmen	76 772	77 871	65 921	2,9	2,9	2,5
Großunternehmen	19 360	18 876	15 769	0,7	0,7	0,6
Insgesamt	2 633 965	2 700 562	2 600 925	100	100	100
	Beschäftigte					
KMU	18 461 424	19 303 049	17 814 614	60,5	62,0	57,2
Kleinstunternehmen	5 595 402	5 867 836	5 721 920	18,3	18,9	18,4
Kleine Unternehmen	6 936 891	7 342 412	6 890 040	22,7	23,6	22,1
Mittlere Unternehmen	5 929 130	6 092 801	5 202 654	19,4	19,6	16,7
Großunternehmen	12 046 515	11 817 791	13 306 226	39,5	38,0	42,8
Insgesamt	30 507 940	31 120 840	31 120 840	100	100	100
	Umsatz					
	Mill. EUR			%		
KMU	2 254 411	2 346 557	2 058 510	32,4	33,2	30,1
Kleinstunternehmen	449 735	476 073	453 405	6,5	6,7	6,6
Kleine Unternehmen	800 329	847 022	769 758	11,5	12,0	11,3
Mittlere Unternehmen	1 004 347	1 023 462	835 347	14,4	14,5	12,2
Großunternehmen	4 696 939	4 724 715	4 771 891	67,6	66,8	69,9
Insgesamt	6 951 350	7 071 272	6 830 401	100	100	100

mittleren Unternehmen. Jedoch geht auch die Anzahl der Großunternehmen zurück, was auf den ersten Blick nicht plausibel erscheint. Immerhin gehen einige der rechtlichen Einheiten, die aufgrund ihrer Größe zu den kleineren und mittleren Einheiten zählen, in komplexen Unternehmen auf, die aufgrund ihrer Größe zu den Großunternehmen zählen. Damit sollte die Zahl der großen Einheiten ansteigen. Da sich gleichzeitig der Profilingeffekt negativ auf die Anzahl großer Einheiten auswirkt, kommt es in der Summe zu einer geringeren Anzahl großer Unternehmen nach EU-Definition. Der Profilingeffekt betrifft die großen Unternehmen stärker als die KMU. Daher nimmt die relative Bedeutung der KMU gemessen an der Anzahl der Einheiten sogar um 0,1 Prozentpunkte zu.

Gemessen an der Anzahl der Beschäftigten und dem Umsatz wird die Bedeutung der KMU nach dem Umstieg auf die neue Unternehmensdefinition um mehrere Prozentpunkte relativiert. So repräsentieren KMU nach der

neuen Definition mit 17,8 Millionen nur noch 57,2% der tätigen Personen (zuvor 60,5%). Der Anteil am Umsatz sinkt von 32,4 auf 30,1%.

Grafik 3 verdeutlicht die Gegenläufigkeit der Profiling- und Konsolidierungseffekte gegenüber den Imputations- und Hochrechnungseffekten bei kleinen und mittleren sowie Großunternehmen für ausgewählte Merkmale. Dadurch wird der Gesamteffekt insbesondere bei KMU jeweils deutlich relativiert.

➤ **Tabelle 5** auf Seite 78 zeigt den Profiling- und Konsolidierungseffekt für KMU und Großunternehmen anhand ausgewählter Merkmale nach aggregierten Wirtschaftszweigen. Demnach nimmt die Bedeutung der KMU insgesamt und im Produzierenden Gewerbe, besonders jedoch im Bereich der Dienstleistungen ab.

Im Vergleich dazu vergrößert sich erwartungsgemäß die Bedeutung von Großunternehmen nach dem Übergang



**Tabelle 5**

**Profiling- und Konsolidierungseffekt für KMU und Großunternehmen 2018**

	Produzierendes Gewerbe <sup>1</sup>		Dienstleistungen <sup>2</sup>	
	KMU	Großunternehmen	KMU	Großunternehmen
	%			
Anzahl der Einheiten	- 2,9	- 12,1	- 3,8	- 19,3
Anzahl der Beschäftigten	- 7,1	+ 14,8	- 8,0	+ 11,0
Umsatz	- 11,3	+ 6,4	- 12,8	- 5,2
Waren- und Dienstleistungskäufe insgesamt	- 12,8	+ 4,8	- 13,6	- 7,9
Produktionswert	- 11,5	+ 4,1	- 12,6	+ 1,3
Bruttowertschöpfung zu Faktorkosten	- 9,1	+ 13,8	- 11,3	+ 5,3
Bruttobetriebsüberschuss	- 7,4	+ 17,9	- 8,6	+ 1,6
Bruttoinvestitionen in Sachanlagen	- 9,2	+ 13,8	- 15,5	+ 9,8

In den WZ-Abteilungen D und E (WZ-2-Steller) findet keine Hochrechnung statt. Deshalb entspricht der Gesamteffekt dem Profiling- und Konsolidierungseffekt.

1 Abschnitte B bis F der Klassifikation der Wirtschaftszweige, Ausgabe 2008 (WZ 2008).  
 2 Abschnitte G bis J, L bis N, S95 der WZ 2008.

auf die neue Unternehmensdefinition. Während Großunternehmen im Produzierenden Gewerbe deutliche Zuwächse aufweisen, fallen diese im Bereich der Dienstleistungen geringer aus. Bei den Merkmalen Umsatz und Waren- und Dienstleistungskäufe insgesamt verringert sich die Bedeutung von Großunternehmen im Dienstleistungsbereich sogar deutlich.

## 4

### Einfluss auf Produktivitätskennziffern

Produktivitätskennziffern beschreiben Maße für die Leistungsfähigkeit von unternehmerischem Handeln, indem der Output dem Input gegenübergestellt wird. In der Praxis findet eine Vielzahl von Produktivitätsmaßen Anwendung (für einen Überblick siehe Syverson, 2011). Neben multifaktoriellen Betrachtungsformen, die mehrere Inputfaktoren gleichzeitig einbeziehen<sup>6</sup>, sind auch vergleichsweise einfache Maße weit verbreitet, die dem Output (zum Beispiel Umsatz, Brutto-

wertschöpfung) nur einen Inputfaktor (zum Beispiel Arbeit, Kapital) gegenüberstellen. Im Folgenden werden zwei Maße für die Arbeitsproduktivität betrachtet.

Analog zu den vorab in den Kapiteln 2 und 3 dargestellten Effekten sind auch die Produktivitätsmaße von der Umstellung auf die neue Unternehmensdefinition betroffen. In der bisherigen Ergebnisdarstellung nach alter Unternehmensdefinition konnten weder die Produktions- beziehungsweise Inputfaktoren, noch der Gesamtoutput komplexer Unternehmen konsolidiert dargestellt werden. Daher trägt die neue Definition (Unternehmen) auch bei den Produktivitätskennziffern zu einer realitätsgetreueren Abbildung der Unternehmensstrukturen bei.

➤ **Tabelle 6** stellt den Profiling- und Konsolidierungseffekt für drei Merkmale und zwei daraus abgelei-

**Tabelle 6**

**Profiling- und Konsolidierungseffekt für Produktivitätskennziffern nach Wirtschaftsabschnitten 2018**

	Umsatz	Bruttowertschöpfung zu Faktorkosten	Lohn- und Gehaltsempfänger/-innen in Vollzeiteinheiten	Umsatzproduktivität	Sichtbare Arbeitsproduktivität
	%				
Verarbeitendes Gewerbe <sup>1</sup>	+ 3,4	+ 5,9	+ 3,1	+ 0,2	+ 2,7
Wasser-, Energieversorgung <sup>2</sup>	- 1,3	+ 2,7	+ 5,4	- 6,3	- 2,5
Baugewerbe	+ 2,3	+ 1,9	+ 0,9	+ 1,4	+ 1,0
Handel	- 9,9	- 2,8	+ 1,0	- 10,7	- 3,7
Gastgewerbe	+ 0,5	+ 2,4	- 0,1	+ 0,6	+ 2,5
Dienstleistungen <sup>3</sup>	- 6,6	- 5,1	- 4,0	- 2,7	- 1,1
Insgesamt	- 3,4	± 0,0	± 0,0	- 3,4	± 0,0

1 Abschnitte B und C der Klassifikation der Wirtschaftszweige, Ausgabe 2008 (WZ 2008).

2 Abschnitte D und E der WZ 2008.

3 Abschnitte H, J, L, M, N, S95 der WZ 2008.

6 Für eine Anwendung mit amtlichen Mikrodaten siehe Zimmermann (2020).

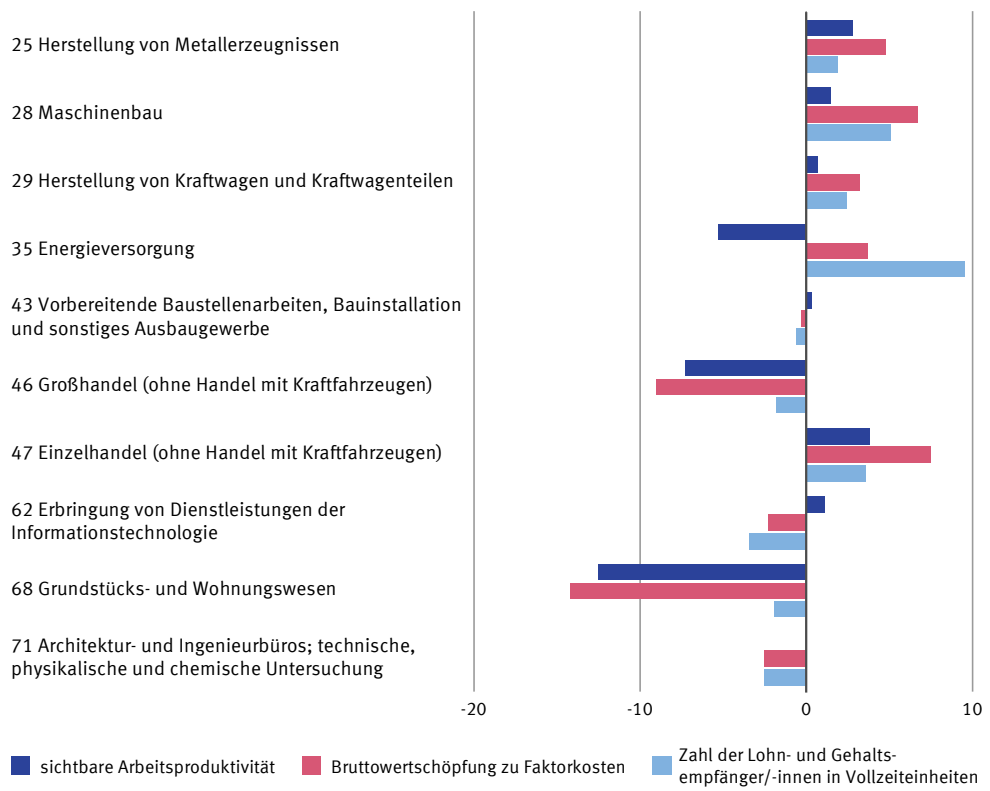
tete Produktivitätsmaße dar. Die Umsatzproduktivität wird als Umsatz geteilt durch die Zahl der Lohn- und Gehaltsempfänger/-innen in Vollzeiteinheiten operationalisiert. Die sichtbare Arbeitsproduktivität wird als Bruttowertschöpfung zu Faktorkosten geteilt durch die Zahl der Lohn- und Gehaltsempfänger/-innen in Vollzeiteinheiten berechnet. Ob sich die Umstellung der neuen Unternehmensdefinition auf die Produktivitätsmaße positiv oder negativ niederschlägt, hängt davon ab, wie sich dies auf die relevanten Merkmale je Wirtschaftszweig auswirkt: Zu einer Verringerung der nachgewiesenen Produktivität kommt es dann, wenn der Profiling- und Konsolidierungseffekt des Inputmerkmals den Profiling- und Konsolidierungseffekt des Outputmerkmals übersteigt.

Da sowohl Bruttowertschöpfung als auch Beschäftigung additive Merkmale sind, verändern sie sich auf gesamt-

wirtschaftlicher Ebene nicht. Somit bleibt gesamtwirtschaftlich die sichtbare Arbeitsproduktivität als Quotient aus beiden ebenfalls konstant. Jedoch zeigen sich strukturelle Verschiebungen nach Wirtschaftszweigen auf disaggregierter Ebene. So nimmt beispielsweise im Verarbeitenden Gewerbe die Bruttowertschöpfung stärker zu (um 5,9 %) als die Beschäftigung (um 3,1 %), sodass die sichtbare Arbeitsproduktivität um 2,7 % steigt. Ein Anstieg der nachgewiesenen sichtbaren Arbeitsproduktivität ist ebenfalls für das Baugewerbe (+1,0 %) und das Gastgewerbe (+2,5 %) zu verzeichnen, während sie sich in den Bereichen Wasser- und Energieversorgung (-2,5 %), Handel (-3,7 %) und Dienstleistungen (-1,1 %) verringert.

Die Stärke des Profiling- und Konsolidierungseffekts variiert auf tiefer disaggregierter Ebene stark. [➔ Grafik 4](#) stellt die Effekte für die sichtbare Arbeitsproduktivität

**Grafik 4**  
**Profiling- und Konsolidierungseffekt ausgewählter Merkmale nach ausgewählten Wirtschaftszweigen<sup>1</sup> 2018**  
 in %



<sup>1</sup> Abteilungen der Klassifikation der Wirtschaftszweige, Ausgabe 2008 (WZ 2008).

und deren Bestandteile für ausgewählte WZ-Abteilungen (WZ-2-Steller nach der WZ 2008) dar.<sup>17</sup> Demnach wird etwa für den Einzelhandel (+ 3,8%) oder die Dienstleistungen in der Informationstechnologie (+ 1,1%) eine höhere Produktivität ermittelt, obwohl für Handel und die Dienstleistungen insgesamt eine niedrigere Produktivität nachgewiesen wird.

## 5

### Nationale Veröffentlichung

Mit der Umstellung auf die neue Unternehmensdefinition stellt das Statistische Bundesamt die Ergebnisse der amtlichen Unternehmensstrukturstatistik für alle Wirtschaftsbereiche und deren gemeinsame Merkmale an einer Stelle in GENESIS-Online<sup>18</sup> zentral zur Verfügung. Die Ergebnisse werden nach der WZ 2008 bis zur Gliederungstiefe der Klassen (WZ-4-Steller) dargestellt. Metadaten liefern die Qualitätsberichte, die ebenfalls unter [www.destatis.de](http://www.destatis.de) bereitstehen. Ergebnisse nach der alten Unternehmensdefinition werden in einer Übergangsphase zumindest für die Berichtsjahre 2018 bis 2020 ergänzend in der gewohnten Form und an gewohnter Stelle veröffentlicht. An Eurostat übermittelt werden jedoch nur Ergebnisse für Unternehmen entsprechend der EU-Verordnung Nr. 696/93.

Die Veröffentlichung der Ergebnisse der KMU-Statistik erfolgt weiterhin in GENESIS-Online im Quader 48121. In der Zeitreihe ergibt sich dabei ein Strukturbruch, da bis zum Berichtsjahr 2017 ausschließlich die rechtliche Einheit die verwendete Unternehmensdefinition war. Ab dem Berichtsjahr 2018 gibt es zusätzlich Ergebnisse für die neue Definition „Unternehmen (EU)“, nach der Ergebnisse veröffentlicht werden. Darüber hinaus stellt die KMU-Statistik für eine Übergangszeit weiterhin Tabellen für die alte Unternehmensdefinition (rechtliche Einheiten) auf den Internetseiten der KMU-Statistik unter [www.destatis.de](http://www.destatis.de) zum Download bereit.

---

7 Ausgewählt wurden WZ-Abteilungen mit einem Anteil von 3% oder mehr an der Gesamtwertschöpfung der Unternehmensstrukturstatistiken bei Unternehmen. Darunter fallen die Wirtschaftszweige B bis N, S95 (ohne K) der WZ 2008.

8 Die Datenbank GENESIS-Online ist über die Homepage des Statistischen Bundesamtes [www.destatis.de](http://www.destatis.de) zu erreichen, die bereichsübergreifende Unternehmensstrukturstatistik über einen neuen Quader mit dem Code 48112.

## 6

### Zusammenfassung

Nach langer Vorbereitung wurden Mitte 2020 erstmals Ergebnisse der Unternehmensstrukturstatistik nach der neuen Unternehmensdefinition erzeugt, an Eurostat übermittelt und anschließend national veröffentlicht. Diese notwendige Änderung in der Unternehmensstatistik konnte belastungsarm vollzogen werden. Wie bei einer großen methodischen Umstellung unvermeidbar, kommt es dabei zu einem Strukturbruch in den Zeitreihen. Um diesen besser einzuordnen und interpretierbar zu machen, ist dieser Beitrag ausführlich auf die Auswirkungen der Umstellung eingegangen. Insbesondere bei einem Vergleich der Ergebnisse zwischen alter und neuer Unternehmensdefinition sind diese Hinweise bedeutsam.

Auch die KMU-Statistik wurde auf die neue Unternehmensdefinition umgestellt und hat damit ihre Aussagekraft verbessert. In der Vergangenheit kam es bei der Darstellung der kleinen und mittleren Unternehmen zu Ungenauigkeiten, da etwa die Unternehmensgruppenzugehörigkeit bisher keine Berücksichtigung fand. In der Folge wurde die Bedeutung der KMU überschätzt. Durch die Konsolidierung innerhalb komplexer Unternehmen werden KMU mit der neuen Unternehmensdefinition nun realitätsgetreuer abgebildet. Bei einem Vergleich der Ergebnisse zwischen alter und neuer Unternehmensdefinition sind jedoch auch hier die Auswirkungen durch die Anpassungen bei der Ergebnisermittlung und die damit verbundenen gegenläufigen Effekte zu beachten.

Mit der Umstellung auf die neue Unternehmensdefinition veröffentlicht das Statistische Bundesamt die Ergebnisse der bereichsübergreifenden Unternehmensstrukturstatistiken nun für alle Wirtschaftsbereiche an einer Stelle in einheitlicher Darstellung. Sie sind damit für Datennutzende einfacher abzurufen. Für einen Übergangszeitraum werden Ergebnisse nach alter Unternehmensdefinition parallel dazu veröffentlicht, und zwar in der gewohnten Form. [uu](#)

## 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. (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)

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.

Syverson, Chad. *What Determines Productivity?* In: Journal of Economic Literature. Jahrgang 49. Ausgabe 2/2011, Seite 326 ff. [Zugriff am 30. September 2020].  
Verfügbar unter: <https://www.aeaweb.org>

Zimmermann, Markus. [Immaterielles Kapital und Produktivität im Verarbeitenden Gewerbe. Ergebnisse auf Basis verknüpfter Einzeldaten \(Micro Data Linking\)](#). In: WISTA Wirtschaft und Statistik. Ausgabe 3/2020, Seite 61 ff.

## RECHTSGRUNDLAGEN

---

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).

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



**Rainer Opfermann**

ist Diplom-Verwaltungswirt und Leiter des Referats „Güterverkehr, Luftverkehr“ des Statistischen Bundesamtes. Zu seinem Arbeitsgebiet gehören die Güterverkehrsstatistik der Binnenschifffahrt und die Seeverkehrsstatistik.

# PERSONENBEFÖRDERUNG UND UNFÄLLE AUF BINNENWASSERSTRASSEN

Rainer Opfermann

↘ **Schlüsselwörter:** Binnenschifffahrt – Fahrgastschifffahrt – Binnenwasserstraße – Schiffsunfälle – Verkehrsstatistik

## ZUSAMMENFASSUNG

Auf europäischer Ebene gibt es einen Informationsbedarf über den Personenverkehr sowie über Schiffsunfälle auf Binnenwasserstraßen, zu dem derzeit keine Daten der deutschen amtlichen Statistik vorliegen. Wie und mit welchem Aufwand solche Informationen gewonnen werden können, hat das Statistische Bundesamt in einer von der Europäischen Kommission finanziell geförderten Pilotstudie untersucht. Der Beitrag beschreibt Methodik und Durchführung dieser Pilotstudie; ihre Ergebnisse fließen in einen Bericht der Europäischen Kommission zu einer möglichen Anpassung der EU-Verordnung über die Statistik des Güterverkehrs auf Binnenwasserstraßen ein.

↘ **Keywords:** inland waterways transport – inland waterway passenger transport – inland waterway – ship accidents – transport statistics

## ABSTRACT

*At European level, there is a need for information on passenger transport and ship accidents on inland waterways. Information of this kind is currently not available from German official statistics. The Federal Statistical Office has conducted a pilot study, with funds from the European Commission, to investigate how such information can be obtained and what effort will be required. This article describes the methodology and implementation of the pilot study. Its results will be incorporated into a report of the European Commission concerning possible amendments to the EU Regulation on statistics of goods transport by inland waterways.*

## 1

---

### Einleitung

---

Die EU-Verordnung von 2018 über die Statistik des Güterverkehrs auf Binnenwasserstraßen legt die Notwendigkeit von Pilotstudien über die Verfügbarkeit statistischer Daten über den Personenverkehr auf Binnenwasserstraßen (einschließlich des grenzüberschreitenden Verkehrs) fest. Sie verpflichtet die Kommission, gemeinsam mit den Mitgliedstaaten die Entwicklung von entsprechenden Statistiken zu untersuchen. Von Interesse sind dabei europäische und nationale Daten zur Anzahl der Passagiere nach Marktsegmenten (Schiffstypen) sowie zu den Fahrgastkilometern. Da für alle anderen Verkehrsmittel europaweit Statistiken über Unfälle vorliegen und lediglich der Verkehrsträger „Binnenschiff“ fehlt, soll auch diese Datenlücke auf europäischer Ebene geschlossen werden.

In der deutschen amtlichen Statistik liegen derzeit für den Verkehrsträger „Binnenschiff“ weder Angaben zu dessen Bedeutung für den Personenverkehr noch zu den Unfällen vor. Wie und mit welchem Aufwand solche Informationen gewonnen werden können, hat das Statistische Bundesamt in einer von der Europäischen Kommission finanziell geförderten Pilotstudie untersucht. Die Pilotstudie umfasst dabei zwei Module:

- › Personenbeförderung auf Binnenwasserstraßen
- › Unfälle auf Binnenwasserstraßen

Eurostat, das Statistische Amt der Europäischen Union (EU), hat zusammen mit sieben EU-Mitgliedstaaten, darunter Deutschland, sowie der Zentralkommission für die Rheinschifffahrt und der Generaldirektion Mobilität und Verkehr der Europäischen Kommission eine Methodik für die vorgenannten Untersuchungsgegenstände entwickelt. Diese diente als Grundlage für die Pilotstudie und ist Thema des zweiten Kapitels.

Der vorliegende Aufsatz stellt Vorgehen und Ergebnisse der Pilotstudie vor. Nicht näher eingegangen wird auf das vorhandene statistische Datenangebot für die Unternehmen des Wirtschaftszweigs „Personenbeförderung in der Binnenschifffahrt“. Hierzu sei allgemein auf die umfangreichen Informationen zur Produktion, Wertschöpfung, zum Anlagevermögen und zur Wettbewerbsfähigkeit verwiesen, die über die [Homepage](#)

[des Statistischen Bundesamtes](#) sowie die [Datenbank GENESIS-Online](#) zugänglich sind.

Kapitel 3 beschreibt die Durchführung der Pilotstudie im Einzelnen, Kapitel 4 befasst sich mit den voraussichtlichen Kosten einer Erhebung zur Personenbeförderung und zu Unfällen auf Binnenwasserstraßen. Der Beitrag schließt mit einem Fazit und Ausblick.

## 2

---

### Methodische Grundlagen der Pilotstudie

---

Eurostat und die EU-Mitgliedstaaten haben in der EU-Task-Force Inland Waterways gemeinsam eine Methodik zur statistischen Erfassung des Personenverkehrs sowie der Unfälle auf Binnenwasserstraßen entwickelt. Mitgearbeitet haben neben Eurostat und dem Statistischen Bundesamt auch Frankreich, Kroatien, die Niederlande, Österreich, Polen, Rumänien sowie die Zentralkommission für die Rheinschifffahrt und die Generaldirektion Mobilität und Verkehr der Europäischen Kommission. Auf Basis dieser Methodik haben dann Deutschland sowie sechs weitere Mitgliedstaaten, von der Kommission unterstützt, Pilotstudien in den Jahren 2019/2020 durchgeführt.

Bei den methodischen Arbeiten stellte sich heraus, dass 15 Mitgliedstaaten bereits heute in unterschiedlichem Umfang Daten zur Passagierschifffahrt auf Binnenwasserstraßen ermitteln. Dabei ist zu berücksichtigen, dass Binnenschifffahrt nur in 22 EU-Mitgliedstaaten stattfindet. Ein Teil der Mitgliedstaaten mit Binnenschifffahrt erfasst bereits Unfälle auf Binnenwasserstraßen. So ermitteln 13 Mitgliedstaaten die Zahl der Unfälle, nicht alle aber beispielsweise die damit verbundenen Personenschäden. Auch die Ansätze zur Gewinnung der Daten unterscheiden sich.

Die Task-Force hat außerdem den bis dato nur allgemein formulierten Informationsbedarf in sogenannten Datenanforderungen (Tabellen) konkretisiert. Für jedes Merkmal und für jede Gliederungsebene hat sie Definitionen erarbeitet.

## 2.1 Datenbedarf zur Passagierschifffahrt auf Binnenwasserstraßen

Der Datenbedarf wurde in vier Tabellen näher benannt:

- › Tabelle A1: Anzahl der beförderten Fahrgäste nach Einschiffungsland, Ausschiffungsland, Beförderungsart und Schiffstyp
- › Tabelle A2: Flusskreuzfahrten – Anzahl der Passagiere, Passagierkilometer und Anzahl der Fahrten nach Einschiffungsregion, Ausschiffungsregion, Nationalität des Schiffs und Art der Kreuzfahrt
- › Tabelle B1: Anzahl der gefahrenen Personenkilometer je Schiffstyp (ohne Flusskreuzfahrtschiffe)
- › Tabelle C1: Anzahl der in einem Land registrierten Schiffe (unabhängig davon, ob sie auf nationalen Binnenwasserstraßen fahren oder nicht) nach Schiffstyp, Schiffskapazität und Schiffsalter

Das europäisch abgestimmte Methodendokument definiert und erläutert umfassend Merkmale und Gliederungen. [↗ Übersicht 1](#) enthält die Festlegungen zu ausgewählten Begriffen. Die sachliche und regionale Aufgliederung der Tabellen kann den in Kapitel 3 dargestellten Spezifikationen entnommen werden.

## 2.2 Datenbedarf zu Unfällen auf Binnenwasserstraßen

Der Datenbedarf wurde in drei Tabellen konkretisiert:

- › Tabelle D1: Anzahl der Unfälle nach Schweregrad, Unfallart und beteiligten Schiffen
- › Tabelle D2: Anzahl der Unfälle nach Schweregrad, Unfallursache und Beteiligung von Gefahrgütern

### Übersicht 1

#### Definition ausgewählter Begriffe zur Passagierschifffahrt auf Binnenwasserstraßen

Anzahl der Fahrgäste:	Als Fahrgast zählt jede Person, die eine „Reise“ an Bord eines Fahrgastschiffs unternimmt. Servicepersonal an Bord der Schiffe zählt nicht dazu.
Fahrgastkilometer:	Maßeinheit für die Beförderung eines Fahrgasts auf Binnenwasserstraßen über eine Entfernung von einem Kilometer. Zu erfassen ist die tatsächlich vom Fahrgast zurückgelegte Entfernung.
Einstiegsland/-region:	Benennt das Einstiegsland des Fahrgasts nach NUTS 0 oder die Hafenregion nach NUTS 2.
Ausstiegsland/-region:	Benennt das Ausstiegsland des Fahrgasts nach NUTS 0 oder die Hafenregion nach NUTS 2.
Fahrgastbinnenschiff:	Schiff, das speziell für die Beförderung von mehr als 12 zahlenden Fahrgästen auf Binnenwasserstraßen konzipiert ist.
Nationalität des Schiffes:	Angegeben wird das Land (Staat), in dem das Schiff registriert ist.
Schiffskapazität:	Die Schiffskapazität beschreibt, wie viele Fahrgäste ein Schiff maximal aufnehmen kann.

- › Tabelle D3: Anzahl der Personen nach Schwere der Verletzung (getötet oder verletzt) und Personengruppen (Passagier, Besatzungsmitglied)

Den Merkmalen „Schwerer Unfall“ und „Gefahrgüter“ liegen folgende Definitionen zugrunde:

- › Schwerer Unfall: Verlust von Menschenleben, Schiffsverlust, schwere Umweltverschmutzung, erhebliche Infrastrukturschäden
- › Gefahrgüter: Stoffe und Gegenstände, deren Beförderung durch die EU-Richtlinie 2008/68 über die Beförderung gefährlicher Güter im Binnenland geregelt ist. Die Gefahrgutklassen sind die in den Empfehlungen der Vereinten Nationen für die Beförderung gefährlicher Güter definierten Klassen.

Die Definitionen und Abgrenzungen dieses Projektteils stützen sich auf die bei der Untersuchung von Seeunfällen (Unfällen) verwendeten Definitionen der Europäischen Agentur für die Sicherheit des Seeverkehrs (EMSA, 2017).

## 3

### Durchführung der Pilotstudie

#### 3.1 Recherche bereits vorhandener Datenquellen

Grundsätzlich können die benötigten Informationen bei den beteiligten Wirtschaftseinheiten selbst durch Primärerhebungen erfasst werden. Belastungsärmer und auch aus Wirtschaftlichkeitsgründen vorzuziehen ist jedoch, bestehende Datenquellen, beispielsweise

Verwaltungsdaten, zu nutzen, in denen die gewünschten Informationen bereits vorliegen. Die Daten müssen jedoch den relevanten Qualitätsanforderungen der amtlichen Statistik (Relevanz, Genauigkeit, Aktualität, Pünktlichkeit, Vergleichbarkeit, Kohärenz) genügen.

Zusammen mit der Wasserstraßen- und Schifffahrtsverwaltung konnten einige Datenquellen identifiziert werden, in denen die benötigten Informationen bereits heute beziehungsweise in der näheren Zukunft vorliegen. Das gilt insbesondere für die Datenanforderung C1 (siehe Abschnitt 2.1) sowie die Datenanforderungen im Modul „Unfälle auf Binnenwasserstraßen“ (siehe Abschnitt 2.2; Näheres dazu folgt in den Abschnitten 3.3 und 3.4).

Ansonsten wurden keine Datenquellen identifiziert, die die benötigten Daten in ausreichender Qualität enthalten. Die Informationen wären also zu erfragen.

### 3.2 Unternehmensbefragung zur Verfügbarkeit von Angaben

Unternehmen, die Personenbeförderung in der Binnenschifffahrt betreiben, wurden dazu befragt, ob die gewünschten Daten bei ihnen verfügbar sind und mit welchem Aufwand sie zu ermitteln wären. Um das Rechercheergebnis abzusichern und um in einem potenziellen Echtbetrieb einer Doppelbelastung der Unternehmen begegnen zu können, wurde zusätzlich erfragt, ob die Informationen bereits an andere Stellen zu melden sind. Die Teilnahme der Unternehmen an der Erhebung nach § 7 Absatz 2 Bundesstatistikgesetz war freiwillig.<sup>1</sup>

Es fehlte somit die Möglichkeit, die Unternehmen zu einer Auskunft zu verpflichten und damit die Vollständigkeit der Antworten und die Qualität der Ergebnisse sicherzustellen. Das war ein wichtiger Grund dafür, keine quantitativen Informationen zu erheben, sondern die Unternehmen ausschließlich zur Verfügbarkeit der gewünschten Daten und zu den Kosten für die Zusammenstellung der Daten zu befragen. Dieses Vorgehen begrenzte zudem die mit der Erhebung verbundene Belastung der Unternehmen. Dahinter steht auch die Erfahrung, dass zwischen dem Aufwand für die Antwort-

ten und der Teilnahmebereitschaft an einer freiwilligen Erhebung ein enger Zusammenhang besteht.

Das Statistische Bundesamt hat für die Befragung einen Onlinefragebogen entwickelt, den alle Unternehmen verwenden konnten, unabhängig davon, in welchen Marktsegmenten sie tätig sind. Filterfragen verbesserten die Übersichtlichkeit des Fragebogens für die heterogene Gesamtheit der angeschriebenen Unternehmen, der aufwendige Einsatz mehrerer spezifischer Fragebogen konnte somit vermieden werden.

#### Berichtskreisaufbau

Es gibt kein Register, das alle Personenverkehr auf Binnenwasserstraßen betreibenden Unternehmen enthält und identifiziert. Der Berichtskreisaufbau für die Pilotstudie erfolgte daher mehrstufig. Zunächst wurden alle aktiven Unternehmen des Berichtsjahrs 2018 aus dem statistischen Unternehmensregister<sup>2</sup> ausgewählt, die schwerpunktmäßig dem Wirtschaftszweig 50.30 „Personenbeförderung in der Binnenschifffahrt“<sup>3</sup> zugeordnet sind. Der Berichtskreis wurde um Unternehmen ergänzt, die selbst nicht diesem Wirtschaftszweig zuzurechnen sind, die aber eine zugehörige örtliche Einheit (Betrieb) mit der entsprechenden Zuordnung zum Wirtschaftszweig 50.30 ausweisen. Im Zuge der Durchführung der Pilotstudie wurden zudem noch einige bedeutende Charterer<sup>4</sup> einbezogen, die Flusskreuzfahrten anbieten.

Damit fehlten insbesondere ausländische Unternehmen, die Personenschifffahrt auf deutschen Binnenwasserstraßen betreiben. Sie sind im deutschen statistischen Unternehmensregister nicht enthalten.

#### Daten gewinnen und Response

Alle so ausgewählten Unternehmen erhielten Schreiben, die die Ziele der Pilotstudie erläuterten und in denen um Teilnahme gebeten wurde sowie die Zugangsdaten zum Online-Fragebogen aufgeführt waren. Von den insgesamt 573 angeschriebenen Unternehmen haben 157

1 § 7 Absatz 2 Bundesstatistikgesetz: „Zur Klärung wissenschaftlich-methodischer Fragestellungen auf dem Gebiet der Statistik dürfen Bundesstatistiken ohne Auskunftspflicht durchgeführt werden.“

2 Das statistische Unternehmensregister führt unter anderem Informationen zu Unternehmen und deren örtlichen Einheiten. Neben Identifizierungsmerkmalen enthält es auch die Zuordnung der statistischen Einheiten zu einem Wirtschaftszweig der WZ 2008, und zwar nach der Haupttätigkeit.

3 Wirtschaftszweig der Klassifikation der Wirtschaftszweige, Ausgabe 2008 (WZ 2008; Statistisches Bundesamt, 2008).

4 Das sind Unternehmen, die ein Schiff vom Eigentümer ausleihen.



den elektronischen Fragebogen ausgefüllt. Für Erhebungen ohne Auskunftspflicht ist die damit erzielte Rücklaufquote von 27,4 % ein guter Wert. Wie bei freiwilligen Befragungen üblich, ist eine tendenziell höhere Antwortbereitschaft bei den größeren Unternehmen zu verzeichnen. Während von den Kleinstunternehmen (unter zehn Beschäftigte) rund ein Viertel den Erhebungsbogen ausgefüllt hat, war es von den Großunternehmen (250 und mehr Beschäftigte) über die Hälfte.

### Ergebnisse der Unternehmensbefragung

Die in den Tabellen spezifizierten Datenanforderungen zielen auf eine nach Marktsegmenten (Schiffstypen) und Regionen gegliederte jährliche statistische Darstellung der Fahrgastschiffahrt. Für alle Schiffstypen soll das Merkmal „Anzahl der Fahrgäste“ ermittelt und danach unterschieden werden, ob ein grenzüberschreitender (internationaler) Verkehr stattgefunden hat oder nicht (Tabelle A1). Die Datenanforderung A2 bezieht sich nur auf das Segment der Flusskreuzfahrten. Gewünscht sind hier nähere Informationen zur Ein- und Ausstiegsregion (Regierungsbezirksebene), eine Unterscheidung nach der Art der Kreuzfahrt (Rundfahrt, keine Rundfahrt) und zusätzlich die Merkmale „Zahl der Fahrten“ und „Fahrgastkilometer“. Für die weiteren Schiffstypen (Datenanforderung B1) sollen die Fahrgastkilometer ermittelt werden.

Es folgen einige spezifische Ergebnisse für die einzelnen Datenanforderungen sowie ein generelles Fazit für den mit der Unternehmensbefragung untersuchten Teil der Pilotstudie.

**Tabelle A1: Anzahl der beförderten Fahrgäste nach Einschiffungsland, Ausschiffungsland, Beförderungsart und Schiffstyp**

### Übersicht 2

Für Tabelle A1 formulierte Datenanforderungen

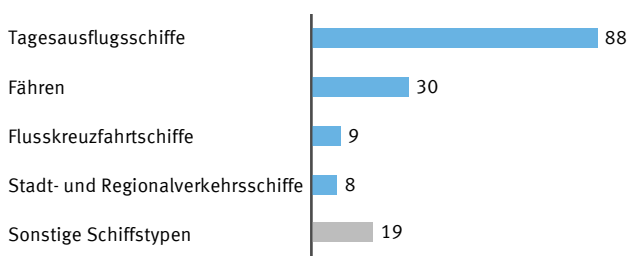
Gliederung/Merkmal	Ausprägung/Einheit
Land der Einschiffung	NUTS 0 oder ISO 3166 <sup>1</sup>
Land der Ausschiffung	NUTS 0 oder ISO 3166
Transportart	1 = national 2 = international
Schiffstyp	1 = Flusskreuzfahrtschiff 2 = Tagesausflugsschiff 3 = Fähre 4 = Stadt- und Regionalverkehrsschiff 5 = Sonstiger Schiffstyp 6 = Unbekannter Schiffstyp
Beförderte Passagiere	Anzahl

<sup>1</sup> Die ISO 3166 ist ein Standard für die Kodierung von geografischen Einheiten, herausgegeben von der Internationalen Organisation für Normung (ISO).

Für die einzelnen Marktsegmente (Schiffstypen) liegen insgesamt 154 Meldungen vor. [➤ Grafik 1](#) Einzelne Unternehmen haben für mehrere Marktsegmente Angaben gemacht.

### Grafik 1

Meldungen nach Marktsegmenten

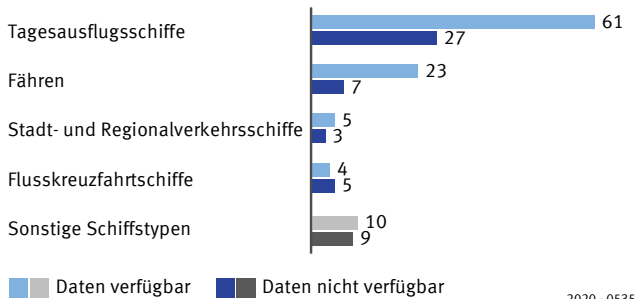


2020 - 0534

Rund drei Viertel der Unternehmen (116 von 154 Fällen) können Angaben zur Transportart sowie zum Einstiegsland und zum Ausstiegsland machen, etwa 67 % (103 Unternehmen) können Informationen zur Anzahl der Fahrgäste bereitstellen. Angaben zur Anzahl der Fahrgäste und zum Einstiegs- und Ausstiegsland (in Kombination) sind rund 55 % (84) der Unternehmen möglich. Damit sieht sich nur etwas mehr als die Hälfte der Unternehmen in der Lage, die gewünschten Angaben zu liefern. [➤ Grafiken 2 bis 4](#)

**Grafik 2**

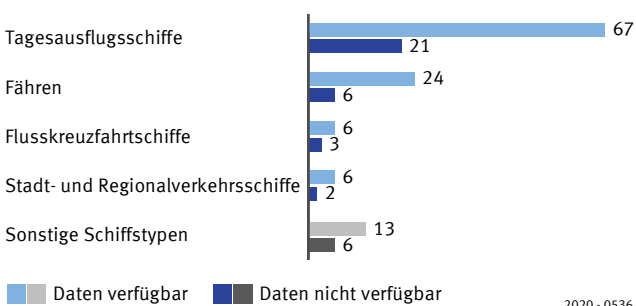
Datenverfügbarkeit zur Anzahl der Fahrgäste nach Marktsegmenten



Für den Bereich der Tagesausflugsschiffe und der Fähren liegen 88 beziehungsweise 30 Rückmeldungen vor. Diese Unternehmen sind überwiegend in der Lage, die gewünschten Angaben zu den Dimensionen Transportart und Schiffstyp zu machen. Allerdings gab knapp ein Drittel (Tagesausflugsschiffe) beziehungsweise ein Viertel (Fähren) der Unternehmen an, über keine Informationen zur Zahl der Fahrgäste zu verfügen. Das trifft auch auf die Segmente „Flusskreuzfahrtschiffe“ sowie „Stadt- und Regionalverkehr“ zu, wobei in diesen Segmenten eine geringere Anzahl an Rückmeldungen vorliegt. Trotzdem bleibt festzuhalten, dass 5 von 9 Unternehmen (Flusskreuzfahrtschiffe) beziehungsweise 3 von 8 Unternehmen (Stadt- und Regionalverkehr) sich nicht in der Lage sehen, Passagierzahlen bereitzustellen.

**Grafik 3**

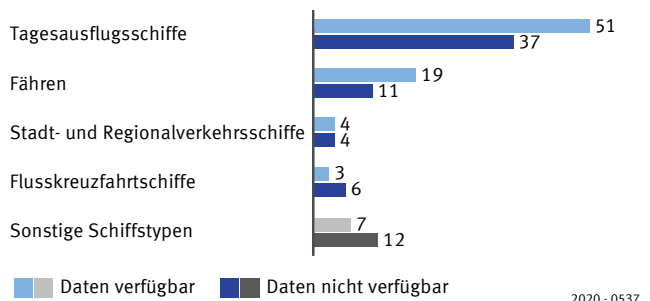
Datenverfügbarkeit zum Ein- und Ausstiegsland nach Marktsegmenten



Einige Ergebnisse verwundern auf den ersten Blick. So war zu erwarten, dass Binnenfähren, die grundsätzlich einen geregelten Schiffsbetrieb zwischen feststehenden Anlegepunkten bedienen, Angaben zum Einstiegsland, Ausstiegsland und zur Transportart liefern können. Auch die hohe Zahl der Fälle, in denen keine Informationen zur

**Grafik 4**

Datenverfügbarkeit zur Kombination nach Ein- und Ausstiegsland und der Anzahl der Fahrgäste nach Marktsegmenten



Anzahl der Fahrgäste vorliegen, überraschte zunächst. Das Statistische Bundesamt hat daraufhin bei einigen Unternehmen rückgefragt mit folgenden Ergebnissen:

- › Bei Fähren, Tagesausflugsschiffen und im Regionalverkehr ist das Land (NUTS 0) des Ein-/Ausstiegs grundsätzlich bekannt, da es hier in der Regel feststehende Anlegestellen gibt. Teilweise haben die Unternehmen die Frage missverstanden und auf die Nationalität des einzelnen Fahrgasts bezogen. Nach den Rückfragen ist davon auszugehen, dass Unternehmen dieser Marktsegmente in der Regel Angaben zum Land machen können. Ebenso kann bei Kreuzfahrtschiffen in der Regel das Land des Ein-/Ausstiegs angegeben werden.
- › Die Ermittlung der Fahrgastzahlen stellt die Unternehmen der Fähr- und der Tagesausflugsschiffahrt sowie die des Regionalverkehrs teilweise vor Probleme. So liegen den Unternehmen bei Gruppentickets keine näheren Informationen zu den tatsächlichen Fahrgastzahlen vor, da diese lediglich die maximale Anzahl begrenzen. Passagiere von Fähren nutzen zudem Zeitkarten (zum Beispiel Monatskarten), die für eine bestimmte Dauer zu einer unbestimmten Zahl von Fahrten berechtigen. Eine Schätzung der Fahrgastzahlen wäre den Unternehmen oft möglich (wenn auch teilweise nur grob).
- › Für Kreuzfahrtschiffe liegen nur wenige Rückmeldungen der Unternehmen vor. Von den fünf Unternehmen, die angaben, dass sie die Fahrgastzahl nicht liefern können, wiesen vier darauf hin, dass die Schiffe an Reiseveranstalter (teilweise ausländische Unternehmen) verchartert wurden und die Angaben vermutlich dort vorlägen. Das Phänomen tritt teilweise auch außerhalb des Marktsegments der Flusskreuzfahrt auf.

**Tabelle A2: Flusskreuzfahrten – Anzahl der Passagiere, Passagierkilometer und Anzahl der Fahrten nach Einschiffungsregion, Ausschiffungsregion, Nationalität des Schiffes und Art der Kreuzfahrt**

Diese Datenanforderung bezieht sich nur auf das Segment der Flusskreuzfahrten. Im Vergleich zur Datenanforderung A1 werden nähere Informationen zur Ein- und Ausstiegsregion verlangt und es wird die Art der Kreuzfahrt unterschieden. Des Weiteren wurde das Merkmal „Zahl der Fahrten“ ergänzt. [↘ Übersicht 3](#)

**Übersicht 3**

Für Tabelle A2 formulierte Datenanforderungen

Gliederung/Merkmal	Ausprägung/Einheit
Region der Einschiffung	NUTS 0 oder ISO
Region der Ausschiffung	NUTS 0 oder ISO
Flagge des Schiffes	NUTS 0 oder ISO
Art der Kreuzfahrt	1 = Rundreise 2 = keine Rundreise
Beförderte Passagiere	Anzahl
Passagierkilometer	in 1 000
Fahrten	Anzahl

Zur Datenanforderung A2 liegen neun Rückmeldungen von Unternehmen vor, wobei es drei Unternehmen nicht möglich ist, Rundfahrten von sonstigen Kreuzfahrten zu unterscheiden sowie Angaben zur Einstiegs- und zur Ausstiegsregion zu machen. Unterscheidet man nach der Art der Kreuzfahrt, so sind bei Rundreisen drei von sechs Unternehmen nicht in der Lage, die entfernteste Region entweder nach NUTS 2, nach Angabe eines Hafens oder Ortes oder einer sonstigen räumlichen Angabe zu bestimmen. In fünf von neun Fällen verfügen die Kreuzfahrtunternehmen nicht über die notwendigen Informationen zur Zahl der Fahrgäste. Die Informationen zur Anzahl der Fahrten liegen überwiegend (in acht von neun Fällen), die zur Nationalität des Schiffes (Schiffsflagge) vollständig vor.

Bei Rückfragen gaben die Unternehmen an, dass ihnen im Falle der Schiffscharter die gewünschten Informationen nicht vorliegen und verwiesen auf die Charterer.

Das Merkmal Fahrgastkilometer soll nach dem europäischen Konzept nicht unbedingt erhoben, sondern anhand der Informationen zur Region des Ein- beziehungsweise Ausschiffens mithilfe einer Distanzmatrix und den sonstigen Informationen zur Passagierzahl berechnet werden. Dazu wäre aber notwendig, dass die

Unternehmen vollständige Angaben zur Anzahl der Passagiere und der Einschiffungs- beziehungsweise Ausschiffungsregion machen könnten.

**Tabelle B1: Anzahl der gefahrenen Personenkilometer pro Schiffstyp (ohne Flusskreuzfahrtschiffe)**

Diese Datenanforderung bezieht alle Marktsegmente mit Ausnahme der Flusskreuzfahrten ein. Einzige Variable sind die Fahrgastkilometer. [↘ Übersicht 4](#)

**Übersicht 4**

Für Tabelle B1 formulierte Datenanforderungen

Gliederung/Merkmal	Ausprägung/Einheit
Schiffstyp	2 = Tagesausflugsschiff 3 = Fähre 4 = Stadt- und Regionalverkehrsschiff 5 = Sonstiger Schiffstyp
Passagierkilometer	in 1 000

Um den Unternehmen die Ermittlung zu erleichtern, wurde auf europäischer Ebene eine Schätzmethode entwickelt. Anhand der Anzahl der Fahrten im Jahr, der durchschnittlichen prozentualen Fahrgastauslastung und der durchschnittlichen Länge je Fahrt sollen die Fahrgastkilometer von den Unternehmen näherungsweise ermittelt werden.

**↘ Beispiel:**

Angenommen sei: Ein Unternehmen meldet 4 Fahrten je Tag für eine bestimmte Strecke und fährt auf dieser Strecke durchschnittlich 270 Tage je Jahr (4 Fahrten x 270 Tage = 1 080 Fahrten im Jahr).

Die durchschnittliche Fahrgastkapazität je Fahrt beträgt 150 Fahrgäste und der Auslastungsfaktor 70 % (150 Fahrgastkapazität x 70 % = 105 Fahrgäste je Fahrt). Mit diesen Annahmen ergeben sich 113 400 Fahrgäste im Jahr (1 080 Fahrten x 105 Fahrgäste je Fahrt). Beträgt die durchschnittlich zurückgelegte Entfernung je Fahrt 5 Kilometer, so errechnen sich 567 000 Fahrgastkilometer im Jahr (113 400 Fahrgäste x 5 Kilometer je Route = 567 000 Fahrgastkilometer je Jahr).

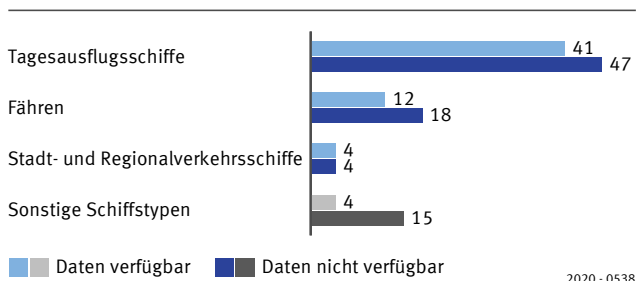
Im Zuge der Pilotstudie wurden die Unternehmen befragt, ob ihnen die notwendigen Informationen zur Berechnung der Fahrgastkilometer vorliegen oder ob sie diese ermitteln können.

Nur gut zwei Fünftel (42 %) der Unternehmen können nach eigenen Angaben die für die Berechnung notwendigen Angaben komplett bereitstellen. Knapp drei Fünf-

tel der Unternehmen (58%) können dies nicht. Besonders hoch ist dieser Anteil im Bereich der sonstigen Schiffstypen (79%). Bei den anderen Marktsegmenten liegen die entsprechenden Anteile zwischen 50 und 60%. Damit ist bei jedem Schiffstyp nicht einmal die Hälfte der Unternehmen in der Lage, die notwendigen Informationen zur Berechnung der Fahrgastkilometer zu liefern. [↪ Grafik 5](#)

### Grafik 5

Datenverfügbarkeit zur Berechnung der Fahrgastkilometer nach Marktsegmenten



Auch hier hat das Statistische Bundesamt die Unternehmen kontaktiert und um nähere Erläuterungen gebeten.

- › Rückfragen bei Fähren ergaben, dass die Länge der Strecke grundsätzlich bekannt ist, die für die Anwendung des Schätzmodells notwendigen Informationen zur Anzahl der Fahrten und zum Fahrgastaufkommen aber nicht. Die Unternehmen haben darauf hingewiesen, dass die Anzahl der Fahrten teilweise vom Fahrgastaufkommen abhängt und daher nicht immer bekannt ist. Auch Angaben zur Fahrgastzahl sind nicht immer vorhanden (siehe oben).
- › Bei Tagesausflügen gibt es das Problem, dass die Fahrgäste bei Linienfahrten teilweise die Route unterbrechen und mit dem nächsten Schiff weiterfahren können. Allerdings gibt es hier einen Start- und Zielpunkt, weil das Ticket für eine bestimmte Strecke gekauft wird. Dieser „gekaufte“ Reiseverlauf könnte als Näherungswert erfasst werden.
- › Im Regionalverkehr stellt sich verstärkt das Problem der Monats- und Sammelkarten (siehe oben).
- › Die Gliederung „Sonstige Schiffstypen“ umfasst sehr verschiedene Angebote, beispielsweise auch Charterfahrten mit historischen Schiffen. Für diese Angebote gibt es zum Teil überhaupt keine Aufzeichnungen in der benötigten Form.

### Fazit

Für die im Bereich der Passagierschifffahrt interessierenden Merkmale und Dimensionen stehen keine (administrativen) Quellen zur Verfügung, mit denen die notwendigen Angaben gewonnen werden können und somit eine Erhebung verzichtbar würde. Daher wäre eine Befragung der Unternehmen notwendig.

Die Pilotstudie hat gezeigt, dass ein Teil der zu befragenden Unternehmen nicht über alle Informationen in der gewünschten Gliederung verfügt. Die Probleme bei der Bereitstellung der Daten betreffen insbesondere drei Punkte:

1. Im Falle der Schiffscharter stehen den Eignern der Schiffe nicht alle Informationen zur Verfügung.
2. Bei spezifischen Fahrkartenangeboten fehlen Informationen zur Zahl der Fahrgäste, zum Teil auch zu den Routen.
3. Es fehlen Informationen von ausländischen Unternehmen, die Fahrgastschifffahrt im Inland anbieten.

Für diese Probleme sind auf europäischer Ebene Lösungsansätze zu entwickeln. Um den Meldeaufwand insbesondere hinsichtlich der Erfassung der Fahrgastkilometer zu begrenzen und der eingeschränkten Datenverfügbarkeit zu begegnen, käme zum Beispiel ein vereinfachter methodischer Ansatz infrage. Dieser müsste auf die gesamte Schiffsreise abstellen und keine Zu- und Ausstiege von Fahrgästen während der Fahrt berücksichtigen.<sup>5</sup> Damit einher geht natürlich ein Verlust an detaillierter Information.

Eine weitere Möglichkeit zur Begrenzung von Berichtspflichten ist, kleinere Unternehmen von der Erfassung auszunehmen. Der Pilotstudie lag als untere Erfassungsgrenze eine Schiffskapazität von zwölf Fahrgästen zugrunde. Knapp 4% der in Deutschland registrierten Schiffe haben eine Fahrgastkapazität von 12 bis maximal 20 Fahrgästen, weitere rund 17% eine von 21 bis 50 Fahrgästen. Bei Fähren liegen die entsprechenden Anteile erheblich höher (gut 9 beziehungsweise 37%). Gleiches gilt grundsätzlich für Personenbarkassen. Auch wenn damit keine Aussage zur Größe der dahinterste-

<sup>5</sup> Der vereinfachte methodische Ansatz geht beispielsweise bei einer Rundreise davon aus, dass alle Passagiere die Standardroute zurücklegen, also die volle Strecke. Er berücksichtigt nicht, dass nicht alle Passagiere die gesamte Strecke zurücklegen, sondern unterwegs beziehungsweise aussteigen.

henden Unternehmen getroffen wird, wäre zu bedenken, diese kleinen Schiffe von der Erhebung zu befreien.

Besonderes Augenmerk ist auf die Erfassung ausländischer Unternehmen zu legen, die im Inland Passagierdienstleistungen anbieten. Für die vollständige Darstellung der inländischen Passagierschifffahrt ist deren Einbeziehung relevant, wirft aber Erhebungsprobleme auf. Sollen die von ausländischen Unternehmen erbrachten Verkehrsleistungen einbezogen werden, sind auf europäischer Ebene Lösungsansätze zu deren Erfassung zu entwickeln.<sup>16</sup>

### 3.3 Datenbedarf zum Schiffsbestand

Hinsichtlich der Datenanforderung zum Schiffsbestand (Anzahl der in einem Land registrierten Schiffe nach Schiffstyp, Schiffskapazität und Schiffsalter; [↪ Übersicht 5](#)) hat das Statistische Bundesamt Kontakt mit der Wasserstraßen- und Schifffahrtsverwaltung des Bundes aufgenommen. Diese führt die Zentrale Binnenschiffsbestandsdatei, die unter anderem technisches Hilfsmittel für die Fortschreibung der deutschen Binnenflotte ist.

#### Übersicht 5

##### Datenanforderungen zum Schiffsbestand

Gliederung/Merkmal	Ausprägung/Einheit
Schiffstyp	1 = Flusskreuzfahrtschiff 2 = Tagesausflugsschiff 3 = Fähre 4 = Stadt- und Regionalverkehrsschiff 5 = Sonstiger Schiffstyp
Passagierkapazität	1 = 12 – 20 Passagiere 2 = 21 – 50 Passagiere 3 = 51 – 200 Passagiere 4 = > 200 Passagiere
Schiffsalter	1 = 0 bis 5 Jahre 2 = 6 bis 10 Jahre 3 = 11 bis 20 Jahre 4 = 21 bis 30 Jahre 5 = älter als 30 Jahre 6 = unbekannt
Schiffe	Anzahl

Binnenschiffe, die in Deutschland registriert werden sollen, werden zunächst in Binnenschiffsregister eingetragen, welche von den Amtsgerichten geführt werden.

<sup>6</sup> Zum Beispiel durch den Aufbau eines europaweiten Registers von Unternehmen, die Personenbeförderung (auch) in anderen Mitgliedstaaten anbieten, und auf das die Mitgliedstaaten bei der Statistikerstellung zugreifen könnten.

Diese Registereinträge sind die Basis für die Zentrale Binnenschiffsbestandsdatei. Die Wasserstraßen- und Schifffahrtsverwaltung ergänzt weitere Eintragungen, die sie beim Eigentümer des Schiffs erfragt.

Genau geregelt ist, welche Schiffe in den Binnenschiffsregistern zu erfassen sind. So können Personenschiffe eingetragen werden, deren Wasserverdrängung bei größter Eintauchung mindestens fünf Kubikmeter beträgt. Ab einer Wasserverdrängung von zehn Kubikmetern ist die Eintragung verpflichtend. Schiffe im Eigentum der öffentlichen Hand müssen nicht zur Eintragung angemeldet werden.

Die in der Zentralen Binnenschiffsbestandsdatei registrierten Schiffe können inländischen oder ausländischen Eignern gehören. Enthalten sind auch Schiffe, die zwar in Deutschland registriert sind, aber nicht auf deutschen Binnenwasserstraßen fahren. Umgekehrt sind nicht alle auf deutschen Binnenwasserstraßen fahrenden Binnenschiffe enthalten, nämlich dann, wenn sie im Ausland registriert sind. Auf den Nachweis der im Inland verkehrenden Schiffe stellt dieser Teil der Datenanforderungen aber auch nicht ab.

Die Zentrale Binnenschiffsbestandsdatei ist damit geeignet, den spezifischen Datenbedarf abzudecken, eine Unternehmensbefragung ist nicht notwendig.

Die Wasserstraßen- und Schifffahrtsverwaltung wertet die Veränderung des Schiffsbestands der deutschen Binnenflotte regelmäßig statistisch aus und publiziert sie.<sup>17</sup> So hat sich im Fortschreibungszeitraum 1. Januar 2018 bis 31. Dezember 2018 der Schiffsbestand der in der Zentralen Binnenschiffsbestandsdatei erfassten fortschreibungspflichtigen Binnenflotte auf 4 500 Binnenschiffe erhöht. Zum Jahresende 2018 gab es in der Personenschifffahrt 1 005 Fahrgastschiffe, 184 Personenbarkassen und 267 Fähren. Einen Teil des Datenangebots zeigt [↪ Tabelle 1](#).

<sup>7</sup> [www.gdws.wsv.bund.de](http://www.gdws.wsv.bund.de)

**Tabelle 1**

Schiffsbestand in der Personenschifffahrt am 31. Dezember 2018

	Schiffe je Schiffsgattung				
	zusammen	Flusskreuzfahrt	Tagesausflug	Fähren	Stadt- und Regionalverkehr
Alter des Schiffs					
0 bis 5 Jahre	54	10	29	15	0
6 bis 10 Jahre	61	16	39	6	0
11 bis 20 Jahre	173	18	111	42	2
21 bis 30 Jahre	222	8	156	49	9
älter als 30 Jahre	942	8	607	155	172
Alter unbekannt	4	0	3	0	1
Insgesamt	1 456	60	945	267	184

### 3.4 Unfälle auf Binnenwasserstraßen

Schiffsunfälle auf Binnenwasserstraßen werden in Deutschland von den Wasserschutzpolizeien der Länder erfasst. Es gibt aber aktuell keine zentrale Datenbank, in der die von den Wasserschutzpolizeien ermittelten Informationen digitalisiert, strukturiert und harmonisiert erfasst wären. Gesetzlich beschlossen ist der Aufbau einer bundesweiten Schiffsunfalldatenbank; sie soll detaillierte und relevante Daten zu den Schiffsunfällen enthalten und nach aktueller Planung bis Ende 2021 fertiggestellt sein.<sup>18</sup> Mit dem Aufbau ist die Generaldirektion Wasserstraßen und Schifffahrt beauftragt.

Die zentrale Schiffsunfalldatenbank HAVARIS wird nach ihrer Inbetriebnahme die geeignete Quelle zur Erfüllung der in den Tabellen D1 bis D3 formulierten Datenanforderungen sein. ➔ **Übersichten 6 bis 8** Sie wird detaillierte Informationen enthalten über

- › Besatzungsmitglieder, Bordpersonal, Lotsen, Be- und Entlader der an einem Unfall beteiligten Wasserfahrzeuge,
- › die Position und Fahrtrichtung der an einem Unfall beteiligten Wasserfahrzeuge,
- › den Unfall (Unfallzeitpunkt, Unfallort, Unfallart, Verkehrssituation, äußere Bedingungen, insbesondere Wetterlage, Unfallursachen, Unfallfolgen) und
- › die an einem Unfall beteiligten Wasserfahrzeuge.

<sup>18</sup> Rechtsgrundlage ist das Schiffsunfalldatenbankgesetz von 2013.

### Übersicht 6

Für Tabelle D1 formulierte Datenanforderungen

Gliederung/Merkmal	Ausprägung/Einheit
Schweregrad des Unfalls	1 = Sehr schwerer Unfall 2 = Schwerer Unfall
Art des Unfalls	1 = Kollision mit einem anderen Wasserfahrzeug 2 = Kollision mit einem festen Objekt 3 = Auf Grund gelaufen 4 = Feuer oder Explosion 5 = Wassereintritt 6 = Gekentert 7 = Sonstiger Unfall 8 = Unbekannt
Beteiligte Schiffe	1 = Frachtschiff 2 = Passagierschiff 3 = Frachtschiffe 4 = Passagierschiffe 5 = Fracht- und Passagierschiffe 6 = Sonstige
Unfälle	Anzahl

Dies ermöglicht, das Unfallgeschehen auf Bundeswasserstraßen vollständig und kohärent darzustellen. Die Bundeswasserstraßen sind abzugrenzen von den Binnenwasserstraßen der Länder, die ebenfalls eine allgemeine, meist aber nachgeordnete Verkehrsfunktion

### Übersicht 7

Für Tabelle D2 formulierte Datenanforderungen

Gliederung/Merkmal	Ausprägung/Einheit
Schweregrad des Unfalls	1 = Sehr schwerer Unfall 2 = Schwerer Unfall
Unfallgrund	1 = Menschlicher Fehler 2 = Technisches Problem 3 = Wetter- oder Wasserbedingungen 4 = Sonstiger Grund 5 = Unbekannt
Gefahrgut	1 = Gefahrgut transportiert 2 = Gefahrgut ausgetreten 3 = Unbekannt 4 = Kein Gefahrgut transportiert
Unfälle	Anzahl

### Übersicht 8

Für Tabelle D3 formulierte Datenanforderungen

Gliederung/Merkmal	Ausprägung/Einheit
Schweregrad der Verletzung	1 = Personen getötet 2 = Personen verletzt
Personentyp	1 = Passagier 2 = Crew-Mitglied 3 = Sonstige Person 4 = Unbekannt
Personen	Anzahl

erfüllen können. Als Lücke verbleiben damit aus aktueller Sicht Unfälle auf Landeswasserstraßen, weil sie nicht im Anwendungsbereich von HAVARIS enthalten sind.

## 4

### Kosten

In der Pilotstudie hat das Statistische Bundesamt versucht, den Aufwand für die Unternehmen und die Verwaltung (statistisches Amt und datenzuliefernde Verwaltung) grob abzuschätzen. Das geschah für die Abschätzung der Unternehmensbelastung auf Basis von erfragten Informationen zum benötigten Zeitaufwand und zur Qualifikation der Person, die den Erhebungsbogen in einem potenziellen Echtbetrieb ausfüllen würde.

Danach ergäben sich bei einer jährlichen Erhebung für die Unternehmen grob geschätzt Kosten in Höhe von rund 100 000 Euro pro Jahr. Die tatsächlichen Kosten hängen dabei stark von der verwendeten Methodik und der Fähigkeit der Unternehmen ab, die Daten entsprechend zu melden. Die grob geschätzten Kosten stehen damit unter einer schwer einzuschätzenden Unsicherheit.

Für die Verwaltung ergäben sich grob geschätzt insgesamt einmalige Kosten in Höhe von rund 740 000 Euro und laufende jährliche Kosten in Höhe von rund 155 000 Euro.

## 5

### Fazit und Ausblick

Die Pilotstudie führte für die verschiedenen Datenanforderungen zu unterschiedlichen Ergebnissen. Zum Teil sind bereits in der Verwaltung vorliegende Daten verwendbar beziehungsweise stehen in absehbarer Zeit als Datenquelle bereit; in diesen Fällen ist eine Unternehmensbefragung durch die statistischen Ämter verzichtbar. Zu einem anderen Teil sind Erhebungen notwendig, wobei einige der geforderten Angaben den Unternehmen nicht vorliegen.

Für die im Bereich der Passagierschifffahrt interessierenden Merkmale (insbesondere Anzahl der Fahrgäste, Fahrgastkilometer) und Dimensionen (zum Beispiel Ein- und Ausstiegsregion) stehen keine (administrativen) Quellen zur Verfügung. Die Informationen müssten daher über eine Befragung der Unternehmen gewonnen werden.

Einige Unternehmen haben in der Befragung zur Datenverfügbarkeit darauf hingewiesen, dass ihnen im Falle der Schiffscharter nicht alle Informationen vorliegen. Das trifft im Falle spezifischer Fahrkartenangebote auch auf die Informationen zur Zahl der Fahrgäste und zum Teil zu den Routen zu. Fahrgastzählungen sind unserer Kenntnis nach im Bereich der Personenbeförderung auf Binnenwasserstraßen nicht in größerem Umfang etabliert. Drei Fünftel der Unternehmen verfügen nicht über die notwendigen Informationen für die Berechnung der Fahrgastkilometer (Anzahl der Fahrten im Jahr, die durchschnittliche Auslastung in Prozent und die Länge der Route).


Die benötigten Angaben liegen den Unternehmen zum Teil also nicht vor. Daher müssten die Unternehmen mit Schätzungen und Annahmen arbeiten, was sich auf die Qualität der Angaben auswirken wird. Die hinter der Ermittlung der Informationen stehenden Annahmen wären von den Unternehmen regelmäßig zu überprüfen und anzupassen. Das bedeutet Aufwand für die Unternehmen und es ist fraglich, ob dies regelmäßig geschehen würde. Daher ist in den statistischen Gremien und mit den Nutzenden zu diskutieren, ob für die Deckung des Datenbedarfs – zumindest teilweise (Datenanforderungen A2, B1) – eine Erfassung in mehrjährigen Abständen ausreicht. Zur Diskussion steht auch die notwendige sachliche und regionale Gliederungstiefe der

Ergebnisse, ebenso die Spezifikation einer Methodik, die den Unternehmen die Meldungen erleichtert. Eine vereinfachte Methodik zur Begrenzung der Unternehmensbelastung ist gegenüber Nutzeranforderungen an feineren Daten abzuwägen. Der Datenbedarf ist umso stärker zu rechtfertigen, je größer die Meldeprobleme der Unternehmen sind. Die europäischen statistischen Gremien sind der geeignete Rahmen, die notwendigen Diskussionen zu führen.

Etwaige Datenanforderungen müssen ferner die Problematik der von ausländischen Unternehmen erbrachten Verkehrsleistungen im Inland berücksichtigen. Dazu gehört die Festlegung einer Methodik (Erfassung aller oder ausschließlich inländischer Unternehmen) und gegebenenfalls der Aufbau eines europäischen Registers von Unternehmen, die Passagierbeförderung in den EU-Mitgliedstaaten betreiben.

Für die formulierten Datenanforderungen zum Schiffsbestand (siehe Abschnitt 3.3) stehen mit der Zentralen Binnenschiffsbestandsdatei Verwaltungsdaten zur Verfügung, der die Informationen in der benötigten Tiefe entnommen werden können.

Um die Datenanforderungen zu den Schiffsunfällen (siehe Abschnitt 3.4) zu erfüllen, wird die im Aufbau befindliche zentrale Schiffsunfalldatenbank HAVARIS nach ihrer Inbetriebnahme die geeignete Quelle sein.

Die Europäische Kommission wird dem Europäischen Parlament und dem Rat bis Dezember 2020 in einem Bericht die Ergebnisse der Pilotstudien über die Durchführbarkeit einer neuen Datenerhebung, ihre Kosten für die Mitgliedstaaten und ihre statistische Qualität darlegen. Abhängig von diesem Bericht wird die Kommission gegebenenfalls einen Vorschlag zur Anpassung der geltenden Verordnung (EU) 2018/974 in Bezug auf die statistische Erfassung des Personenverkehrs und der Unfälle auf Binnenwasserstraßen vorlegen. 



## LITERATURVERZEICHNIS

---

EMSA – European Maritime Safety Agency. *Annual Overview of Marine Casualties and Incidents 2017*. [Zugriff am 21. Oktober 2020]. Verfügbar unter: [emsa.europa.eu](https://www.emsa.europa.eu)

Statistisches Bundesamt. *Klassifikation der Wirtschaftszweige, Ausgabe 2008*. Verfügbar unter: [www.destatis.de](https://www.destatis.de)

## 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 6 des Gesetzes vom 10. Juli 2020 (BGBl. I Seite 1648) geändert worden ist.

Richtlinie 2008/68/EG des Europäischen Parlaments und des Rates vom 24. September 2008 über die Beförderung gefährlicher Güter im Binnenland (Amtsblatt der EU Nr. L 260, Seite 13).

Schiffsunfalldatenbankgesetz (SchUnfDatG) vom 7. August 2013 (BGBl. I Seite 3118), das zuletzt durch Artikel 150 des Gesetzes vom 20. November 2019 (BGBl. I Seite 1626) geändert worden ist.

Verordnung (EG) Nr. 1059/2003 des Europäischen Parlaments und des Rates vom 26. Mai 2003 über die Schaffung einer gemeinsamen Klassifikation der Gebietseinheiten für die Statistik (NUTS) (Amtsblatt der EU Nr. L 154, Seite 1).

Verordnung (EU) 2018/974 des Europäischen Parlaments und des Rates vom 4. Juli 2018 über die Statistik des Güterverkehrs auf Binnenwasserstraßen (Amtsblatt der EU Nr. L 179, Seite 14).

**Herausgeber**

Statistisches Bundesamt (Destatis), Wiesbaden

---

**Schriftleitung**

Dr. Daniel Vorgrimler

Redaktionsleitung: Juliane Gude

Redaktion: Ellen Römer

---

**Ihr Kontakt zu uns**

[www.destatis.de/kontakt](http://www.destatis.de/kontakt)

---

**Erscheinungsfolge**

zweimonatlich, erschienen im Dezember 2020

Das Archiv älterer Ausgaben finden Sie unter [www.destatis.de](http://www.destatis.de)

---

Artikelnummer: 1010200-20006-4, ISSN 1619-2907

---

© Statistisches Bundesamt (Destatis), 2020

Vervielfältigung und Verbreitung, auch auszugsweise, mit Quellenangabe gestattet.