
WISTA

Wirtschaft und Statistik

Arne Ackermann | Xaver Dickopf |
Tanja Mucha

Andrea Buschner | Eva Kibele |
Ulrike Winkelmann | Olaf Eckert

Albert Braakmann

Florian Dumpert

Peter Kuntze | Benedikt Kuckelkorn

Jannik Schaller

Eva-Maria Diehl-Wolf

Florian Peter

Frauke Mischler

Sabrina Estatico | Tobias Tornow |
Sabrina Walther

Flash und Nowcast: Schnellschätzungen des Bruttoinlandsprodukts in der Corona-Pandemie

Neue Monatsberichte der Todesursachenstatistik — mit Fokus auf dem Nachweis von COVID-19-Sterbefallzahlen

Digitalisierung, Globalisierung sowie Wohlfahrt und Nachhaltigkeit – Kernthemen der aktuellen Überarbeitung der internationalen VGR-Methodik

Machine Learning in der amtlichen Statistik – Ergebnisse und Bewertung eines internationalen Projekts

Multifaktorproduktivität in den Volkswirtschaftlichen Gesamtrechnungen

Datenfusion von EU-SILC und Household Budget Survey – ein Vergleich zweier Fusionsmethoden

Neukonzeption der Bodenmarktstatistiken

Neue Statistik der Strom- und Erdgasdurchschnittspreise

Verdienstunterschiede zwischen Männern und Frauen

Generalprobe für den Zensus 2022: Erkenntnisse aus der Pilotstudie zur Haushaltebefragung

4 | 2021

ABKÜRZUNGEN

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

ZEICHENERKLÄRUNG

–	nichts vorhanden
0	weniger als die Hälfte von 1 in der letzten besetzten Stelle, jedoch mehr als nichts
.	Zahlenwert unbekannt oder geheim zu halten
. . .	Angabe fällt später an
X	Tabellenfach gesperrt, weil Aussage nicht sinnvoll
I oder —	grundsätzliche Änderung innerhalb einer Reihe, die den zeitlichen Vergleich beeinträchtigt
/	keine Angaben, da Zahlenwert nicht sicher genug
()	Aussagewert eingeschränkt, da der Zahlenwert statistisch relativ unsicher ist
	Abweichungen in den Summen ergeben sich durch Runden der Zahlen.
	Tiefer gehende Internet-Verlinkungen sind in der Online-Ausgabe hinterlegt.

INHALT

3	Editorial
4	Kennzahlen — wichtige Indikatoren zur Coronakrise
8	Informationsangebote zur Coronakrise
10	Kurznachrichten
17	Arne Ackermann, Xaver Dickopf, Tanja Mucha Flash und Nowcast: Schnellschätzungen des Bruttoinlandsprodukts in der Corona-Pandemie <i>Flash and nowcast: early estimates of the gross domestic product during the coronavirus pandemic</i>
29	Andrea Buschner, Eva Kibele, Ulrike Winkelmann, Olaf Eckert Neue Monatsberichte der Todesursachenstatistik — mit Fokus auf dem Nachweis von COVID-19-Sterbefallzahlen <i>New monthly reports based on the causes of death statistics, focusing on COVID-19 death figures</i>
42	Albert Braakmann Digitalisierung, Globalisierung sowie Wohlfahrt und Nachhaltigkeit – Kernthemen der aktuellen Überarbeitung der internationalen VGR-Methodik <i>Digitalisation, globalisation as well as well-being and sustainability – core topics of the current revision of the international national accounting methodology</i>

INHALT

53	Florian Dumpert Machine Learning in der amtlichen Statistik – Ergebnisse und Bewertung eines internationalen Projekts <i>Machine learning in official statistics - Results and evaluation of an international project</i>
64	Peter Kuntze, Benedikt Kuckelkorn Multifaktorproduktivität in den Volkswirtschaftlichen Gesamtrechnungen <i>Multifactor productivity in national accounts</i>
76	Jannik Schaller Datenfusion von EU-SILC und Household Budget Survey – ein Vergleich zweier Fusionsmethoden <i>Fusion of data from EU-SILC and the household budget survey – A comparison of two fusion methods</i>
87	Eva-Maria Diehl-Wolf Neukonzeption der Bodenmarktstatistiken <i>Redesigned land market statistics</i>
96	Florian Peter Neue Statistik der Strom- und Erdgasdurchschnittspreise <i>New statistics on average electricity and natural gas prices</i>
110	Frauke Mischler Verdienstunterschiede zwischen Männern und Frauen <i>The gender pay gap</i>
126	Sabrina Estatico, Tobias Tornow, Sabrina Walther Generalprobe für den Zensus 2022: Erkenntnisse aus der Pilotstudie zur Haushaltebefragung <i>Trial run for the 2022 Census: findings from the household survey pilot study</i>

EDITORIAL

Dr. Georg Thiel



LIEBE LESERIN, LIEBER LESER,

die Themenausgabe „[Amtliche Statistik in Zeiten von Corona, Teil 2](#)“ hat verdeutlicht, dass die amtliche Statistik auf die Herausforderungen im Zusammenhang mit der Coronakrise reagiert und sich kontinuierlich weiterentwickelt. Der erste Beitrag in diesem Heft vergleicht die Revisionsanalysen der Schnellschätzung des Bruttoinlandsprodukts 30 Tage nach Quartalsende (BIP-t+30-Flash) mit denen eines modellgestützten Ansatzes nach 10 Tagen (BIP-t+10+Nowcast) vor und während der Corona-Pandemie. Es zeigt sich, dass die Kombination von Expertenschätzungen mit ökonometrischen Verfahren dem rein modellgestützten Ansatz in Krisenzeiten überlegen ist. Genaue und zeitnahe Informationen über diejenigen Sterbefälle, die im Zusammenhang mit COVID-19 stehen, ermöglichen die neuen Monatsberichte der Todesursachenstatistik mit vorläufigen Daten. Diese gemeinsame Entwicklung der Statistischen Ämter der Länder und des Bundes erlaubt einen ersten Einblick in das Mortalitätsgeschehen.

Die Masterarbeit „Datenfusion von EU-SILC und HBS: Vergleich zwischen Random Hot-Deck und Predictive Mean Matching im Rahmen einer Simulationsstudie“ wurde nicht nur mit dem Gerhard-Fürst-Preis 2020 des Statistischen Bundesamtes ausgezeichnet, sondern auch mit dem EMOS Master Award 2021. Um den sozialen und wirtschaftlichen Lebensstandard in der Europäischen Union zu beurteilen, strebt die amtliche Statistik an, die Einkommen und Konsumausgaben privater Haushalte auf europäischer Ebene mithilfe einer einheitlichen Datengrundlage gemeinsam zu betrachten.

Die deutsche amtliche Statistik ist auf internationaler Ebene gut vernetzt. So begleitet beispielsweise ein Experte des Statistischen Bundesamtes aktiv die anstehende Überarbeitung des weltweiten Systems Volkswirtschaftlicher Gesamtrechnungen mit den Kernthemen Digitalisierung, Globalisierung sowie Wohlfahrt und Nachhaltigkeit. Auch die weiteren Beiträge dieser Ausgabe empfehle ich Ihrer Lektüre. Bleiben Sie gesund!

A handwritten signature in black ink, reading "Georg Thiel". The script is fluid and cursive.

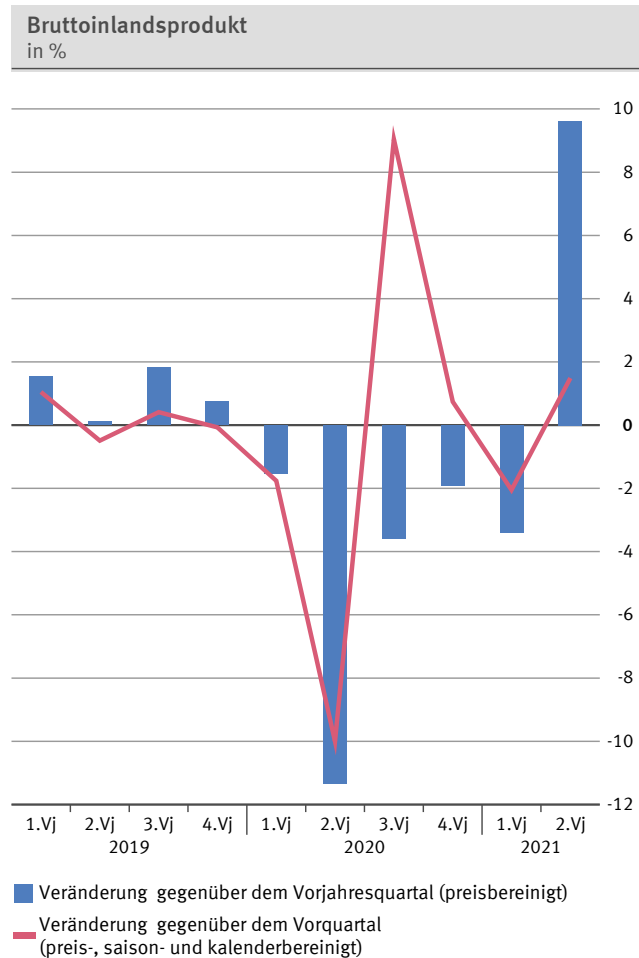
Präsident des Statistischen Bundesamtes

Auswirkungen der Corona-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.

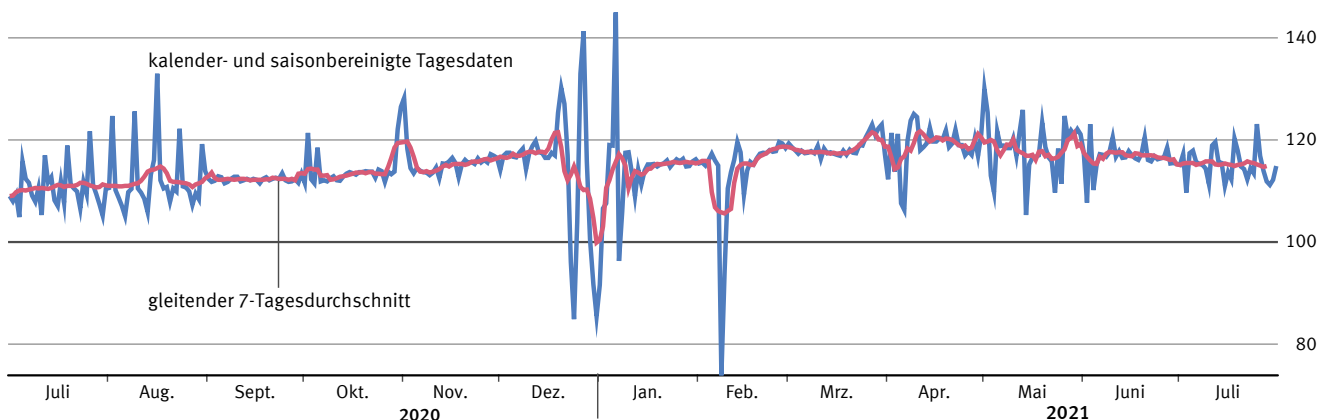
Die Kennzahlen auf dieser und den folgenden Seiten zeigen die Entwicklung einer Auswahl wichtiger Indikatoren, bei denen sich Auswirkungen zeigen.

Am 30. Juli 2021 hat das Statistische Bundesamt erneut das Bruttoinlandsprodukt (BIP) bereits 30 Tage nach Quartalsende veröffentlicht. Der [Podcast BIP t+30](#) 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 Juli 2020

2015 = 100

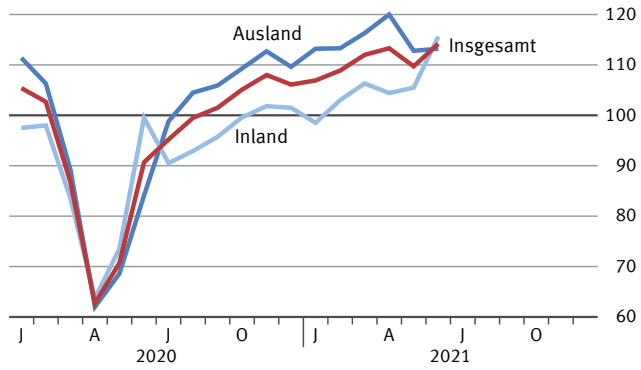


Quellen: Bundesamt für Güterverkehr, Deutsche Bundesbank, Statistisches Bundesamt

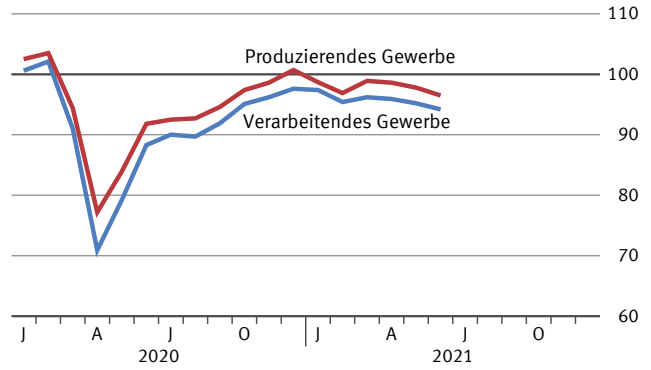
Stand: 11.08.2021

Kennzahlen – wichtige Indikatoren zur Coronakrise

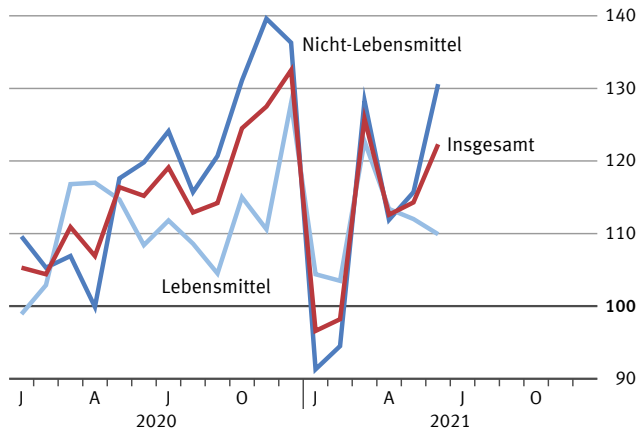
Auftragseingang im Verarbeitenden Gewerbe
Volumenindex 2015 = 100



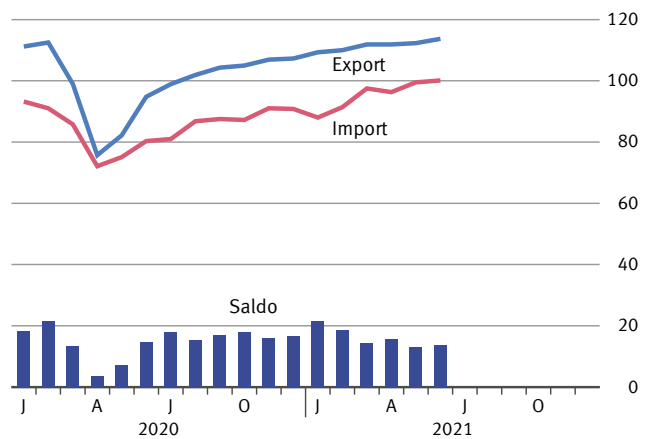
Produktion im Produzierenden und Verarbeitenden Gewerbe
Index 2015 = 100



Umsatz im Einzelhandel
2015 = 100



Außenhandel
in Mrd. EUR



Stand: 11.08.2021

Kennzahlen – wichtige Indikatoren zur Coronakrise

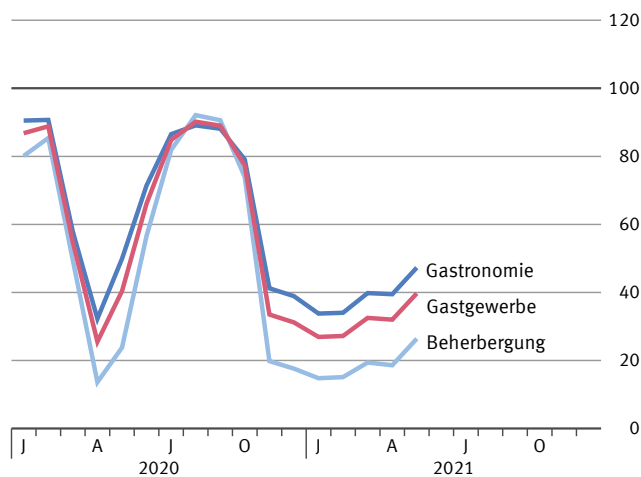
Verbraucherpreisindex 2015 = 100

2020		2021	
Januar	105,2	Januar	106,3
Februar	105,6	Februar	107,0
März	105,7	März	107,5
April	106,1	April	108,2
Mai	106,0	Mai	108,7
Juni	106,6	Juni	109,1
Juli	106,1	Juli	
August	106,0		
September	105,8		
Oktober	105,9		
November	105,0		
Dezember	105,5		

110,1 **↑ 3,8 %**
Veränderung
zum Vorjahresmonat

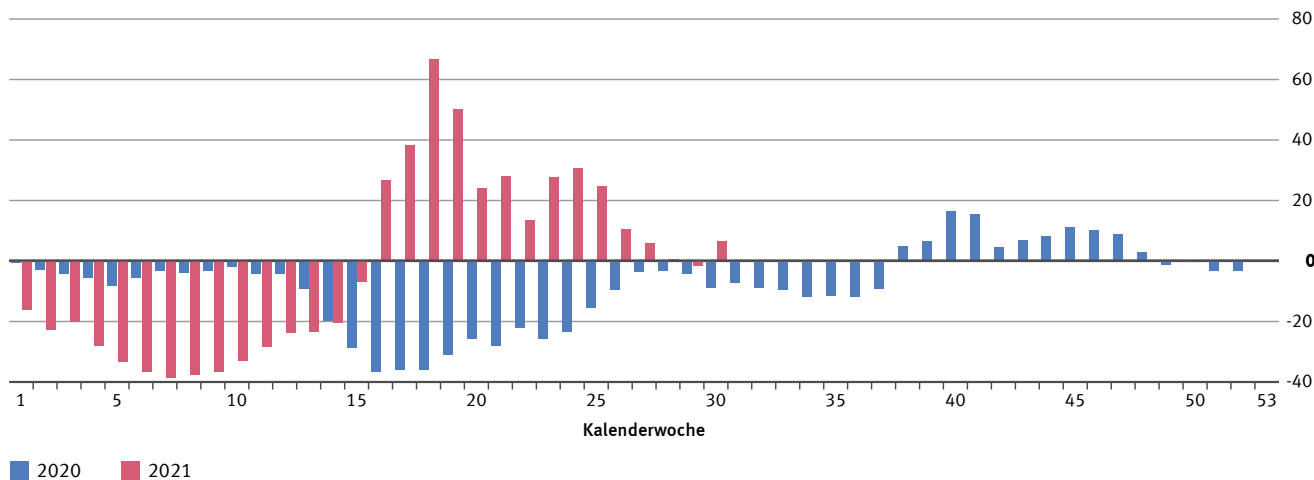
Umsatz des Gastgewerbes

in konstanten Preisen (real), Originalwert, 2015 = 100



Neue Kreditverträge nach Kalenderwochen

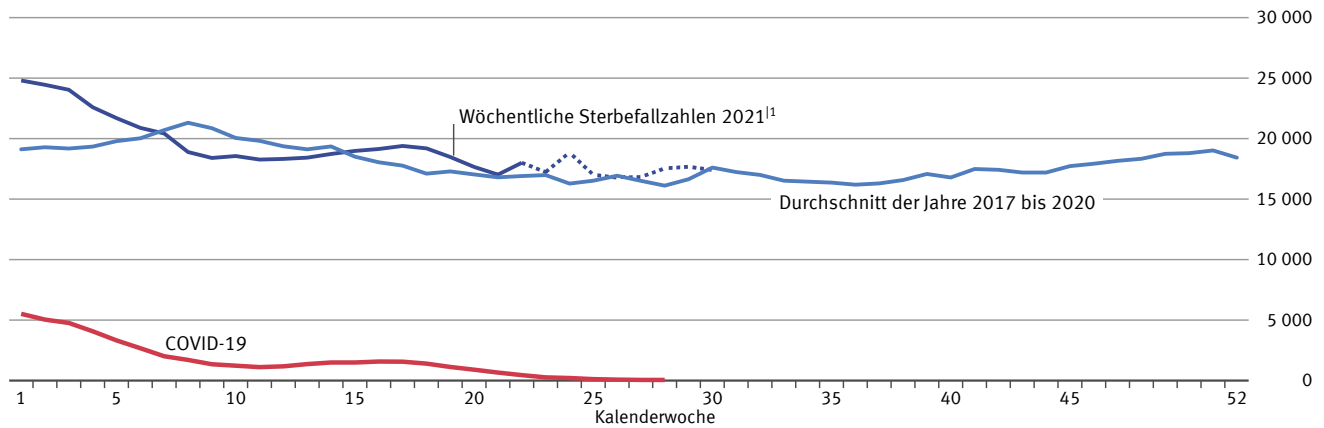
Veränderung gegenüber der entsprechenden Vorjahreswoche in %



Stand: 11.08.2021

Kennzahlen – wichtige Indikatoren zur Coronakrise

Wöchentliche Sterbefallzahlen in Deutschland 2021

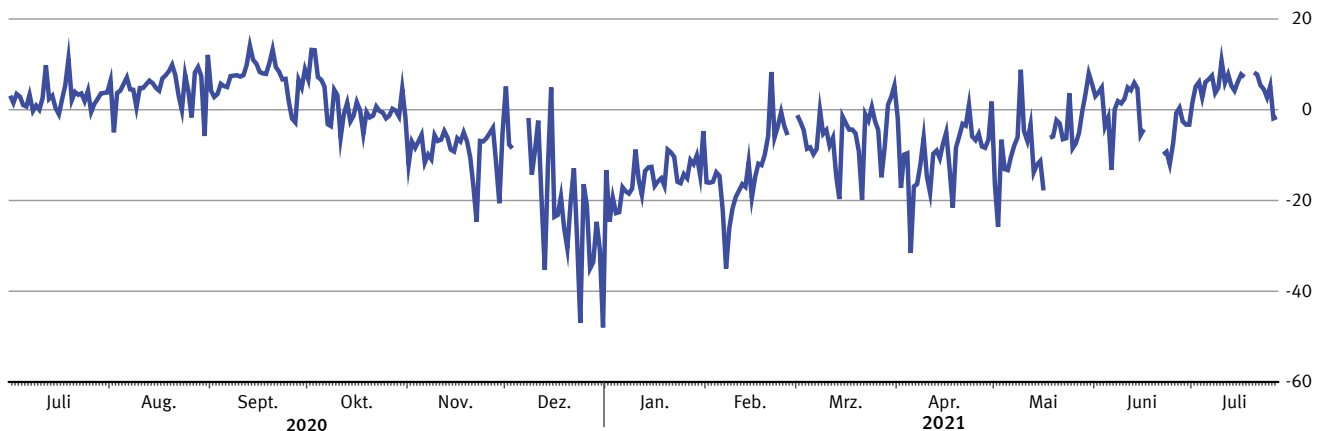


Gestrichelte Werte enthalten Schätzanteil.

1 Sonderauswertung der vorläufigen Sterbefallzahlen.

Quellen: Statistisches Bundesamt (Sterbefallzahlen insgesamt), Robert Koch-Institut (COVID-19-Todesfälle)

Veränderung der Mobilität gegenüber 2019 in %



Datenlücken entstehen in der Regel aufgrund von technischen Problemen beim Mobilfunkanbieter.

Experimentelle Daten; Quelle: Teralytics

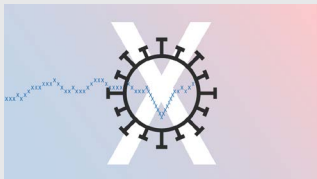
Stand: 11.08.2021



Corona-Statistiken – Auswirkungen der COVID-19-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 seit Beginn der Coronakrise bereitgestellt hat, um den Bedarf an Zahlen am aktuellen Rand zu decken:

➤ www.destatis.de/corona



Corona-Daten Deutschland

Mit Beginn der SARS-CoV-2-Pandemie ist die Relevanz aktueller Daten und Statistiken noch einmal deutlich gestiegen. Vor diesem Hintergrund hat das Statistische Bundesamt gezielt sein Informationsangebot um aktuelle und regional tief gegliederte Daten zum Infektionsgeschehen, zur Krankenhausituation, zur Mobilität und zu den ergriffenen Eindämmungsmaßnahmen der Regierungen erweitert. Die im Rahmen eines im Jahr 2020 vom Bundesministerium für Wirtschaft und Energie initiierten Projekts aufgebaute Datensammlung ist nun für die gesamte Öffentlichkeit frei zugänglich.

➤ www.corona-daten-deutschland.de



EXDAT – Experimentelle Daten

In der 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



Datenportal: Dashboard Deutschland

Das vom Statistischen Bundesamt neu entwickelte Datenportal präsentiert hochaktuelle und hochfrequente Daten aus verschiedenen Themenbereichen, wie Gesundheit, Mobilität und Wirtschaft, auf anschauliche Art und Weise. Es trägt damit zu einem faktenbasierten demokratischen Diskurs der Öffentlichkeit und zur evidenzbasierten Entscheidungsfindung durch Politik und Verwaltung bei. In der aktuellen Lage dient das Dashboard Deutschland auch als Gradmesser für die Coronakrise mit aktuellen Zahlen zu gewährten Corona-Hilfen für Unternehmen, täglich aktualisierten Mobilitätsindikatoren und den wichtigsten Informationen des Robert Koch-Instituts zur aktuellen Pandemiesituation.

➤ www.dashboard-deutschland.de



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



European Statistical Recovery Dashboard

Das während der deutschen EU-Ratspräsidentschaft im zweiten Halbjahr 2020 gemeinsam entwickelte europäische Dashboard bildet die wirtschaftlichen und sozialen Entwicklungen infolge der Corona-Pandemie in der Europäischen Union insgesamt und ihren einzelnen Mitgliedstaaten ab. Mehr als 20 Indikatoren ermöglichen Vergleiche zwischen den EU-Mitgliedstaaten. Die interaktive Anwendung erlaubt eine personalisierte Darstellung sowie das Herunterladen und Weiterverarbeiten der statistischen Daten.

➤ ec.europa.eu/eurostat

KURZNACHRICHTEN

IN EIGENER SACHE

Neues Datenangebot „Corona-Daten Deutschland“

Von der 7-Tage-Inzidenz bis zur Arbeitsmarktentwicklung – das neue Informationsangebot „Corona-Daten Deutschland“ des Statistischen Bundesamtes bietet aktuelle, hochfrequente und regional tief gegliederte Daten aus amtlichen und nichtamtlichen Quellen, die im Zusammenhang mit der Corona-Pandemie relevant sind. Zentral dazu gehören die ergriffenen Maßnahmen zur Eindämmung der Corona-Pandemie seit März 2020 auf Bundesland- und Kreisebene. Als bislang einzige Datensammlung stellt „Corona-Daten Deutschland“ strukturiert Informationen über die Maßnahmen zur Eindämmung der Pandemie bereit. Darüber hinaus bietet die Plattform unter anderem Zeitreihen sowie aktuelle Daten zum Infektionsgeschehen, zur Krankenhaus-situation und zur Mobilität.

„Corona-Daten Deutschland“ basiert auf einem Projekt, das im Jahr 2020 vom Bundesministerium für Wirtschaft und Energie initiiert wurde und bisher dezentrale Daten über Ländergrenzen vergleichbar machen soll. Das Statistische Bundesamt stellt die Datensammlung im Rahmen der Rubrik „EXDAT – Experimentelle Daten“ erstmals der interessierten Öffentlichkeit zur Verfügung. Das Informationsangebot soll sukzessive verbessert und weiter ausgebaut werden.

➤ www.corona-daten-deutschland.de

Freischaltung der Verwaltungsdaten- Informationsplattform

Die Verwaltungsdaten-Informationsplattform (VIP) des Statistischen Bundesamtes steht seit Ende Juli 2021 auch der Öffentlichkeit zur Verfügung. Sie ermöglicht Bürgerinnen und Bürgern, der Wirtschaft und der Wissenschaft wie auch der amtlichen Statistik selbst einen umfassenden Überblick über die in Registern geführten Datenbestände der deutschen Verwaltung.

Die integrierte Suchfunktion ermöglicht eine umfangreiche Recherche zu spezifischen Datenbeständen, wie auch zu Inhalt und Form der dort geführten Merkmale. Eine Kommentarfunktion bietet registrierten Nutzerinnen und Nutzern den fachlichen Austausch über Zugriffs- und Auswertungsmöglichkeiten. Das soll den dynamischen Diskurs zwischen amtlicher Statistik und Wissenschaft fördern und zu einer steten Fortentwicklung der Verwaltungsdaten-Informationsplattform beitragen.

Politik und Verwaltung insgesamt können die VIP als Werkzeug nutzen, um existierende Datenbestände besser miteinander zu verknüpfen, Mehrfachmeldungen entsprechend des Once-Only-Prinzips zu reduzieren und Nutzungspotenziale für neue Datenprojekte frühzeitig zu erkennen. Die amtliche Statistik wird die systematische Eignungsprüfung von Verwaltungsdatenquellen in ihr Geschäftsprozessmodell integrieren und wo immer möglich Belastungen durch Primärerhebungen reduzieren. Mit dem Aufbau der VIP leistet das Statistische Bundesamt einen Beitrag entsprechend der Datenstrategie der Bundesregierung für eine transparente und bürgerfreundliche Datenpolitik sowie der Open-Data-Strategie zur Verbesserung des Open-Data-Ökosystems des Bundes.

➤ www.verwaltungsdaten-info.de

AUS ALLER WELT

ISI-Weltkongress 2021

Der 63. Weltkongress des Internationalen Statistischen Instituts (ISI) im Juli 2021 wurde in virtueller Form durchgeführt. Die in zweijährlichem Turnus stattfindenden ISI-Weltkongresse sind ein Forum des weltweiten Austauschs insbesondere von Statistikerinnen und Statistikern aus Wissenschaft, amtlicher Statistik, Gesundheitssektor und Wirtschaft.

Ein besonderer Schwerpunkt lag in diesem Jahr auf der statistischen Messung der Auswirkungen der Corona-Pandemie auf die Menschen sowie den Chancen und Möglichkeiten, die sich hierdurch für die Weiterentwicklung der Statistik im methodischen Bereich ergeben.

Das Statistische Bundesamt hat sich mit mehreren Beiträgen und Präsentationen beteiligt:

- › Trusted Smart Surveys: Solutions for the European Statistical System – An overview of the objectives and the main challenges (Trusted Smart Surveys: Lösungen für das Europäische Statistische System. Ein Überblick über die Ziele und größten Herausforderungen)
- › The German Council Presidency in the field of statistics during the COVID-19 crisis (Die deutsche EU-Ratspräsidentschaft im Bereich Statistik in der Corona-Pandemie)
- › Mobility indicators based on mobile phone data (Mobilitätsindikatoren auf Basis von anonymisierten und aggregierten Mobilfunkdaten)
- › The potentials of data fusion methods to enhance income predictions for microsimulations (Die Potenziale von Datenfusionsmethoden zur Verbesserung der Einkommensprognosen für Mikrosimulationen)
- › The use of scanner data to compile regional price indices in Germany with an outlook to income inequality (Die Verwendung von Scannerdaten zur Erstellung regionaler Preisindizes in Deutschland mit einem Ausblick auf Einkommensungleichheit)
- › The new seasonally adjusted monthly cash tax revenues (Die neuen saisonbereinigten monatlichen kassenmäßigen Steuereinnahmen)

- › Improvement of the German cause of death statistics by electronic data recording and evaluation of multiple-cause-of-death data (Verbesserung der deutschen Todesursachenstatistik durch elektronische Datenerfassung und Auswertung von multikausalen Todesursachen)

Wie nachhaltig ist Deutschland?

Beim diesjährigen High Level Political Forum für Nachhaltige Entwicklung der Vereinten Nationen berichtete die Bundesregierung zum zweiten Mal über die Fortschritte Deutschlands beim Erreichen der Nachhaltigkeitsziele. Die Daten dazu liefert das Statistische Bundesamt seit nunmehr fünf Jahren. Das Forum spielt eine zentrale Rolle bei der Überprüfung der Agenda 2030 auf globaler Ebene.

Der diesjährige Fortschrittsbericht Deutschlands setzt sich hauptsächlich mit den Fragen auseinander, wie der doppelten Herausforderung der Pandemiebewältigung und der Transformation zu nachhaltiger Entwicklung begegnet werden kann, wie die Entwicklung der nationalen Nachhaltigkeitsindikatoren voranschreitet und wie Deutschland seiner globalen Verantwortung bei der Umsetzung der Sustainable Development Goals (SDGs) gerecht werden kann.

Der umfassende Bericht der Bundesregierung zum diesjährigen Forum, der sogenannte [Voluntary National Review](#), stützt sich in weiten Teilen auf die im März 2021 beschlossene Weiterentwicklung der Deutschen Nachhaltigkeitsstrategie und die vom Statistischen Bundesamt im [Indikatorenbericht](#) veröffentlichten Daten und deren Analysen.

18. Sitzung des Committee on Statistics and Statistical Policy der OECD

Nach einem Überblick über die gesamten geleisteten und geplanten statistischen Arbeiten der Organisation für wirtschaftliche Zusammenarbeit und Entwicklung (OECD) befasste sich das Committee on Statistics and Statistical Policy in seiner Sitzung im Juni 2021 mit folgenden Themen:

Politik und Gesellschaft haben inzwischen einen sehr hohen Bedarf an aktuellen und detaillierten Daten zur Verdichtung von Land durch neue Wohngebiete, insbesondere in den Ballungszentren. Die OECD hat beschlossen, Pilotprojekte zu starten, die eine Initiative zur Entwicklung qualitätsbereinigter Häuserpreisindizes auf regionaler Ebene darstellen. Dabei sollen Indizes möglichst nach Wohnungstypen (Ein- und Mehrfamilienhäuser) untergliedert werden. Deutschland veröffentlicht Häuserpreisindizes für fünf (segmentierte) subnationale Gliederungen. Darüber hinaus hat das Statistische Bundesamt ein von der Europäischen Union (EU) finanziertes Projekt zur Entwicklung von Indizes für Metropolregionen gestartet.

Mit dem Bericht „Measuring What Matters for Child Well-Being and Policies“ sollen Verbesserungen hinsichtlich der statistischen Erfassung der Lebensumstände von Kindern erzielt werden. Dazu zählen unter anderem das materielle Wohlbefinden und die Zeitverwendung von Kindern. Betroffen sind die einschlägigen Haushaltserhebungen, beispielsweise EU-SILC und die Zeitverwendungserhebung. Ein OECD-Vertreter stellte die wichtigsten Ergebnisse des Berichts und die nächsten Schritte vor. Die Implementierung eines neuen Messrahmens ist für das dritte oder vierte Quartal 2021 geplant, sowie bis November 2021 eine Überarbeitung des aktuellen OECD-Datenportals einschließlich des Dashboards zum Wohlbefinden von Kindern. Deutschland unterstützt diesen neuen Messrahmen im Grundsatz, sieht aber neue Datenanforderungen aufgrund der Belastung von Auskunftspflichtigen kritisch.

Innerhalb der OECD wurde im letzten Jahr ein eigener Politikbereich eingerichtet, der die Themen Wohlergehen und Abbau von sozialen Ungleichheiten behandelt. Das hierzu neu geschaffene Zentrum WISE (Well-being, Inclusion, Sustainability and Equal Opportunity) stellte seine Aktivitäten und die Planungen für die kommende Zeit vor. Eines der wesentlichen Arbeitsergebnisse ist die Entwicklung des Dashboards „For a strong, resilient, sustainable and inclusive recovery“, dessen Veröffentlichung für Oktober 2021 anlässlich des OECD-Ministertreffens geplant ist. Es enthält insgesamt 20 Indikatoren zu den Themen wirtschaftliche Entwicklung, Umwelt (Nachhaltigkeit) und Soziales (Inklusion, Wohlergehen).

Das Plenum diskutierte, wie die gesellschaftlichen, nicht-finanziellen Aktivitäten von Unternehmen gemessen werden können.

Faktoren, wie die Job-Zufriedenheit, ein soziales Sicherheitsgefühl oder höhere Bildung durch Fortbildung haben einen positiven Einfluss auf das Wohlbefinden („Well-Being“) einer Gesellschaft. Das WISE schlägt eine Liste von möglichen Indikatoren des Wohlbefindens für verschiedene Gruppen vor (Mitarbeiterinnen und Mitarbeiter, Firmeninhaberinnen und -inhaber, Konsumentinnen und Konsumenten sowie die Gesellschaft). Aus der Sicht Deutschlands ist die Initiative grundsätzlich zu begrüßen, aber in der Gesamtbewertung aufgrund der Belastung der Auskunftspflichtigen unter anderem bei den Haushaltsstatistiken und hinsichtlich der Messbarkeit sogenannter weicher Indikatoren in den Volkswirtschaftlichen Gesamtrechnungen auch als kritisch zu sehen.

Behandelt wurden des Weiteren die strategische Ausrichtung der Weiterentwicklung von SDMX, einem Standard zur Übertragung von Daten und Metadaten an internationale Organisationen, und ein Bericht zum neuen OECD-Mitglied Kolumbien.

AUS EUROPA

69. Plenarsitzung der Konferenz Europäischer Statistiker (CES)

Im Mittelpunkt der diesjährigen CES-Konferenz vom 23. bis 25. Juni 2021 standen die beiden Seminare „Climate Action“ und „Core Values“ sowie das zusammen mit der OECD durchgeführte gemeinsame Seminar zu COVID-19 und die Folgen und Herausforderungen für die Statistik.

Der Klimawandel stellt die Statistik vor neue Herausforderungen. Der Datenbedarf an klimabezogenen Daten ist rasant gestiegen und wird weiterhin steigen. Insgesamt beschließt die CES eine stärkere internationale Zusammenarbeit. Insbesondere die Entwicklung von Klassifikationen, Handbüchern und die Verknüpfung verschiedener geltender Standards soll zu zwischen den Ländern vergleichbaren Daten führen. Vereinbart wurde die Implementierung einer regelmäßigen Berichterstattung zur Messung der Klimaziele des Pariser Klimaschutzabkommens von 2015. Demnach wird mit einem von der Statistischen Abteilung der Vereinten Nationen

noch zu entwickelnden und mit den nationalen statistischen Ämtern abzustimmenden Indikatorenset aus den Bereichen Energie, Transport, Produktion, Handel, Volkswirtschaftliche Gesamtrechnungen, Land- und Forstwirtschaft, Landverbrauch, Abfall und Abwasser, Wohnen und Soziales, Tourismus und Umwelt zum ersten Mal am 31. Dezember 2024 und dann zweijährlich berichtet. Deutschland begrüßte den Kernsatz (core set) an Indikatoren, soweit diese vor allem auf den Bestimmungen des allgemein gültigen Rahmenwerks der Umweltgesamtrechnung, dem SEEA (System of Environmental-Economic Accounting), beruhen. Die Bedeutung aktueller Daten und die wichtige Rolle des SEEA wurde von der CES-Konferenz unterstrichen. Das Thema Verknüpfung von Klimadaten mit georeferenzierten Ansätzen wird voraussichtlich erst in der Plenartagung im Jahr 2022 behandelt werden.

Das zweite Seminar „core values“ beschäftigte sich mit den Grundwerten der Statistik. Neu ist der Kontext der Relevanz amtlicher Statistiken. Chancen bieten sich durch die Einführung neuer Produkte und durch eine systematische Umstellung auf einen ganzheitlichen Dienstleisteransatz einschließlich der federführenden Rolle der amtlichen Statistik gegenüber allen datenproduzierenden Stellen (Data Stewardship). Diskutiert wurde auch die Notwendigkeit einer besseren Kommunikationspolitik. In einem nächsten Schritt wird eine Liste mit Kernwerten der amtlichen Statistik entwickelt werden. Das CES-Büro wird diese Liste in der kommenden Plenarsitzung 2022 vorlegen und zur Diskussion stellen.

In einem gemeinsamen Seminar mit der OECD beriet die CES Schlussfolgerungen und Konsequenzen aus der Corona-Pandemie für die amtliche Statistik. Konsens aller Beteiligten war, dass die Pandemie die Gesamtentwicklung der amtlichen Statistik beschleunigt hat und das Ziel aller nationalen Statistikämter ist, das Erreichte, dazu gehören neue Produkte, neue Arbeitsformen, neue Nutzergruppen, zu sichern und weiter auszubauen. Der zweite Teil des Seminars befasste sich vor allem mit Innovationen in der Organisation und den Arbeitsstrukturen der nationalen statistischen Ämter.

Weitere wichtige Themen waren die Arbeitsprogramme sowie die Modernisierung der Statistik (Bericht über die Sitzung der High-Level Group for the Modernisation of Official Statistics), neue Formen der Arbeit und Kreislauf- und Abfallwirtschaft. Ziel beim Thema „Neue Formen der Arbeit“ ist es, für die bestehenden statistischen Rah-

menwerke verschiedener Organisationen (ILO, OECD, UNECE) einen konzeptionellen Überbau zu erarbeiten. Dabei geht es um die Definition von neu entstehenden Erwerbsformen, deren Einordnung und um die statistische Erfassung. Das Statistische Bundesamt wird sich an der speziell hierfür eingerichteten Task Force aktiv beteiligen.

Es besteht allgemein großes Interesse daran zu untersuchen, wie sich Initiativen zum Thema „Kreislaufwirtschaft“ auf die Klimaschutzpolitik auswirken, beispielsweise Treibhausgasemissionen bei der Herstellung von Produkten, Energieeinsparungen durch energieeffizientere Produkte, Abfallvermeidung, Erhöhung der Lebensdauer von Produkten, insgesamt die Reduzierung des Materialverbrauchs. Außerdem bieten vorhandene und neue Geschäftsmodelle (zum Beispiel Sharing-Economy-Ansätze wie Carsharing, Reparatur und Wiederaufarbeitung sowie Dienstleistungen und so weiter) das Potenzial, Arbeitsplätze zu schaffen und einen Beitrag zum Wirtschaftswachstum zu leisten. Deutschland unterstützt die Ausarbeitung von praktischen Leitlinien, einschließlich der Klärung von Schlüsselbegriffen und Definitionen. Die zuständige Task Force soll der CES-Plenarsitzung im Jahr 2023 über die Fortschritte berichten.

AUS DEM INLAND

Regelmäßige Daten zur Zeitverwendung ab 2022

Am 1. Juli 2021 ist das Gesetz über die statistische Erhebung der Zeitverwendung in Kraft getreten. Bereits seit den 1990er-Jahren erhebt das Statistische Bundesamt gemeinsam mit den Statistischen Ämtern der Länder etwa alle zehn Jahre Daten zur Zeitverwendung der in Deutschland lebenden Menschen, zuletzt 2012/13. Die neue gesetzliche Grundlage sichert nun die regelmäßige Durchführung dieser freiwilligen Erhebung. Damit findet die Zeitverwendungserhebung im Jahr 2022 erstmals nicht mehr als Erhebung für besondere Zwecke nach § 7 Bundesstatistikgesetz, sondern mit einer eigenen Gesetzesgrundlage statt.

Bei der Zeitverwendungserhebung erfassen alle Teilnehmenden ab zehn Jahren vollständige Tagesabläufe über 24 Stunden von Arbeit oder Schule über Hobbies

und Internetnutzung bis zu Einkaufen, Kinderbetreuung, ehrenamtlichen Tätigkeiten sowie den dazugehörigen Wegezeiten mit dem Auto, Bus, Bahn, Fahrrad oder zu Fuß. Aus den zusammengefassten Daten kann abgeleitet werden, wie Personen aus unterschiedlichen Bevölkerungsgruppen und Haushaltskonstellationen ihre Zeit auf verschiedene Lebensbereiche aufteilen. Die Daten sind eine wichtige Grundlage, insbesondere um gesellschaftspolitische Maßnahmen zum Beispiel zur besseren Vereinbarkeit von Familie und Beruf vorzubereiten und zu bewerten.

Besonderes Interesse gilt dem Umfang der „Care-Arbeit“, das heißt der unbezahlten Arbeit wie Kinderbetreuung, Hausarbeit, Ehrenamt und freiwilliges Engagement. Die Daten zur „Care-Arbeit“ ergänzen die klassischen Volkswirtschaftlichen Gesamtrechnungen, die bei der regelmäßigen Berichterstattung zu Wertschöpfung und Wohlstand die unentgeltlichen Leistungen der privaten Haushalte nicht einbeziehen.

Aber auch Informationen über die Dauer von Bildungs- oder Freizeitaktivitäten sowie zum Medienkonsum werden bei der Zeitverwendungserhebung erhoben. Somit kann aus den Daten ein umfassendes Bild über die durchschnittliche Zeitverwendung unterschiedlicher Bevölkerungsgruppen abgeleitet werden.

Da Erhebungen über die Zeitverwendung der Menschen in den meisten Mitgliedstaaten der Europäischen Union durchgeführt werden, sind auch EU-Vergleiche möglich.

➤ www.zve2022.de.

Stadt.Land.Zahl

Die Statistischen Ämter des Bundes und der Länder haben mit dem neuen Produkt Stadt.Land.Zahl eine Anwendung entwickelt, mit der man auf regionaler Ebene deutschlandweit statistische Ergebnisse vergleichen kann. Auf einen Blick sind wichtige Kennzahlen zu finden, wie die Bevölkerungszahl, das Bruttoinlandsprodukt oder die Arbeitslosenquote. Die Nutzerinnen und Nutzer können Ergebnisse für zwei Kreise und kreisfreie Städte übersichtlich auf einer Seite miteinander vergleichen. Ein Ranking hilft dabei, die eigene Region deutschlandweit einzuordnen. Datenquelle ist die Regionaldatenbank Deutschland.

➤ www.statistikportal.de

Update für den Verkehrsunfallatlas

Mit Mecklenburg-Vorpommern liegen für 2020 jetzt die Unfalldaten erstmals für ganz Deutschland im Open-Data-Format vor. Alle Unfälle mit Personenschaden – tödliche Unfälle, Unfälle mit Pkw, Lkw, Motorrädern, Fahrrädern oder Fußgängerinnen und Fußgängern – können nach Unfallstellen und Straßenabschnitten recherchiert werden. Der Unfallatlas umfasst für einige Bundesländer Angaben aus der Statistik der Straßenverkehrsunfälle seit dem Jahr 2016. Die genauen Angaben zum Unfallort (Geokoordinaten) werden von der Polizei bei der Unfallaufnahme erfasst und von den Statistischen Ämtern der Länder und dem Statistischen Bundesamt veröffentlicht.

➤ unfallatlas.statistikportal.de

Neues Basisregister für Unternehmen

Das Statistische Bundesamt hat den gesetzlichen Auftrag zum Aufbau eines Registers über Unternehmensbasisdaten erhalten. Das am 15. Juli 2021 in Kraft getretene Unternehmensbasisdatenregistergesetz soll Unternehmen und Verwaltung deutlich entlasten. Durch die Einführung einer bundeseinheitlichen und behördenübergreifenden Identifikationsnummer müssen Unternehmen ihre Stammdaten künftig nur noch einmal gegenüber der Verwaltung angeben (Once-Only-Prinzip). Das Register leistet einen wichtigen Beitrag zur Umsetzung des Onlinezugangsgesetzes.

Im Basisregister werden künftig Stammdaten wie Name, Sitz, Geschäftsanschrift, Rechtsform und Wirtschaftszweig sowie Identifikatoren aller in der deutschen Verwaltung geführten Unternehmen zusammengeführt, qualitätsgesichert und zentral vorgehalten. Dabei bestehen hohe Anforderungen an Datenschutz und Datensicherheit, die beim Aufbau und Betrieb des Registers stets berücksichtigt werden müssen. Jedes Unternehmen bekommt eine bundeseinheitliche Wirtschaftsnummer, mit der weitere Verwaltungsregister auf diese Unternehmensdaten zugreifen können. Angebundene Register werden direkt über neue Unternehmen und Änderungen an den Stammdaten bestehender Unternehmen informiert.

Zugriff auf Mikrodaten der Lebenslagenbefragungen zu Forschungszwecken

Seit 2015 erhebt das Statistische Bundesamt alle zwei Jahre im Auftrag der Bundesregierung in repräsentativen telefonischen Befragungen die Zufriedenheit von Bürgerinnen und Bürgern sowie Unternehmen mit Behördenkontakten in insgesamt 31 Lebenslagen. Diese reichen für Bürgerinnen und Bürger von der Ausbildung bis zum Erwerbsende und für Unternehmen von der Gründung bis zur Geschäftsaufgabe. In den letzten Lebenslagenbefragungen von 2019 gaben rund 6 000 Personen und mehr als 2 600 Firmen Auskunft zu ihrer Zufriedenheit mit der Verständlichkeit von Formularen, den Öffnungs- und Wartezeiten, dem E-Government-Angebot, den genutzten Informationsquellen und der Komplexität der Verfahren bis hin zum Vertrauen in die Ämter.

Die Ergebnisse aller Erhebungen sind unter [amtlich-einfach.de](https://www.amtlich-einfach.de) veröffentlicht. Nun stellen die Forschungsdatenzentren der amtlichen Statistik erstmals den Zugang zu anonymisierten Mikrodaten der Lebenslagenbefragungen von 2019 für Forschungszwecke bereit. Zudem erfolgte Anfang 2021 eine weitere Lebenslagenbefragung, deren Resultate im Herbst 2021 veröffentlicht werden sollen. Auch für diese Erhebung ist die Bereitstellung eines Scientific-Use-Files mit anonymisierten Mikrodaten geplant.

➤ www.forschungsdatenzentrum.de

INFORMATIONEN ZUM ZENSUS 2022

Neue Kontaktseite auf www.zensus2022.de

Ende Juli 2021 erfolgte die Freischaltung der neuen Kontaktseite auf der Zensus-Webseite. Mit der Anbindung aller Zensus-Auskunftsdienste des Bundes und der Länder wird sie der Haupteingangskanal für schriftliche Anfragen zum Zensus 2022 sein. Für die Anfragenden und die interessierte Öffentlichkeit wird der gemeinsame Auftritt im Verbund die Kommunikation mit den Ämtern erleichtern und zu mehr Transparenz führen.

Um eine zügige Bearbeitung aller Anfragen zu gewährleisten und gleichzeitig den manuellen Bearbeitungsaufwand für die Auskunftsdienste zu reduzieren, steht den Anfragenden eine interaktive FAQ-Suche zur automatisierten Beantwortung häufig gestellter Fragen zur Verfügung.

Sollten diese Antworten nicht ausreichend sein, können Anfragen über ein Kontaktformular direkt an eines der Statistischen Landesämter oder das Statistische Bundesamt versendet werden. Alternativ ist für Hörgeschädigte das Gebärdentelefon der Behördennummer 115 für Fragen zum Zensus zugänglich.

➤ www.zensus2022.de/kontakt

VERANSTALTUNGEN

Tagungsdokumentation zur 14. Wissenschaftlichen Tagung „Smart Surveys“ veröffentlicht

Zum Thema „Smart-Surveys – Neue Technologien bei Befragungen“ trafen sich am 24. und 25. Juni 2021 online rund 230 Vertreterinnen und Vertreter aus der amtlichen Statistik, Wissenschaft und Privatwirtschaft zu einem fachlichen Austausch über Smart Survey-Ansätze und neue Projekte.

Die Tagungsdokumentation ist nun unter www.destatis.de/wisstag21 verfügbar. Die Folienpräsentationen und Videoaufnahmen der einzelnen Vorträge und Diskussionen ermöglichen es, die Veranstaltung bei Interesse noch im Nachgang verfolgen zu können.

NEUERSCHEINUNGEN

OECD Skills Outlook 2021: Learning for Life

Schon vor der Pandemie nahmen nur zwei von zehn Erwachsenen mit niedrigem Bildungsstand an einer formalen oder berufsbegleitenden Weiterbildung teil,

verglichen mit sechs von zehn Erwachsenen mit hohem Bildungsstand. Die Teilnahme an Bildungsangeboten für Erwachsene ist auch von Land zu Land sehr unterschiedlich: Weniger als 25 % der Erwachsenen in Griechenland, Italien, Mexiko und der Türkei geben an, an der Weiterbildung teilzunehmen, verglichen mit über 55 % in Dänemark, Finnland, Neuseeland, Norwegen und Schweden.

Die aktuelle Ausgabe des „OECD Skills Outlook 2021: Learning for Life“ untersucht, wie politische Maßnahmen lebenslanges Lernen für alle am besten fördern können. Der Bericht nutzt vergleichende quantitative Daten, um hervorzuheben, welche Schlüsselrolle sozio-emotionale Faktoren und Motivation für Teilnahme am lebenslangen Lernen und den Erfolg spielen.

➤ manage.oecd-berlin.de

Implications of Remote Working Adoption on Place Based Policies: A Focus on G7 Countries

COVID-19 und die globalen Lockdowns haben die Digitalisierung von Arbeit und sozialen Interaktionen beschleunigt. Viele waren in dieser Zeit quasi gezwungen, tägliche Aufgaben auf virtuellem Wege zu erledigen. Dies hat zu einer größeren Verbreitung und Akzeptanz von Telearbeit geführt. Jetzt gilt es, Arbeitnehmerinnen und Arbeitnehmer sowie Firmen dabei zu unterstützen, dass diese Akzeptanz auch nach Ende der Pandemie bestehen bleibt.

Der neue OECD-Bericht „Implications of Remote Working Adoption on Place Based Policies: A Focus on G7 Countries“ empfiehlt eine Reihe von politischen Maßnahmen, um Regionen besser auf das vorzubereiten, was eine „neue Normalität“ sein könnte. Der Bericht stützt sich auf subnationale Echtzeitdaten, um Veränderungen in den Mobilitätsmustern der Menschen zu analysieren sowie die Faktoren zu untersuchen, die die Akzeptanz von Telearbeit über verschiedene Branchen und Regionen hinweg bestimmen.

➤ manage.oecd-berlin.de

OECD-Ausblick: kleine und mittlere Unternehmen

Die Coronakrise hat kleine und mittlere Unternehmen (KMU) weltweit stark strapaziert, doch dank staatlicher Unterstützungspakete haben sie die Krise weitgehend gemeistert. Bereits jetzt gibt es klare Zeichen der Erholung: In Australien und Frankreich beispielsweise liegt die Unternehmensgründungsrate mittlerweile 20 % über Vorkrisenniveau. Auch hat sich die Resilienz vieler KMU erhöht. So nutzten über 50 % der KMU in den OECD-Ländern während der Pandemie verstärkt digitale Anwendungen und konnten so ihren digitalen Rückstand gegenüber größeren Unternehmen verringern.

Der „SME and Entrepreneurship Outlook 2021“ analysiert die Maßnahmen, die KMU während der Krise das Überleben gesichert haben. Die Studie unterstreicht die enorme Bedeutung der staatlichen Hilfspakete und untersucht zudem, welchen Einfluss die Krise längerfristig haben könnte. Sie zeigt, welche Weichen Regierungen mit Blick auf KMU für einen grünen, nachhaltigen und inklusiven Aufschwung stellen sollten.

➤ manage.oecd-berlin.de

FLASH UND NOWCAST: SCHNELLSCHÄTZUNGEN DES BRUTTOINLANDSPRODUKTS IN DER CORONA-PANDEMIE

Arne Ackermann, Xaver Dickopf, Tanja Mucha

➤ **Schlüsselwörter:** Bruttoinlandsprodukt – Schnellschätzung – Nowcast – Flash Estimate – Corona

ZUSAMMENFASSUNG

Das Statistische Bundesamt veröffentlicht seit Juli 2020 eine erste Schnellschätzung des Bruttoinlandsprodukts (BIP) bereits 30 Tage nach Quartalsende. Vorausgegangen waren umfangreiche Tests zur Qualitätssicherung. Gleichzeitig wurden ab 2018 im Rahmen einer Machbarkeitsstudie die Möglichkeiten einer weiteren Beschleunigung zu einem BIP-t+10-Nowcast getestet. Dieser Beitrag stellt deren Ergebnisse vor und Revisionsanalysen der Schätzungen des BIP-t+10-Nowcast sowie des BIP-t+30-Flash vor und während der Corona-Pandemie gegenüber. Dabei zeigt sich, dass die BIP-Schnellschätzung nach 30 Tagen, die Expertenschätzungen mit ökonometrischen Verfahren kombiniert, dem rein modellgestützten Ansatz nach 10 Tagen vor allem in Krisenzeiten deutlich überlegen ist.

➤ **Keywords:** gross domestic product – early estimate – nowcast – flash estimate – corona

ABSTRACT

Since July 2020, the Federal Statistical Office has released its first estimate of the gross domestic product (GDP) as early as 30 days after the end of the quarter. This was preceded by comprehensive quality assurance tests. At the same time, a feasibility study, which had started in 2018, examined whether a further acceleration was possible in order to provide a t+10 GDP nowcast. This article presents the results of the feasibility study and compares revisions analyses of the t+10 GDP nowcast and the t+30 GDP flash estimate before and during the coronavirus pandemic. The analyses show that the GDP early estimate after 30 days, which combines expert estimations with econometric methods, is far superior to the purely model-based approach after 10 days, especially in times of crisis.

Arne Ackermann

hat Volkswirtschaftslehre in Marburg, Gießen sowie Lodz (Polen) studiert und ist wissenschaftlicher Mitarbeiter im Referat „Koordination, Veröffentlichung und Saisonbereinigung des Inlandsprodukts“ des Statistischen Bundesamtes. Neben seiner Mitwirkung an der Weiterentwicklung des „Dashboard Deutschland“ unterstützt er die methodische Weiterentwicklung und technische Umsetzung der ökonometrischen Schnellschätzung des Bruttoinlandsprodukts.

Xaver Dickopf

ist Volkswirt und Referent im Referat „Koordination, Veröffentlichung und Saisonbereinigung des Inlandsprodukts“ des Statistischen Bundesamtes. Schwerpunkt seiner Arbeit ist die ökonometrische Schnellschätzung des Bruttoinlandsprodukts mittels aktueller Konjunkturindikatoren und neuer digitaler Datenquellen sowie die Konjunkturbeobachtung und -analyse.

Tanja Mucha

ist Diplom-Volkswirtin und leitet das Referat „Koordination, Veröffentlichung und Saisonbereinigung des Inlandsprodukts“ im Statistischen Bundesamt. Neben der Schnellschätzung des Bruttoinlandsprodukts liegt der aktuelle Fokus ihrer Tätigkeit vor allem in der externen Kommunikation und Verbreitung von Daten der Volkswirtschaftlichen Gesamtrechnungen.

1

Einleitung

In wirtschaftlichen Krisenzeiten sind hochaktuelle Daten zur konjunkturellen Entwicklung besonders gefragt – das hat das Jahr 2020 mit dem Beginn der Corona-Pandemie deutlich gezeigt. Allerdings ist es gerade dann besonders schwierig, möglichst exakte Schnellschätzungen der gesamtwirtschaftlichen Entwicklung zu erstellen. Dies gilt insbesondere für Schätzungen mittels ökonomischer beziehungsweise zeitreihenanalytischer Verfahren. Solche modellgestützten Schätzungen beruhen in der Regel zwar auf einer Vielzahl von hochaktuellen Konjunkturindikatoren, aber auch auf den Daten beziehungsweise Zeitreihen der Vergangenheit. Dabei zeigt nicht zuletzt die Coronakrise, dass plötzlich und unerwartet eintretende Krisen durch solche Modelle nur schwer abgebildet werden können.

Die Schnellschätzung des Bruttoinlandsprodukts (BIP) $t+30$ Tage nach Ende eines Berichtsquartals, der sogenannte BIP- $t+30$ -Flash, kombiniert ökonomische Schätzverfahren mit Expertenschätzungen. Mit diesem bisher internen Verfahren hat das Statistische Bundesamt langjährige Erfahrungen gesammelt, auch während der weltweiten Finanz- und Wirtschaftskrise 2008/2009. Parallel dazu wurde seit 2018 als Reaktion auf das gestiegene Nutzerinteresse an schneller verfügbaren und gleichzeitig verlässlichen Daten zur Wirtschaftslage die weitere Beschleunigung der BIP-Schnellschätzung zu einem gesamtwirtschaftlichen Nowcast zum Zeitpunkt $t+10$ getestet. Erste Ergebnisse dieser Machbarkeitsstudie wurden Ende 2019 vorgestellt (Dickopf und andere, 2019; Rubrik „Experimentelle Daten“ unter www.destatis.de).

Die Analysen beider Schnellschätzungsverfahren im Zuge der Machbarkeitsstudie haben die hohe Qualität des BIP- $t+30$ -Flash bestätigt. Die einsetzende Coronakrise verstärkte zwar einerseits die Unsicherheiten der frühen BIP-Schätzungen. Andererseits war der Bedarf an schnell verfügbaren Indikatoren für die Konjunkturbeobachtung in dieser Zeit besonders hoch. Da die Genauigkeit der BIP-Ergebnisse nach $t+30$ Tagen weiter verbessert werden konnte und die gesetzten Qualitätskriterien erfüllt, veröffentlichte das Statistische Bundesamt im Juli 2020 wie vorher zu Jahresbeginn angekündigt erstmals

die beschleunigten BIP-Ergebnisse nach $t+30$ Tagen (Statistisches Bundesamt, 2020). Gleichzeitig zeigten die Ergebnisse der Machbarkeitsstudie, dass die frühe und rein modellgestützte Nowcast-Schätzung zum Zeitpunkt $t+10$ in „normalen“ Zeiten ohne Krise einen guten frühen Anhaltspunkt für die gesamtwirtschaftliche Entwicklung bietet. Die Coronakrise hat jedoch deutlich die Grenzen einer solchen gänzlich ökonomischen Schätzung der konjunkturellen Entwicklung aufgezeigt.

Die folgenden Kapitel stellen die Ergebnisse der Machbarkeitsstudie des BIP- $t+10$ -Nowcast vor und vergleichen diese mit den Ergebnissen des BIP- $t+30$ -Flash: Datengrundlage und -verfügbarkeit (Kapitel 2), statistische Methodik (Kapitel 3) und Qualitätsbeurteilung (Kapitel 4). Der Fokus der Qualitätsbeurteilung liegt dabei auf der Betrachtung der Auswirkungen der Coronapandemie auf die Schätzgenauigkeit der Modelle. Der Beitrag schließt mit einem Fazit und Ausblick auf die künftige Weiterentwicklung der BIP-Schnellschätzungen.

2

Datengrundlage und -verfügbarkeit

Im Zuge der Machbarkeitsstudie zum BIP- $t+10$ -Nowcast wurden insgesamt mehr als 120 Indikatoren auf ihre Eignung untersucht. Nach Abschluss der Studie fließen aktuell rund 30 dieser Konjunkturindikatoren in die ökonomischen Schätzungen des BIP- $t+10$ -Nowcast sowie des BIP- $t+30$ -Flash ein. Dazu zählen neben den amtlichen Konjunkturindikatoren auch einige nicht amtliche Stimmungs- und Umfrageindikatoren sowie auf neuen digitalen Daten basierende Indikatoren. Für künftige Weiterentwicklungen der BIP-Schnellschätzung ist vorgesehen, besonders im Bereich der neuen digitalen Daten weitere geeignete Indikatoren zu identifizieren und zu testen.

2.1 Amtliche Konjunkturindikatoren

Die Basis der BIP-Schnellschätzung bilden immer die jeweils verfügbaren amtlichen Konjunkturindikatoren. Dabei gilt grundsätzlich, je früher die Schätzung, desto weniger Basisdaten sind verfügbar und müssen modellgestützt hinzugeschätzt werden. Zum Zeitpunkt von BIP-

Flash und Nowcast liegen für den größten Teil der monatlich veröffentlichten amtlichen Ausgangsstatistiken noch nicht alle Monatswerte für das letzte Berichtsquartal vor. Dieses Problem betrifft vor allem den BIP-t+10-Nowcast, da zu diesem frühen Termin für viele Indikatoren zwei oder sogar alle drei Monatswerte der Basisstatistik fehlen und entsprechend geschätzt werden müssen. Beim BIP-t+30-Flash fehlt für die meisten Indikatoren nur ein Monatswert. Teilweise kommen auch Indikatoren auf Quartalsbasis zum Einsatz. Für diese Indikatoren wird der fehlende Quartalswert am aktuellen Rand mittels eines ökonometrischen Zeitreihenmodells hinzugeschätzt. Die in den derzeit verwendeten Modellen berücksichtigten amtlichen Konjunkturindikatoren lassen sich in folgende Gruppen zusammenfassen:¹

- › Auftragseingangsindizes (Industrie, Bauhauptgewerbe und andere)
- › Produktionsindizes (Produzierendes Gewerbe, Bauhauptgewerbe, Energie und andere)
- › Umsatzindizes (Einzelhandel, Handel mit Kraftfahrzeugen, Großhandel, Gastgewerbe und andere)
- › Außenhandelsstatistik des Statistischen Bundesamtes und [Zahlungsbilanzstatistik der Deutschen Bundesbank](#) (Importe und Exporte von Waren und Dienstleistungen)
- › [Lkw-Maut-Fahrleistungsindex](#) des Statistischen Bundesamtes und des Bundesamtes für Güterverkehr (BAG)
- › [Pkw-Neuzulassungen](#) des Kraftfahrt-Bundesamtes (KBA)

Insbesondere die Veröffentlichung der Daten der Außenhandelsstatistik wurde im Verlauf der Machbarkeitsstudie beschleunigt. Dies hat zusammen mit Vorabmeldungen zum Extra- und Intrahandel zur Verbesserung der Qualität der BIP-Schnellschätzungen nach t+10 und t+30 Tagen beigetragen. Perspektivisch wird die geplante Umstellung der amtlichen Statistiken im Dienstleistungsbereich von quartalsweiser auf monatliche Frequenz die BIP-Schnellschätzung auch nochmals deutlich voranbringen können.

¹ Die meisten dieser Indikatoren werden im Zeitverlauf revidiert. Nicht für alle verwendeten Indikatoren liegen historische Datenstände vor. Sofern für einen Indikator Echtzeitdaten in der Datenbank der Deutschen Bundesbank vorliegen, wurden diese einbezogen, um den jeweiligen Informationsstand in den Rück-Testrechnungen möglichst exakt abzubilden.

2.2 Nicht amtliche Stimmungs- und Umfrageindikatoren

Viele wissenschaftliche Studien verweisen auf die Prognosegüte konjunktureller Stimmungs- und Umfrageindikatoren für das BIP (Henzel/Rast, 2013). Zum Schätzzeitpunkt nach t+10 beziehungsweise t+30 Tagen liegen die Werte solcher Indikatoren in der Regel bereits für alle drei Monate des Berichtsquartals vor, was die Indikatoren zusätzlich interessant macht. Sie fließen nicht nur direkt in die BIP-Schätzmodelle ein, sondern dienen auch als externe Regressoren für die Schätzung der fehlenden Monate der amtlichen Konjunkturindikatoren. Derzeit werden folgende nicht amtliche Stimmungs- und Umfrageindikatoren für den Nowcast nach t+10 beziehungsweise den Flash nach t+30 Tagen verwendet:

- › [ifo Geschäftsklimaindex](#) (einschließlich Teilindizes im Verarbeitenden Gewerbe, im Bauhauptgewerbe, im Handel und im Dienstleistungsbereich), ifo Produktionserwartungen, ifo Exporterwartungen, ifo Importklima
- › [GfK Konsumklima](#)

Neben den oben beschriebenen konjunkturellen Stimmungs- und Umfrageindikatoren wurde im Zuge der Coronakrise der [Oxford-Stringency-Index](#) in die Modelle aufgenommen. Dieser soll die Stringenz (beziehungsweise Härte) der staatlichen Maßnahmen zur Eindämmung der Corona-Pandemie abbilden.

2.3 Neue digitale Daten

Die Bedeutung von neuen digitalen Daten hat für die amtliche Statistik in den letzten Jahren zugenommen (Wiengarten/Zwick, 2017). Auch im Bereich der BIP-Schnellschätzung gibt es verschiedene neue digitale Datenquellen, die die Schätzungen potenziell verbessern könnten und deshalb in die Testrechnungen der Machbarkeitsstudie einbezogen wurden. Im Ergebnis werden derzeit folgende Indikatoren neuer digitaler Datenquellen für die BIP-Schnellschätzungen verwendet:

- › [RWI/ISL-Containerumschlag-Index](#)
- › [Stromproduktionsdaten](#) des Fraunhofer-Instituts für Solare Energiesysteme (ISE)
- › [Wetterdaten](#) des Deutschen Wetterdiensts (DWD)

- › Trassenkilometer der DB Netz AG²
- › Passagier- und Güterumschlagzahlen im Flugverkehr ([ADV](#))
- › [Umsatzsteuervoranmeldungen](#) des Statistischen Bundesamtes (EXDAT)

Im Statistischen Bundesamt gibt es darüber hinaus derzeit viele Projekte, die darauf abzielen, mithilfe neuer digitaler Datenquellen zeitnah verfügbare neue Konjunkturindikatoren für den aktuellen Rand zu entwickeln. Die untersuchten Datenquellen reichen dabei von Scanner-, Fernerkundungs-, Mobilitäts- und Finanztransaktionsdaten bis hin zu neuen digitalen Daten im Bereich des Gütertransports. Beim seit Kurzem als experimentelle Statistik veröffentlichten [Transportindex](#) finden die letztgenannten Daten bereits Anwendung. Mit dem National Single Window der Wasserstraßen- und Schifffahrtsverwaltung wird beim Projekt Transportindex eine weitere auch für die BIP-Schnellschätzung potenziell interessante Datenquelle getestet. Weitere perspektivisch für die BIP-Schnellschätzung relevante Quellen aus dem Bereich der neuen digitalen Daten könnten auch solche aus Automatic Identification Systems (AIS) des Schiffsverkehrs sein, wie sie der [Kiel Trade Indicator](#) nutzt, oder Daten von Reservierungsplattformen wie [OpenTable](#). Diese und ähnliche neue digitale Datenquellen könnten für die künftige Weiterentwicklung der BIP-Schnellschätzungen auf ihre Eignung untersucht werden.

3

Statistische Methodik

Das Statistische Bundesamt greift für die Schätzungen des BIP-t+10-Nowcast und des BIP-t+30-Flash auf sogenannte ARIMA-Modelle³ zurück. Nicht nur beim ökonomischen Ansatz der BIP-Schnellschätzung, sondern auch im Bereich der Expertenschätzungen für den BIP-t+30-Flash finden solche zeitreihenanalytische Verfahren mittlerweile Anwendung (Rosengart, 2021). Beim rein ökonomischen Ansatz erfolgt die Modellierung in Anlehnung an die Bottom-up-Struktur der detaillier-

ten BIP-Berechnung. Des Weiteren wurden im Zuge der Machbarkeitsstudie für den BIP-t+10-Nowcast auch BIP-Schnellschätzungen mittels sogenannter dynamischer Faktormodelle (DFM) sowie mit Methoden des Machine Learning (ML) und der Künstlichen Intelligenz (KI) durchgeführt. Ihre Modellierung erfolgte nicht wie bei den ARIMA-Modellen bottom-up, sondern direkt beziehungsweise top-down.

3.1 Bottom-up-ARIMA-Modelle

Der ökonometrische Teil der BIP-Schnellschätzung nach t+10 beziehungsweise nach t+30 Tagen erfolgt im Statistischen Bundesamt standardmäßig mithilfe von ARIMA-Modellen, durch die das Bruttoinlandsprodukt indirekt beziehungsweise bottom-up geschätzt wird. Die zunächst geschätzten Werte für die Teilaggregate werden anschließend zu einem Gesamtergebnis aggregiert. Im Fall der BIP-Schnellschätzungen werden also erst die Bruttowertschöpfung einzelner Wirtschaftsbereiche beziehungsweise die einzelnen Aggregate der Verwendungsseite berechnet, die dann zu einem Entstehungs- beziehungsweise Verwendungsergebnis des Bruttoinlandsprodukts aggregiert werden (Dickopf und andere, 2019, hier: Seite 50 ff.). Die Schätzungen basieren dabei auf den realen Ursprungswerten der BIP-Aggregate (preisbereinigt, verkettet). Ein solcher Bottom-up-Ansatz ist einem Top-down-Ansatz, bei dem direkt das BIP geschätzt und anschließend auf die Teilaggregate heruntergebrochen wird, erfahrungsgemäß überlegen.

Um das optimale ARIMA-Modell für jedes Aggregat zu finden, wird der „Automatic Model Identification“-Algorithmus (AIM) verwendet, der auf Basis des korrigierten Informationskriteriums nach Akaike (AICC) das beste Modell für die jeweilige Schätzung bestimmt. Anschließend werden die ausgewählten Modelle mit der X-13-ARIMA-Methode geschätzt, die zeitgleich Ausreißer identifiziert und mögliche saisonale und kalenderbedingte Effekte modelliert (Grudkowska, 2016, hier: Seite 216 f.). Beide Methoden sind Teil des vom Statistischen Amt der Europäischen Union (Eurostat) empfohlenen Tools zur Saisonbereinigung JDemetra+ und können mithilfe des R-Pakets RJDemetra in der Programmiersprache R verwendet werden.

² Der Indikator bildet die monatlich bei der DB Netz AG gebuchten Streckenkilometer ab. Er wird nicht veröffentlicht, sondern liegt dem Statistischen Bundesamt zur internen Verwendung vor.

³ Modelle autoregressiver integrierter gleitender Mittelwerte (englisch: autoregressive integrated moving average, ARIMA).

Sofern für eine Zeitreihe geeignete Indikatoren zur Verfügung stehen, werden diese als externe Regressoren in das entsprechende Modell einbezogen. Je Zeitreihe gehen dabei bis zu fünf Indikatoren ein.¹⁴ Da die meisten Indikatoren auf Monatsbasis zur Verfügung stehen, während die Entstehungs- beziehungsweise Verwendungsaggregate auf Quartalsbasis geschätzt werden, sind die verwendeten Modelle als sogenannte Brückengleichungs-Modelle zu verstehen. Sie schlagen eine „Brücke“ zwischen erklärenden Monats- und erklärten Quartalsdaten. Eurostat empfiehlt die Verwendung dieser Modelle im Bereich der BIP-Schnellschätzung, wenn nur ein bis zwei Monatswerte der zugrundeliegenden Indikatoren zum Zeitpunkt der Schätzung zur Verfügung stehen und die fehlenden Monatswerte zugeschätzt werden müssen (Eurostat, 2016, hier: Seite 15 ff.).

Zum Schluss sind die verwendungsseitige und die entstehungsseitige BIP-Schätzung zu einem einzigen BIP-Ergebnis zusammenzufassen. Hier gibt es für die beiden Schnellschätzungen nach $t+10$ beziehungsweise $t+30$ Tagen unterschiedliche Vorgehensweisen. Die Berechnung eines abgestimmten BIP-Ergebnisses für den BIP- $t+30$ -Flash wird mit dem sogenannten Drei-Säulen-Modell durchgeführt, welches aufbauend auf mehreren Machbarkeitsstudien entwickelt wurde (Hartmann und andere, 2005). Die erste Säule bildet die Expertenschätzung, die zweite die ökonometrische Schätzung und die dritte die Abstimmung der ersten beiden Säulen (Dickopf und andere, 2019, hier: Seite 52 ff.). Beim BIP- $t+10$ -Nowcast erfolgt die Festlegung eines abgestimmten BIP-Ergebnisses bisher nur anhand der zweiten Säule. Dazu werden das verwendungsseitige und entstehungsseitige BIP-Ergebnis, wie in Dickopf und andere (2019, hier: Seite 52 ff.) beschrieben, gewichtet und zusammengefasst. Die BIP-Abstimmung mit dem Drei-Säulen-Modell ist ein entscheidender Faktor für die hohe Qualität des BIP- $t+30$ -Flash. Derzeit stellt sie für den BIP- $t+10$ -Nowcast aufgrund der geringeren Datenverfügbarkeit und des hohen Aufwands keine Alternative dar, könnte aber perspektivisch eine weitere Qualitätssteigerung bewirken.

¹⁴ Fehlende Monatswerte amtlicher Konjunkturindikatoren werden dabei teilweise zunächst in einem vorgelagerten Schritt hinzugeschätzt. Hierbei kommen wiederum ARIMA-Modelle unter Verwendung geeigneter früh verfügbarer nicht amtlicher Konjunkturindikatoren als externe Regressoren zum Einsatz.

3.2 Top-down-Faktormodelle

In der wissenschaftlichen Literatur sind dynamische Faktormodelle (DFM) heute Standard im Bereich makroökonomischer Prognosen, und neuere Ansätze übertragen diese Modelle auch in den Bereich des BIP-Nowcasting. Eurostat empfiehlt dynamische Faktormodelle im Bereich der BIP-Schnellschätzung, wenn – wie besonders im Fall des BIP- $t+10$ -Nowcast – nur wenige Basisdaten zur Verfügung stehen und eine detailliertere Modellierung daher nicht möglich ist (Eurostat, 2016, hier: Seite 18 ff.). Ihre Modellstruktur bietet die Möglichkeit, eine Vielzahl von Indikatoren in das Schätzmodell aufzunehmen, ohne dabei die Schätzqualität negativ zu beeinflussen. Bei vielen anderen ökonometrischen Modellen erhöhen zusätzliche Indikatoren auch die Anzahl der zu schätzenden Parameter, was die Schätzunsicherheit vergrößert und die Qualität der Schätzungen verschlechtert. Bei dynamischen Faktormodellen werden jedoch die verwendeten Indikatoren auf wenige Faktoren heruntergebrochen, wodurch zusätzliche Indikatoren nicht automatisch die Anzahl an Modellparametern erhöhen (Bok und andere, 2018, hier: Seite 626 ff.).

In der Machbarkeitsstudie zum BIP- $t+10$ -Nowcast wurden deshalb als Alternative zum ARIMA-Ansatz auch Top-down-Schätzungen mit Faktormodellen durchgeführt. Zielvariablen waren dabei nicht die einzelnen Aggregate der Verwendungs- beziehungsweise Entstehungsseite wie bei den Bottom-up-ARIMA-Modellen. Stattdessen wurde direkt ein BIP-Ergebnis geschätzt. Technisch wurden die Schätzungen mithilfe des [JDemetra+ Nowcasting Plugin](#) umgesetzt. Dieses Tool bietet den Nutzenden die Möglichkeit, die Schätzungen variabel anzupassen. Im Ergebnis lässt sich allerdings feststellen, dass die Qualität der Schätzungen mit dynamischen Faktormodellen noch nicht an die des klassischen Bottom-up-ARIMA-Ansatzes heranreicht. Aus diesem Grund wird im Folgenden von einer detaillierten Analyse der Genauigkeit der Faktormodelle abgesehen.

3.3 Methoden des Machine Learning und der Künstlichen Intelligenz

Wie die oben beschriebenen Verfahren, so ist auch Machine Learning (ML) eine Weiterentwicklung von Methoden der klassischen Regressionsanalyse. Sowohl bei ARIMA-Modellen und dynamischen Faktormodellen

als auch bei Ansätzen des Machine Learning wird der verwendete Datensatz zur Modellentwicklung in einen Trainings- und einen Testdatensatz unterteilt. In der Literatur finden Methoden des Machine Learning – zum Beispiel neuronale Netze, Bootstrapping oder Bagging – auch Anwendung für Kurzfristprognosen beziehungsweise Schnellschätzungen (Hyndman/Athanasopoulos, 2018, hier: Kapitel 11.3 und 11.4). Im Verlauf der Machbarkeitsstudie wurden in Zusammenarbeit mit dem Referat „Künstliche Intelligenz, Big Data“ im Statistischen Bundesamt unterschiedliche ML-Methoden, zum Beispiel [Random Forests](#), [XGBoost](#) und [neuronale Netze](#), für den BIP-t+10-Nowcast getestet. Zwar konnte mit der XGBoost-Methode der vielversprechendste Nowcast aller ML-Methoden erstellt werden. Allerdings war die Genauigkeit noch etwas niedriger als bei der Verwendung dynamischer Faktormodelle. Somit stellen ML-Methoden zum jetzigen Zeitpunkt keine sinnvolle Alternative zur ARIMA-Modellierung dar. Im weiteren Verlauf des Aufsatzes wird daher auf eine detaillierte Analyse der Genauigkeit der Schätzungen mit ML-Methoden verzichtet.

4

Qualitätsbeurteilung

Ein wichtiger Ansatz für die Qualitätsbeurteilung des BIP-t+10-Nowcast stellt Backtesting dar, also die Qualitätsbeurteilung der Regressionsmodelle anhand historischer Datenstände. Mit diesem Ansatz kann mittels Rück-Testrechnungen überprüft werden, wie gut das ökonometrische Modell das BIP-Ergebnis vergangener Quartale getroffen hätte.⁵ Die in der Machbarkeitsstudie zum BIP-t+10-Nowcast verwendeten Daten wurden daher in einen Trainings- und einen Testdatensatz unterteilt. Mit dem Trainingsdatensatz, der den Zeitraum vom ersten Quartal 1994 bis zum vierten Quartal 2015 umfasst, wurde das ökonometrische Modell (in-sample) angepasst beziehungsweise kalibriert. Anhand des Testdatensatzes und der Echtzeitdaten, die den Zeitraum vom ersten Quartal 2016 bis zum zweiten Quartal 2019 umfassen, wurde die Qualität des Nowcast-Modells (out-

of-sample) beurteilt. Das verwendete Verfahren wird in der Literatur auch als „time series cross validation“ oder „rolling window forecasting“ bezeichnet (Hyndman/Athanasopoulos, 2018, hier: Kapitel 3.4). Seit dem dritten Quartal 2019 wird der BIP-t+10-Nowcast im laufenden Betrieb weiterentwickelt. Dieser Aufsatz betrachtet die Ergebnisse bis zum vierten Quartal 2020.

4.1 Qualitätskriterien

Um die Qualität der BIP-Schnellschätzungen nach t+10 beziehungsweise t+30 Tagen zu beurteilen, wurden zwei übliche Revisions- beziehungsweise Abweichungsmaße herangezogen: die mittlere Revision („mean revision“, MR) und die mittlere absolute Revision („mean absolute revision“, MAR).

$$MR_{t+n} = \frac{\sum_{t=1}^T (y_{t,t+n} - \hat{y}_{t,t+z})}{T}$$

$$MAR_{t+n} = \frac{\sum_{t=1}^T |y_{t,t+n} - \hat{y}_{t,t+z}|}{T}$$

Hierbei bezeichnet $\hat{y}_{t,t+z}$ den BIP-t+10-Nowcast beziehungsweise den BIP-t+30-Flash für Quartal t zum Zeitpunkt $t+z$ und $y_{t,t+n}$ das jeweilige BIP-Ergebnis für Quartal t zum Zeitpunkt $t+n$. Der BIP-t+10-Nowcast und der BIP-t+30-Flash werden dabei mit den abgestimmten und veröffentlichten Ergebnissen zu t+45/t+55 verglichen.⁶ Die Zeiträume, für die die Revisionsmaße berechnet wurden, umfassen einmal die $T=16$ Quartale vom ersten Quartal 2016 bis zum vierten Quartal 2019 (Qualitätsbeurteilung vor der Pandemie). Zur Beurteilung der beiden Modelle in Krisenzeiten bezieht der zweite Zeitraum ($T=20$) die Quartale vom ersten Quartal 2016 bis zum vierten Quartal 2020 und damit auch die vier ersten Quartale der Corona-Pandemie mit ein.

⁵ Um den jeweiligen Datenstand zum Schätzzeitpunkt t+10 für zurückliegende Quartale zu simulieren, werden Echtzeitdaten der Vergangenheit genutzt. Im Zuge der Machbarkeitsstudie wurden hierfür die nötigen historischen Datenstände der jeweiligen Zeitreihen der Volkswirtschaftlichen Gesamtrechnungen sowie, soweit vorhanden, die historischen Datenstände der Indikatoren zusammengestellt.

⁶ Bis zum ersten Quartal 2020 wurde das Bruttoinlandsprodukt erstmals nach t+45 Tagen veröffentlicht, die Details nach t+55 Tagen. Seit dem zweiten Quartal 2020 veröffentlicht das Statistische Bundesamt den BIP-t+30-Flash. Die Veröffentlichung nach t+45 Tagen wurde zeitgleich eingestellt. Die erste Veröffentlichung detaillierter BIP-Ergebnisse erfolgt nach wie vor zu t+55. Da das BIP-Ergebnis zwischen t+45 und t+55 grundsätzlich nicht revidiert wurde, handelt es sich dabei bis einschließlich erstes Quartal 2020 jeweils um dieselben Ergebnisse. Für die Revisionsanalysen in diesem Bericht wird daher bis zum ersten Quartal 2020 das BIP-Ergebnis zu t+45/t+55 und ab dem zweiten Quartal 2020 das BIP-Ergebnis zu t+55 als Vergleichsmaßstab herangezogen.

Als Bedingung für die erstmalige Veröffentlichung der BIP-Schnellschätzung wurden dieselben drei Qualitätskriterien gewählt, die auch vor der erstmaligen Veröffentlichung des „GDP Preliminary Flash“ durch Eurostat im Jahr 2016 erfüllt sein mussten (Eurostat, 2016a). Diese waren durch eine europäische Task Force definiert worden, an der auch Deutschland beteiligt war:

1. Die mittlere Revision, MR, von t+30 zu t+45 soll zwischen $-0,05$ und $+0,05$ Prozentpunkten liegen.
2. Die mittlere absolute Revision, MAR, von t+30 zu t+45 soll nicht mehr als $0,10$ Prozentpunkte betragen; zu t+65 nicht mehr als $0,13$ Prozentpunkte.
3. Es sollen nicht mehr als $66,7\%$ der Revisionen in eine Richtung gehen.

Diese Qualitätskriterien wurden zwar mit Blick auf das Bruttoinlandsprodukt der Europäischen Union beziehungsweise der Eurozone entwickelt und beziehen sich auf die Revisionen von t+30 zu t+45 beziehungsweise t+65. Sie sind jedoch grundsätzlich auch auf Deutschland und die Revisionen des BIP-t+10-Nowcast übertragbar. Auch für den BIP-t+30-Flash wurden dieselben Qualitätskriterien angewendet. Das Statistische Bundesamt entschied sich für die erstmalige Veröffentlichung der Ergebnisse mit dem zweiten Quartal 2020, weil alle drei Qualitätskriterien erfüllt waren. Daher erfolgt auch die Qualitätsbeurteilung des BIP-t+10-Nowcast in Anlehnung an diese Qualitätskriterien.

4.2 Qualitätsbeurteilung vor und während der Coronakrise

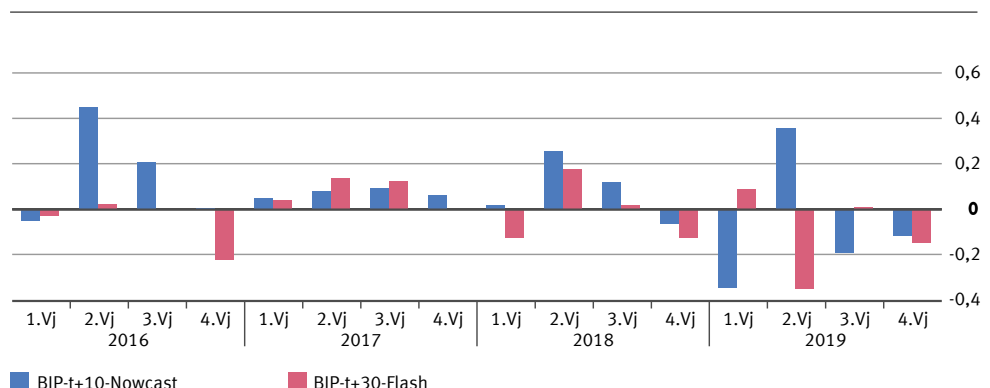
Die im ersten Quartal 2020 einsetzende Corona-Pandemie bedeutete einen deutlichen Einschnitt in der Machbarkeitsstudie zum BIP-t+10-Nowcast. Die Testrechnungen waren bis dahin bei stabiler konjunktureller Entwicklung und unter sehr günstigen gesamtwirtschaftlichen Rahmenbedingungen durchgeführt worden. Die Corona-Pandemie verursachte im Jahr 2020 eine weltweite wirtschaftliche Krise, die eine große Herausforderung für alle Konjunkturschätzungen darstellte. Bei der Qualitätsbeurteilung wird deshalb unterschieden zwischen dem Zeitraum vor Einsetzen der Krise (bis zum vierten Quartal 2019) und dem Gesamtzeitraum einschließlich der Coronakrise (bis zum vierten Quartal 2020).

Für einen ersten Eindruck, wie genau die ARIMA-Modelle vor der Corona-Pandemie t+10 beziehungsweise t+30 Tage nach Ende des Referenzquartals das BIP-Ergebnis nach 55 Tagen getroffen haben, werden zunächst die Revisionen in Prozentpunkten betrachtet. Dazu werden hier und in der nachfolgenden Revisionsanalyse immer die Wachstumsraten des BIP gegenüber dem Vorjahresquartal (beziehungsweise die Abweichungen dieser Wachstumsraten) verwendet.¹⁷ [Grafik 1](#) zeigt die Abweichungen der BIP-Schätzergebnisse nach t+10 beziehungsweise nach t+30 Tagen von den nach t+55 Tagen veröffentlichten Ergebnissen der ersten detaillier-

¹⁷ Eine Revisionsanalyse der saison- und kalenderbereinigten Wachstumsraten gegenüber dem Vorquartal führt zu vergleichbaren Ergebnissen.

Grafik 1

Abweichungen der Ergebnisse der BIP-Schnellschätzungen von den Ergebnissen der ersten detaillierten BIP-Berechnung für den Zeitraum 2016 bis 2019 (vor der Pandemie)
Prozentpunkte



2021 - 0281

ten BIP-Berechnung für den Zeitraum vom ersten Quartal 2016 bis zum vierten Quartal 2019. Das Ergebnis ist zunächst wenig überraschend: Der BIP-t+10-Nowcast weicht in den meisten Quartalen stärker vom BIP-Ergebnis nach t+45/t+55 Tagen ab als der BIP-t+30-Flash. Es gibt aber durchaus auch einige Quartale, in denen der frühere BIP-t+10-Nowcast das erste Rechenergebnis besser geschätzt hat als der spätere BIP-t+30-Flash. Insgesamt trifft der BIP-t+30-Flash die Rechenergebnisse nach t+45/t+55 Tagen offenbar wie erwartet etwas besser als der BIP-t+10-Nowcast.

Diese Vermutung bestätigt sich bei Betrachtung der Revisionsmaße: Die Ergebnisse der Revisionsanalysen des Testzeitraums vor der Corona-Pandemie für den BIP-t+10-Nowcast und den BIP-t+30-Flash im Vergleich zu den BIP-Ergebnissen der ersten detaillierten Berechnung nach t+45/t+55 Tagen beziehungsweise der zweiten detaillierten Berechnung ein Quartal später (t+1Q) enthält [Tabelle 1](#).

Sowohl die MR mit –0,025 Prozentpunkten als auch die MAR mit 0,100 Prozentpunkten fallen beim BIP-t+30-Flash im Vergleich zum BIP-Ergebnis t+45/t+55 Tage nach Ende des Referenzquartals besser aus als beim BIP-t+10-Nowcast (MR: 0,057 Prozentpunkte, MAR: 0,153 Prozentpunkte). Der BIP-t+30-Flash erfüllt damit für den betrachteten Zeitraum die ersten beiden oben beschriebenen Qualitätskriterien, der BIP-t+10-Nowcast dagegen nicht. Gleiches gilt für das dritte Qualitätskriterium. Während der Anteil der Revisionen in eine Richtung (hier: Revisionen nach oben) für den BIP-t+10-Nowcast leicht über dem europäischen Qualitätskrite-

rium von 66,7% liegt, erfüllt der BIP-t+30-Flash auch dieses Kriterium. Beim Vergleich der Revisionsmaße der Schnellschätzungen mit den späteren BIP-Ergebnissen der zweiten detaillierten Berechnung (nach t+1Q) fällt dagegen auf, dass die MAR des BIP-t+30-Flash mit 0,183 deutlich größer ist als beim Vergleich mit dem BIP-Ergebnis nach t+45/t+55 Tagen. Sie ist vergleichbar mit der MAR des BIP-t+10-Nowcast mit 0,182. Zieht man also die zweite detaillierte BIP-Berechnung als Vergleichsmaßstab heran, lässt sich beim Kriterium der MAR kein qualitativer Unterschied mehr zwischen BIP-t+10-Nowcast und BIP-t+30-Flash ausmachen.

Beim Vergleich der Revisionsmaße des BIP-t+10-Nowcast mit den Ergebnissen von 2019 (Dickopf und andere, 2019: MAR 0,16 Prozentpunkte zu der t+55-Veröffentlichung für den Zeitraum vom ersten Quartal 2016 bis zum zweiten Quartal 2019) wird deutlich, dass die Weiterentwicklung des BIP-t+10-Nowcast die Genauigkeit der Schätzergebnisse bereits leicht verbessert hat. Entsprechend könnten künftige Weiterentwicklungen der Methodik und der verwendeten Konjunkturindikatoren perspektivisch zu genaueren Ergebnissen führen, die die festgelegten europäischen Qualitätskriterien erfüllen.

[Tabelle 2](#) stellt die Abweichungen der BIP-Rechnungsergebnisse nach t+10 beziehungsweise nach t+30 Tagen von den nach t+45/t+55 Tagen veröffentlichten Ergebnissen für den gesamten Zeitraum vom ersten Quartal 2016 bis zum vierten Quartal 2020 dar. Sie verdeutlicht, dass sowohl beim BIP-t+10-Nowcast als auch beim BIP-t+30-Flash in den vier von der Corona-Pandemie betroffenen Quartalen 2020 die Genauigkeit der Ergeb-

Tabelle 1

Ergebnisse der Revisionsanalysen des Testzeitraums vor der Corona-Pandemie für den BIP-t+10-Nowcast und den BIP-t+30-Flash

	BIP-t+10-Nowcast gegenüber		BIP-t+30-Flash gegenüber	
	t+45/t+55- Veröffentlichung	t+1Q- Veröffentlichung	t+45/t+55- Veröffentlichung	t+1Q- Veröffentlichung
	in Prozentpunkten			
MR (mittlere Revision)	0,057	0,096	– 0,025	0,015
MAR (mittlere absolute Revision)	0,153	0,182	0,100	0,183
Maximale Revision	+ 0,45 (2. Quartal 2016)	+ 0,45 (2. Quartal 2016)	– 0,35 (2. Quartal 2019)	– 0,44 (2. Quartal 2019)
in %				
Anteil der Revisionen nach oben	68,8	68,8	56,3	62,5

Revisionsmaße der Veränderungsraten gegenüber dem Vorjahresquartal, Ursprungswerte; Bezugszeitraum: 1. Quartal 2016 bis 4. Quartal 2019

nisse erheblich geringer war als im Zeitraum vor der Krise. Die Abweichungen des BIP-t+10-Nowcast in den vier Quartalen 2020 sind dabei wesentlich größer als die des BIP-t+30-Flash. Die BIP-Schätzungen nach t+10 Tagen fielen während des wirtschaftlichen Einbruchs im Zuge der ersten Pandemiewelle im ersten und zweiten Quartal 2020 deutlich zu positiv aus und mussten stark nach unten revidiert werden. Auch die wirtschaftliche Erholung im dritten Quartal 2020 konnte der BIP-t+10-Nowcast nicht präzise prognostizieren, die Schätzung war viel zu negativ und musste um 1,5 Prozentpunkte nach oben revidiert werden. Im vierten Quartal 2020 war der Revisionsbedarf mit knapp 0,5 Prozentpunkten wieder moderater. Dies zeigt die grundsätzliche Problematik ökonometrischer beziehungsweise zeitreihenanalytischer Verfahren, plötzlich eintretende Krisen akkurat abzubilden. Während der BIP-t+10-Nowcast in der Corona-Pandemie je nach Quartal nach unten oder oben revidiert werden musste, waren die Schätzungen des BIP-t+30-Flash in allen vier Quartalen zu pessimis-

tisch, mussten aber nur um maximal 0,4 Prozentpunkte nach oben revidiert werden.

➤ **Tabelle 2** auf Seite 26 enthält die Ergebnisse der Revisionsanalyse des Testzeitraums einschließlich der Corona-Pandemie für den BIP-t+10-Nowcast und den BIP-t+30-Flash im Vergleich zu den nach t+45/t+55 Tagen beziehungsweise den ein Quartal später veröffentlichten BIP-Ergebnissen.

Hier bestätigt sich, was schon in Grafik 2 zu erkennen war: Während der Corona-Pandemie haben sich die Revisionsmaße von BIP-t+10-Nowcast und BIP-t+30-Flash zum Teil deutlich verschlechtert. Das gilt insbesondere für die Schätzung des BIP-t+10-Nowcast (MAR 0,314 Prozentpunkte gegenüber MAR 0,153 Prozentpunkte vor Corona-Pandemie), die sich in der Krise wesentlich stärker verschlechtert hat als die Schätzung des BIP-t+30-Flash (MAR 0,136 Prozentpunkte gegenüber MAR 0,100 Prozentpunkte vor Corona-Pandemie). Auch beim Vergleich der beiden Schnellschätzungen mit den BIP-

Grafik 2

Abweichungen der Ergebnisse der BIP-Schnellschätzungen von den Ergebnissen der ersten detaillierten BIP-Berechnung für den Zeitraum 2016 bis 2020 (vor und während der Pandemie)
Prozentpunkte

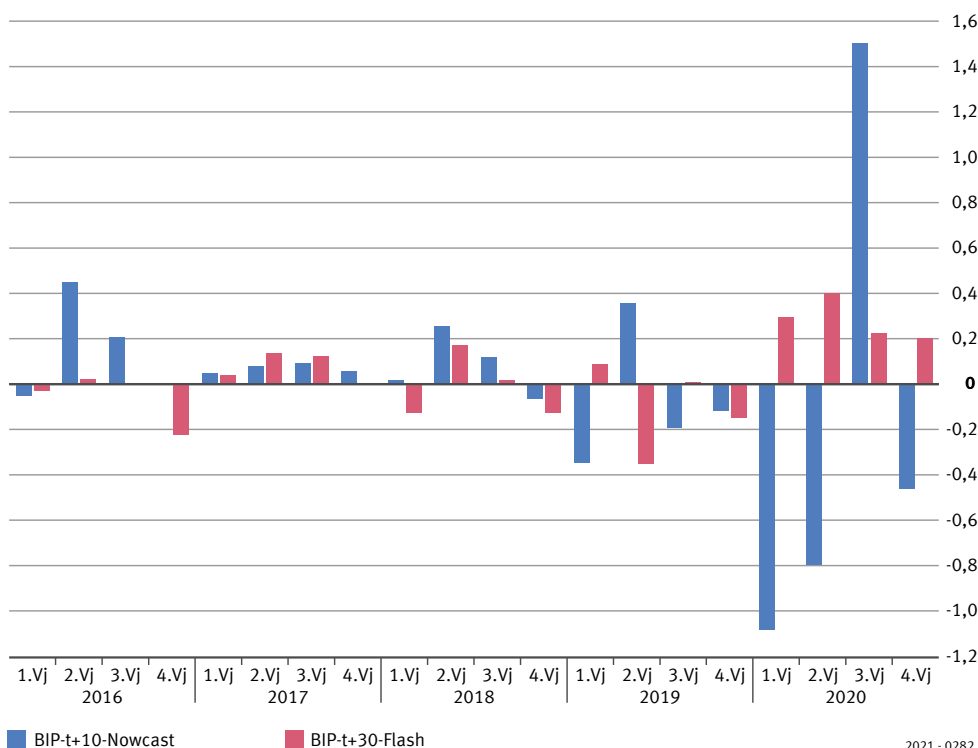


Tabelle 2

Ergebnisse der Revisionsanalyse des Testzeitraums einschließlich der Corona-Pandemie für den BIP-t+10-Nowcast und den BIP-t+30-Flash

	BIP-t+10-Nowcast gegenüber		BIP-t+30-Flash gegenüber	
	t+45/t+55- Veröffentlichung	t+1Q- Veröffentlichung	t+45/t+55- Veröffentlichung	t+1Q- Veröffentlichung
	in Prozentpunkten			
MR (mittlere Revision)	0,004	0,058	0,036	0,091
MAR (mittlere absolute Revision)	0,314	0,315	0,136	0,226
Maximale Revision	+ 1,50 (3. Quartal 2020)	+ 1,51 (3. Quartal 2020)	+ 0,40 (2. Quartal 2020)	+ 0,56 (4. Quartal 2020)
in %				
Anteil der Revisionen nach oben	60,0	60,0	65,0	70,0

Revisionsmaße der Veränderungsraten gegenüber dem Vorjahresquartal, Ursprungswerte; Bezugszeitraum: 1. Quartal 2016 bis 4. Quartal 2020


Ergebnissen nach t+1Q zeigen die Revisionsmaße eine Verschlechterung der Modelle während der Corona-Pandemie. Der Unterschied zwischen den Revisionsmaßen fällt dabei erneut beim BIP-t+10-Nowcast deutlich höher aus (MAR 0,315 Prozentpunkte gegenüber MAR 0,182 Prozentpunkte vor Corona-Pandemie) als beim BIP-t+30-Flash (MAR 0,226 Prozentpunkte gegenüber MAR 0,183 Prozentpunkte vor Corona-Pandemie). Die Ergebnisse bestätigen die These, dass ökonometrische Modelle in Krisenzeiten deutlich an Genauigkeit einbüßen. Allerdings gelingt es der BIP-t+30-Flash-Schätzung, zumindest einen Teil der Schätzunsicherheit in der Corona-Pandemie durch die vollständigere Datenbasis und die Verwendung des Drei-Säulen-Modells zu kompensieren. Zwar werden auch in der BIP-t+30-Flash-Schätzung zeitnah verfügbare konjunkturelle Frühindikatoren aus den Bereichen der Stimmungs- und Umfrageindikatoren sowie der neuen digitalen Daten in die ökonometrischen Modelle einbezogen. Die Expertenschätzung als weitere Säule kann aber auf plötzlich eintretende Ereignisse besser reagieren und die Schätzungen entsprechend besser anpassen, als dies bei ökonometrischen Modellen möglich ist.

5

Fazit und Ausblick

Die vorgestellten Analysen zur Schnellschätzung des BIP haben die hohe Qualität des BIP-t+30-Flash bestätigt. Die Ergebnisse der Machbarkeitsstudie zeigen aber ebenso, dass sowohl der BIP-t+10-Nowcast als auch der BIP-t+30-Flash während der Corona-Pandemie eine geringere Genauigkeit aufwiesen als vorher. Während beim BIP-t+30-Flash in den vier betrachteten Krisenquartalen des Jahres 2020 tendenziell etwas stärker nach oben revidiert werden musste, wies der BIP-t+10-Nowcast insgesamt deutlich größere Abweichungen in beide Richtungen auf. Es bestätigt sich demnach die These, dass ökonometrische Modelle plötzlich eintretende Ereignisse nur unzureichend abbilden können. Gleichzeitig haben die BIP-Ergebnisse während der Corona-Pandemie nochmals verdeutlicht, wie wichtig die Expertenschätzung für die BIP-Schnellschätzung ist. In der Krise war der rein ökonometrische BIP-t+10-Nowcast dem BIP-t+30-Flash, in dem Expertenschätzungen mit ökonometrischen Schätzungen kombiniert werden, qualitativ deutlich unterlegen. Die Schätzungen zum Zeitpunkt t+10 waren sehr viel ungenauer als zum Zeitpunkt t+30, was nur zum Teil der weniger umfangreichen Datengrundlage geschuldet war. Insgesamt konnte die Genauigkeit des BIP-t+10-Nowcast im Verlauf der Machbarkeitsstudie aber leicht verbessert werden. Es hat sich zudem gezeigt, dass die frühe und rein modellgestützte Nowcast-Schätzung zum Zeitpunkt t+10 in „normalen“ Zeiten ohne Krise einen guten frühen Anhaltspunkt für

die gesamtwirtschaftliche Entwicklung bietet. Die Coronakrise hat jedoch auch deutlich die Grenzen einer solchen gänzlich ökonometrischen Schätzung der konjunkturellen Entwicklung aufgezeigt.

Abschließend ist festzustellen, dass die Machbarkeitsstudie zum BIP-t+10-Nowcast die BIP-Schnellschätzung insgesamt vor allem aus zwei Gründen deutlich vorangebracht hat: Zum einen wurden neue nicht amtliche Datenquellen erschlossen, die nun auch in der Expertenschätzung berücksichtigt werden und die BIP-Schnellschätzung dort verbessern, wo noch keine amtlichen Basisstatistiken vorliegen. Zum anderen finden nun ökonometrische Ansätze an bestimmten Stellen innerhalb der Expertenschätzung Anwendung, wo sie die etablierten Schätzansätze bereichern und ergänzen (Rosengart, 2021). Vor diesem Hintergrund sollte sich die künftige Weiterentwicklung der BIP-Schnellschätzungen vor allem auf die weitere Verzahnung von Expertenschätzung und Ökonometrie, die Erschließung weiterer (neuer digitaler) Datenquellen (zum Beispiel Scanner-, Fernerkundungs-, Mobilitäts-, Finanztransaktions- und Transportdaten) und die Ausweitung der Analysen auf die BIP-Aggregate der Entstehungs- und Verwendungsrechnung konzentrieren. 

LITERATURVERZEICHNIS

Bok, Brandyn/Caratelli, Daniele/Giannone, Domenico/Sbordone, Argia M./Tambalotti, Andrea. *Macroeconomic Nowcasting and Forecasting with Big Data*. In: Annual Review of Economics. Ausgabe 10/2018, Seite 615 ff.

Breiman, Leo. *Random Forests*. 2001. [Zugriff am 9. Juli 2021]. Verfügbar unter: link.springer.com

Dickopf, Xaver/Janz, Christian/Mucha, Tanja. *Vom BIP-Flash zum BIP-Nowcast: Erste Ergebnisse einer Machbarkeitsstudie zur weiteren Beschleunigung der BIP-Schnellschätzung*. In: WISTA Wirtschaft und Statistik. Ausgabe 6/2019, Seite 47 ff.

Eurostat. *Overview of GDP flash estimation methods*. In: Eurostat statistical working papers. 2016b. Seite 15 ff. [Zugriff am 9. Juli 2021]. Verfügbar unter: ec.europa.eu

Grudkowska, Sylwia. *JDemetra+ Reference Manual Version 2.1*. 2016. [Zugriff am 9. Juli 2021]. Verfügbar unter: ec.europa.eu

Hartmann, Norbert/Schmidt, Joachim/Oltmanns, Erich. *Schnellschätzungen für das Bruttoinlandsprodukt: Ergebnisse einer Machbarkeitsstudie*. In: Wirtschaft und Statistik. Ausgabe 7/2005, Seite 682 ff.

Henzel, Steffen R./Rast, Sebastian. *Prognoseeigenschaften von Indikatoren zur Vorhersage des Bruttoinlandsprodukts in Deutschland*. In: ifo Schnelldienst. Ausgabe 17/2013, Seite 39 ff. [Zugriff am 9. Juli 2021]. Verfügbar unter: www.ifo.de

Hyndman, Rob J./Athanasopoulos, George. *Forecasting: Principles and practice*. Melbourne 2018. Zweite Auflage. [Zugriff am 9. Juli 2021]. Verfügbar unter: otexts.com

Maggi, Mats. *JDemetra+ Nowcasting*. 2017. [Zugriff am 9. Juli 2021]. Verfügbar unter: github.com

Rosengart, Nathalie. *Schnellschätzung des Außenhandels in den Volkswirtschaftlichen Gesamtrechnungen: ein ökonometrischer Ansatz*. In: WISTA Wirtschaft und Statistik. Ausgabe 2/2021, Seite 19 ff.

Statistisches Bundesamt. *Bruttoinlandsprodukt im 2. Quartal 2020 um 10,1 % niedriger als im Vorquartal*. Pressemitteilung Nr. 287 vom 30. Juli 2020. [Zugriff am 8. Juli 2021]. Verfügbar unter: www.destatis.de

Stava, Tavish Sriva. *How to use XGBoost algorithm in R in easy steps*. 2016. [Zugriff am 1. Juni 2021]. Verfügbar unter: www.analyticsvidhya.com

Wiengarten, Lara/Zwick, Markus. *Neue digitale Daten in der amtlichen Statistik*. In: WISTA Wirtschaft und Statistik. Ausgabe 5/2017, Seite 19 ff.

Wuttke, Laurenz. *Deep Learning Grundlagen*. 2020. [Zugriff am 9. Juli 2021]. Verfügbar unter: datasolut.com

NEUE MONATSBERICHTE DER TODESURSACHENSTATISTIK — MIT FOKUS AUF DEM NACHWEIS VON COVID-19-STERBEFALLZAHLEN

Andrea Buschner, Eva Kibele, Ulrike Winkelmann (geteilte
Erstautorenschaft), Olaf Eckert

📌 **Schlüsselwörter:** Todesursachen – Sterbefälle – Grundleiden – Begleit-
erkrankung – COVID-19

ZUSAMMENFASSUNG

Während der Corona-Pandemie zeigte sich, dass genaue und zeitnahe Informationen über diejenigen Sterbefälle benötigt werden, die im Zusammenhang mit der Infektionskrankheit stehen. Bundesweite Angaben zu den Todesursachen auf Grundlage der Todesursachenstatistik liegen planmäßig Mitte August des Folgejahres vor. Deshalb wurden bislang die Daten der SARS-CoV-2-Meldestatistik des Robert Koch-Instituts für Vergleiche mit den allgemeinen Sterbefallzahlen herangezogen. Die Statistischen Ämter des Bundes und der Länder haben eine monatliche Todesursachenstatistik mit vorläufigen Daten entwickelt, die einen ersten Einblick in das Mortalitätsgeschehen geben soll. Der Beitrag stellt die Entwicklung dieser vorläufigen Monatsberichte vor und präsentiert erste ausgewählte Monatsergebnisse zu COVID-19 als Todesursache.

📌 **Keywords:** *causes of death – deaths – underlying cause of death – concomitant disease – COVID-19*

ABSTRACT

The coronavirus pandemic has shown the need for accurate and timely information on deaths that are related to the infectious disease. Information on causes of death in the whole of Germany is usually available from the causes of death statistics by mid-August of the following year. For this reason, the data of the SARS-CoV-2 statistics reported by the Robert Koch Institute have so far been used for comparisons with the general mortality figures. The Federal Statistical Office and the statistical offices of the Länder have developed monthly causes of death statistics containing provisional data, which are intended to provide a first overview of the mortality situation. This article explains the development of the provisional monthly reports based on the causes of death statistics and presents first selected monthly results on COVID-19 as a cause of death.

Dr. Andrea Buschner

ist Diplom-Soziologin und im Bayerischen Landesamt für Statistik im Sachgebiet „Bevölkerung, Kompetenzzentrum Demographie“ tätig.

Dr. Eva Kibele

ist Diplom-Demografin und leitet das Referat „Bevölkerung und Demografie, Bautätigkeit, Soziales, Gesundheit, Statistische Methoden“ des Statistischen Landesamtes Bremen.

Dr. Ulrike Winkelmann

ist Diplom-Volkswirtin und im Statistischen Landesamt Baden-Württemberg im Referat „Bevölkerung, Gesundheit, Rechtspflege, Insolvenzen“ tätig.

Dr. Olaf Eckert

ist Diplom-Physiker und hat an der Medizinischen Hochschule Hannover im Fach Humanbiologie promoviert. Als Referent im Referat „Gesundheitsstatistiken“ des Statistischen Bundesamtes arbeitet er an der technischen, inhaltlichen und methodischen Weiterentwicklung der Todesursachenstatistik.

1

Einleitung

Änderungen im Morbiditäts- und Mortalitätsgeschehen sind meist erst im langfristigen Vergleich sichtbar. Ausnahmen davon sind beispielsweise Grippewellen oder Hitzewellen sowie Naturkatastrophen oder Pandemien, die zu erhöhten Sterbefallzahlen führen. Die Corona-Pandemie hat nun ein besonders stark erhöhtes öffentliches und wissenschaftliches Interesse an aktuellen Daten der Todesursachenstatistik hervorgerufen. Fragen zur Zahl der durch eine COVID-19-Infektion hervorgerufenen Sterbefälle, zur Demografie der Verstorbenen, zu deren regionaler Verteilung, zum Zeitverlauf des Pandemiegeschehens und zu Vor- und Begleiterkrankungen stehen häufig im Mittelpunkt der Diskussionen. Ebenso interessiert die Frage, ob eine Person [an](#) oder [mit](#) einer COVID-19-Erkrankung verstorben ist.

Teile dieser Fragestellungen können nun mithilfe einer neuen Veröffentlichung, den Monatsberichten der Todesursachenstatistik, beantwortet werden. Dieser Beitrag stellt in Kapitel 2 die Rahmenbedingungen für die Erstellung endgültiger, jährlicher Ergebnisse und die Entwicklung vorläufiger Monatsberichte der Todesursachenstatistik vor. Kapitel 3 zeigt, wie in der Todesursachenstatistik nach dem Regelwerk der Weltgesundheitsorganisation (WHO – World Health Organization) Sterbefälle [an](#) und [mit](#) COVID-19 ermittelt werden. Ausgewählte Ergebnisse ermöglichen im vierten Kapitel eine erste Darstellung der COVID-19-Sterbefälle in den Monaten Januar 2020 bis Dezember 2020 nach der neuen Monatsstatistik. Die beiden Schlusskapitel diskutieren die Vor- und Nachteile der vorläufigen monatlichen Todesursachenstatistik und geben einen Ausblick auf mögliche Weiterentwicklungen.

2

Die Abläufe in der Todesursachenstatistik

2.1 Jährliche Todesursachenstatistik

Die Todesursachenstatistik ist eine gefragte Datenquelle für epidemiologische und gesundheitspolitische Fragestellungen. So bieten beispielsweise Indikatoren der vorzeitigen Sterblichkeit wesentliche Ansatzpunkte für Präventions- und Interventionsmaßnahmen. Änderungen in der Sterblichkeit nach Todesursachen vollziehen sich in der Regel langfristig. Daher liegt der Fokus der etablierten Todesursachenstatistik auf qualitätsgesicherten und sehr differenzierten – also nach bis zu 3 000 verschiedenen Todesursachen gegliederten – Daten. Nach Abschluss eines Berichtsjahrs, planmäßig acht Monate nach Ende des jeweiligen Kalenderjahrs, veröffentlichen Bund und Länder die Ergebnisse mit einem breiten Tabellenprogramm unter Wahrung der statistischen Geheimhaltung.

Der Jahresabschluss der Todesursachenstatistik ist erst möglich, wenn die Statistik der Sterbefälle abgeschlossen ist und alle Statistischen Landesämter die vorliegenden Todesbescheinigungen vollständig signiert und plausibilisiert haben:

- › Die Todesursachenstatistik ist der Statistik der Sterbefälle nachgeordnet. Diese weist alle von einem Standesamt beurkundeten Sterbefälle nach Alter, Geschlecht und Nationalität aus. Die Zahl der Sterbefälle im jeweiligen Berichtsmonat muss daher zwischen beiden Statistiken synchronisiert werden.
- › Die Statistik der Sterbefälle unterliegt seit 2018 der Geheimhaltung durch Vergrößerung und Zellspernung. Daher sind auch die Veröffentlichungen der Todesursachenstatistik entsprechend anzupassen, um die Geheimhaltung der Bevölkerungsstatistik nicht aufzudecken.
- › Der Jahresabschluss der Todesursachenstatistik beinhaltet neben dem Abgleich mit den Sterbefällen der Bevölkerungsstatistik weitere Plausibilitätsprüfungen und den sogenannten Länderaustausch (siehe Abschnitt 2.2).

Störungen dieser Prozesse, beispielsweise der Rückstand eines einzelnen Landes aufgrund eigener Personalengpässe oder ausstehender Datenlieferungen der auskunftspflichtigen Gesundheitsämter können bundesweite Verzögerungen verursachen. Während der Corona-Pandemie kam es teilweise zu großen Rückständen bei der monatlichen Lieferung der Todesbescheinigungen an die Statistischen Ämter der Länder.

2.2 Entwicklung einer monatlichen Todesursachenstatistik mit vorläufigen Daten

Vorgehen und Voraussetzungen

Die Monatsberichte der Todesursachenstatistik sollen – basierend auf vorläufigen Daten – einen ersten Einblick in das Mortalitätsgeschehen geben. Vorläufig sind die Daten vor allem, weil sie zunächst nicht vollzählig sind, aber auch aufgrund ihrer Anbindung an die Sterbefallstatistik. Änderungen dort, beispielsweise eine nachträgliche Korrektur des Wohnorts oder der Staatsangehörigkeit, wirken sich auch auf die Todesursachenstatistik aus. Die monatliche Berichterstattung der Todesursachenstatistik orientiert sich daher am Berichtsmonat der Sterbefallstatistik.¹ Ein vom Standesamt gemeldeter Sterbefall wird dabei nach dem Sterbedatum dem Berichtsmonat zugewiesen, in der Regel entspricht dies dem Ereignismonat. Nach vier Wochen sind bundesweit 97 % der Sterbefälle registriert (zur Nieden und andere, 2020). Erfolgt die Meldung des Sterbefalls mit mehrmonatigem Verzug und ist die Bearbeitung der Sterbefallzahl des Berichtsmonats bereits abgeschlossen, wird der Sterbefall dem nächstmöglichen Berichtsmonat zugeordnet. Für die monatliche Todesursachenstatistik kann anhand dieser vorläufigen monatlichen Sterbefallzahlen berechnet werden, für welchen Anteil bereits eine Todesursachenkodierung vorliegt.

Die Statistischen Landesämter übernehmen die Todesursachenkodierung für die im jeweiligen Bundesland registrierten Sterbefälle. Veröffentlicht wird die Todesursachenstatistik jedoch nach dem Wohnort der Verstorbenen, was auch für die Monatsberichte einen vor-

läufigen Länderaustausch erfordert. Dabei werden alle Sterbefälle den Wohnorten der Verstorbenen zugeordnet, das heißt die Sterbefälle werden unter den Statistischen Landesämtern ausgetauscht. So wird die regional korrekte Abbildung des Mortalitätsgeschehens sichergestellt. Vor allem in Stadtstaaten und den umliegenden Bundesländern entstünden sonst Verzerrungen, da hier überproportional häufig Sterbe- und Wohnort in unterschiedlichen Bundesländern liegen.

Für die Monatsberichte der Todesursachenstatistik gilt Folgendes: Sind für mindestens 50 % der Sterbefälle die Todesursachen plausibel kodiert und können alle genutzten Fälle genau einem Sterbefall aus der Bevölkerungsstatistik zugeordnet werden, erfolgt die Veröffentlichung der Daten auf Bundesebene. Für die Landesebene liegt die Schwelle bei 80 % der Sterbefälle. Dieses Vorgehen soll sicherstellen, dass Abweichungen zur später veröffentlichten Jahresstatistik möglichst gering ausfallen. Die Monatsberichte der Todesursachenstatistik stellen fortlaufend revidierte und vervollständigte Ergebnisse dar. Somit verbessert sich die Qualität der Berichte zu späteren Veröffentlichungszeitpunkten mit höherem Vollzähligkeitsgrad immer mehr. Da es im Laufe des Jahres immer wieder zu Nachmeldungen und Korrekturen kommt, stellen erst die Jahresdaten sicher, dass alle Sterbefälle korrekt und vollzählig mit einer Todesursache versehen sind.

Vorläufig sind die Monatsergebnisse der Todesursachenstatistik zudem durch den reduzierten Merkmalskranz. Zunächst beschränkt sich die Veröffentlichung der monatlichen Todesursachenstatistik auf die unikausale Ausweisung des Grundleidens. Dabei werden die wichtigsten ICD-Kapitel, Diagnosegruppen und Einzeldiagnosen ausgewählt. Zusätzlich werden COVID-19-Sterbefälle ausgewiesen, das heißt die Monatsberichte enthalten sowohl Sterbefälle, in denen COVID-19 das Grundleiden darstellte ([an COVID-19 verstorben](#)), als auch jene Sterbefälle, bei welchen COVID-19 eine Begleiterkrankung war ([mit COVID-19 verstorben](#)). ➔ [Übersicht 1](#)

1 Einzelne Länder veröffentlichen ihren Monatsbericht auf Basis des Sterbemonats.

Übersicht 1

Merkmalskranz mit ausgewählten Todesursachen

Position/Nummer der ICD-10	Todesursache
A00-B99	KAPITEL I: Bestimmte infektiöse und parasitäre Krankheiten
C00-D48 darunter: C00-C97 C15-C26 C30-C39 C50	KAPITEL II: Neubildungen Bösartige Neubildungen Bösartige Neubildungen der Verdauungsorgane Bösartige Neubildungen der Atmungsorgane und sonstiger intrathorakaler Organe Bösartige Neubildung der Brustdrüse
E00-E90 darunter: E10-E14	KAPITEL IV: Endokrine, Ernährungs- und Stoffwechselkrankheiten Diabetes mellitus
F00-F99 darunter: F00-F03 F10-F19	KAPITEL V: Psychische und Verhaltensstörungen Demenz Psychische und Verhaltensstörungen durch psychotrope Substanzen
G00-G99	KAPITEL VI: Krankheiten des Nervensystems
I00-I99 darunter: I10-I15 I20-I25 I21 I30-I52 I60-I69 I64 I70-I79	KAPITEL IX: Krankheiten des Kreislaufsystems Hypertonie (Hochdruckkrankheit) Ischämische Herzkrankheiten Akuter Myokardinfarkt Sonstige Formen der Herzkrankheit Zerebrovaskuläre Krankheiten Schlaganfall, nicht als Blutung oder Infarkt bezeichnet Krankheiten der Arterien, Arteriolen und Kapillaren
J00-J99 darunter: J09-J18 J40-J47	KAPITEL X: Krankheiten des Atmungssystems Grippe und Pneumonie Chronische Krankheiten der unteren Atemwege
K00-K93 darunter: K70-K77	KAPITEL XI: Krankheiten des Verdauungssystems Krankheiten der Leber
N00-N99 darunter: N17-N19	Kapitel XIV: Krankheiten des Urogenitalsystems Niereninsuffizienz
R00-R99	KAPITEL XVIII: Symptome und abnorme klinische und Laborbefunde, die anderenorts nicht klassifiziert sind
V01-Y98 darunter: V01-V99 X60-X84	KAPITEL XX: Äußere Ursachen Transportmittelunfälle Vorsätzliche Selbstbeschädigung
U00- U49 darunter: U07.1 U07.2	KAPITEL XXII: Schlüsselnummern für besondere Zwecke COVID-19, durch einen Labortest nachgewiesen COVID-19, Virus ist klinisch-epidemiologisch bestätigt, jedoch nicht durch einen Labortest nachgewiesen
	sonstige Kapitel (D50-D89; H00-H95; L00-L99; M00-M99; O00-O99; P00-P96; Q00-Q99)
A00-Y98	Todesursachen insgesamt
nachrichtlich: davon: U07.1 U07.2	Anzahl der Fälle, in denen COVID-19 als Begleiterkrankung angegeben wurde, aber nicht ursächlich für den Tod gewesen ist Begleiterkrankung COVID-19, durch einen Labortest nachgewiesen Begleiterkrankung COVID-19, Virus ist klinisch-epidemiologisch bestätigt, jedoch nicht durch einen Labortest nachgewiesen

ICD-10: Internationale statistische Klassifikation der Krankheiten und verwandter Gesundheitsprobleme.

Geheimhaltung

Die Begrenzung auf einen ausgewählten Merkmalskranz bei der erstmaligen Bereitstellung der vorläufigen Monatsberichte der Todesursachenstatistik und die nicht vollzählige Erfassung aller Sterbefälle eines Landes (vorläufige Daten) erfordert in diesem Schritt nur in Einzelfällen eine Geheimhaltung. Je nach Fallzahl in den einzelnen Bundesländern nehmen diese eine Geheimhaltung vor.

Sensitivitätsanalyse: wie aussagekräftig ist die monatliche Todesursachenstatistik?

Die Aussagekraft der vorläufigen monatlichen Todesursachenstatistik ist möglicherweise eingeschränkt, da die Daten noch nicht vollzählig sind. Eine Selektivität der monatlichen Todesursachenstatistik könnte beispielsweise auftreten, wenn Todesbescheinigungen mit bestimmten Todesursachen überproportional häufig verspätet bei den Statistischen Landesämtern eingehen und damit deutlich später in die Monatsberichte der Todesursachenstatistik einfließen.

In ersten Selektivitätsanalysen wurde daher überprüft, ob die zunächst nicht in den Monatsberichten enthaltenen Fälle aufgrund verzögerter Lieferungen der Gesundheitsämter über- oder unterproportional häufig bestimmte Todesursachen enthielten. Bei COVID-19-Sterbefällen war dies jedoch nicht der Fall: Der Anteil der Sterbefälle mit Grundleiden COVID-19 blieb relativ stabil, unabhängig von der Vollzähligkeit der Lieferungen der Todesbescheinigungen. Jedoch kann Selektivität zu anderen Zeiten oder in bestimmten Regionen nicht grundsätzlich ausgeschlossen werden.

3

Von der Todesbescheinigung zum Grundleiden

3.1 Angaben auf der Todesbescheinigung

Datenbasis der Todesursachenstatistik ist der sogenannte vertrauliche Teil der Todesbescheinigung, die die leichenschauenden Ärztinnen und Ärzte bei jedem

Sterbefall ausfüllen. Dabei ist im Abschnitt „Todesursache“, Teil 1, der Krankheitsverlauf von der zugrundeliegenden Erkrankung oder äußeren Ursache bis hin zur unmittelbaren Todesursache in einer Kausalkette zu dokumentieren. Das Grundleiden ist definiert als „a) die Krankheit oder Verletzung, die den Ablauf der direkt zum Tode führenden Krankheitszustände auslöste, oder b) die Umstände des Unfalls oder der Gewalteinwirkung, die den tödlichen Ausgang verursachten“ (DIMDI, 2019). In Teil 2 der Todesbescheinigung sind andere wesentliche Krankheiten und Zustände aufzuführen, die nicht im Zusammenhang mit dem Grundleiden stehen. Soweit vorhanden fließen auch über Obduktionsscheine Informationen zu präziseren Todesursachen in die Todesursachenstatistik ein.

Die vertraulichen Teile der Todesbescheinigung gehen über das für den Sterbeort zuständige Standesamt an das Gesundheitsamt, wo sie geprüft und anschließend an die Statistischen Landesämter übermittelt werden. Je nach Bundesland erfolgt diese Übermittlung in Form der handschriftlich ausgefüllten Papier-Todesbescheinigung oder elektronisch. Damit ist die Todesursachenstatistik eine Vollerhebung für die Wohnbevölkerung Deutschlands.

3.2 Kodierung der Todesursachen nach ICD-10

In den Statistischen Landesämtern wird anhand der Angaben auf der Todesbescheinigung unter Anwendung des komplexen Regelwerks der Internationalen statistischen Klassifikation der Krankheiten und verwandten Gesundheitsprobleme (ICD-10) das Grundleiden bestimmt. Rund 75 % der Sterbefälle des Jahres 2020 wurden mithilfe des elektronischen Kodiersystems „Iris/MUSE“ kodiert², welches das Regelwerk der ICD-10 anwendet und eine Texterkennung für medizinische Terme enthält (Eckert, 2019). Obwohl die Bestimmung des Grundleidens multikausal erfolgt, das heißt unter Berücksichtigung aller auf der Todesbescheinigung vermerkten Erkrankungen (siehe Abschnitt 3.1), ist die Darstellung der Todesursachenstatistik bislang unikausal: Es wird ausschließlich das Grundleiden veröffentlicht.

² Todesbescheinigungen mit der Angabe COVID-19 wurden in 13 von 16 Bundesländern multikausal mit Iris/MUSE kodiert.

3.3 Kodierung der Todesursachen am Beispiel COVID-19

Am 20. April 2020 hat die WHO kurzfristig neue ICD-10-Kodes eingeführt, mit denen in der Todesursachenstatistik eine COVID-19-Erkrankung mit Virusnachweis (U07.1) oder ohne Virusnachweis (U07.2) kodiert wird. Die Schlüsselnummer U07.1 wird auch dann angewandt, wenn ärztlicherseits „COVID-19“ ohne expliziten Hinweis auf einen Virusnachweis dokumentiert ist. Die Schlüsselnummer U07.2 soll dagegen verwendet werden, wenn COVID-19 klinisch-epidemiologisch bestätigt ist und das Virus nicht durch Labortest nachgewiesen wurde oder kein Labortest zur Verfügung stand.

Ob COVID-19 als Grundleiden in die amtliche Statistik einfließt, hängt entscheidend davon ab, an welcher Position auf der Todesbescheinigung, mit welcher Diagnosesicherheit und wie plausibel der leichenschauende Arzt beziehungsweise die leichenschauende Ärztin dieses Krankheitsbild bescheinigt hat (WHO, 2020). Die WHO-Kodierregeln für COVID-19 sind analog zu denjenigen für Influenza aufgebaut: Ist auf dem Totenschein COVID-19 als Startpunkt einer plausiblen Kausalkette von Folgeerkrankungen (beispielsweise Pneumonie) verzeichnet, so wird COVID-19 als Grundleiden festgelegt. Laut ICD-10 gilt der Grundsatz, dass COVID-19 (ebenso wie Influenza) nicht durch andere Krankheiten verursacht werden kann.

Die typische Bearbeitung von Todesbescheinigungen mit Erwähnung einer COVID-19-Infektion und die damit verbundene Entscheidung, ob eine Person **an** oder **mit** COVID-19 verstorben ist (Möbius, 2020), wird für vier Fallbeispiele erläutert.

(1) **An** COVID-19 Verstorbene mit Virusnachweis

➤ **Übersicht 2** zeigt eine vereinfachte Todesbescheinigung (linke Spalte), bei welcher nach zeilenweiser Zuweisung der ICD-10-Kodes (rechte Spalte) das Kodiersystem Iris/MUSE den ICD-10-Kode U07.1 als Grundleiden ermittelt. Dabei wird das von der WHO vorgegebene Regelwerk angewendet, das in diesem Fall die in Teil 1 angegebene Kausalkette als plausibel akzeptiert.

Während in diesem Beispiel das Grundleiden in der dafür vorgesehenen Zeile 1c zu finden ist, geben die Zeilen 1a und 1b die direkt zum Tode führenden Krankheiten oder

Übersicht 2

An COVID-19 verstorben:
Beispiel einer Todesbescheinigung (links)
mit multikausaler Kodierung nach ICD-10 (rechts)

81 Jahre, weiblich, natürlicher Tod		Multikausale ICD-10-Kodierung
1a	Akutes Lungenversagen	J96.0
1b	Lungenentzündung	J18.9
1c	COVID-19	U07.1 → Grundleiden
2	Diabetes mellitus Typ 2	E11.9

Zustände (1a) sowie vorausgegangene Ursachen und Krankheitszustände (1b) an. Die in den Zeilen 1a und 1b auf der Todesbescheinigung angegebenen Krankheiten sind Folgen der COVID-19-Erkrankung und demnach Begleiterkrankungen. Die in Zeile 2 angegebene Diabetes-Erkrankung ist eine Erkrankung, die nicht ursächlich zum Tode beigetragen hat. Für diesen Sterbefall wird der Kode U07.1 in die unikausale Todesursachenstatistik aufgenommen.

(2) **Mit** COVID-19 Verstorbene mit Virusnachweis

Bei einer **mit** COVID-19 verstorbenen Person wird nach Auswertung der Todesbescheinigung ein abweichender ICD-10-Kode als Grundleiden ausgewählt, obwohl U07.1 als multikausaler ICD-10-Kode dokumentiert ist. Die Person ist „**mit**“ COVID-19 verstorben, da COVID-19 in diesem Fall eine Begleiterkrankung ist.

➤ **Übersicht 3** zeigt eine Todesbescheinigung, bei der ein Herzinfarkt (ICD-10-Kode I21.9) der Startpunkt einer Kausalkette in Teil 1 ist und damit auch als Grundleiden in die Todesursachenstatistik einfließt. Die Angabe von COVID-19 als Begleiterkrankung in Teil 2 hat hier keinen Einfluss auf die Auswahl des Grundleidens.

Übersicht 3

Mit COVID-19 verstorben:
Beispiel einer Todesbescheinigung (links)
mit multikausaler Kodierung nach ICD-10 (rechts)

81 Jahre, weiblich, natürlicher Tod		Multikausale ICD-10-Kodierung
1a	Herzversagen	I46.9
1b	Herzinfarkt	I21.9 → Grundleiden
1c		
2	COVID-19	U07.1

- (3) Verdachtsfälle: „an“ oder „mit“ COVID-19 Verstorbene ohne Virusnachweis in der Todesursachenstatistik

Ein medizinischer Term **mit** COVID-19 wird dann mit U07.2 kodiert, wenn aus der Formulierung ableitbar ist, dass entweder kein PCR-Test durchgeführt wurde oder das Ergebnis eines veranlassten PCR-Tests zum Zeitpunkt der Ausstellung der Todesbescheinigung nicht bekannt war. Beispielsweise wird bei den folgenden Nennungen auf der Todesbescheinigung mit dem Kode U07.2 kodiert:

- › Verdacht auf COVID-19
- › Infektion durch SARS-CoV-2, Virus nicht nachgewiesen
- › Verdacht auf COVID-19, kommt aus Pflegeheim mit Corona-Ausbruch und so weiter

Bei der in [Übersicht 4](#) dargestellten Todesbescheinigung wurde in der Kausalkette eine fragliche COVID-19-Erkrankung angegeben, die zu einem Grundleiden U07.2 führt. Hätte der Arzt oder die Ärztin nicht den Verdacht in Klammern auf dem Totenschein aufgeschrieben, so wäre die Pneumonie als Grundleiden in die Todesursachenstatistik eingegangen. Das Grundleiden ist jedoch COVID-19, da hierzu eine schlüssige Kausalkette vorliegt.

Übersicht 4

An COVID-19 verstorben ohne Virusnachweis:
Beispiel einer Todesbescheinigung (links)
mit multikausaler Kodierung nach ICD-10 (rechts)

79 Jahre, weiblich, natürlicher Tod		Multikausale ICD-10-Kodierung
1a	Respiratorische Globalinsuffizienz	R68.8
1b	Pneumonie (fragl. COVID-19)	J18.9, U07.2 → Grundleiden
1c		
2	Chronisch venöse Insuffizienz, pAVK	I87.2, I70.2

- (4) Fehlerhafte oder unvollständige Todesbescheinigungen

Eine rechtsmedizinische Publikation zeigt auf, dass mehr als 27 % der Todesbescheinigungen schwerwiegende Dokumentationsfehler enthalten (Zack und andere, 2017). Falls beispielsweise die Kausalkette in Teil 1 fehlerhaft ist, so wird nach den Vorgaben der WHO lediglich der plausible obere Teil für die Ermittlung

des Grundleidens herangezogen. Das folgende Beispiel weist daher I25.9 als Grundleiden aus, da COVID-19 aus medizinischer Sicht keine koronare Herzkrankheit verursachen kann. Dieser Fall geht als **mit** COVID-19 verstorben in die Todesursachenstatistik ein. [↗ Übersicht 5](#)

Übersicht 5

Unplausible Kausalkette eines fehlerhaft ausgefüllten Totenscheins

88 Jahre, weiblich, natürlicher Tod		Multikausale ICD-10-Kodierung
1a	Herzversagen	I46.9
1b	Koronare Herzkrankheit	I25.9 → Grundleiden
1c	COVID-19	U07.1
2	Diabetes mellitus Typ 2	E11.9

Daher ist davon auszugehen, dass bei einem relativ hohen Prozentsatz der Todesbescheinigungen die Unterscheidung zwischen **an** und **mit** COVID-19 verstorben aufgrund von Dokumentationsmängeln nicht sicher möglich ist.

Sterbefälle, die im Zusammenhang mit einer COVID-19-Erkrankung stehen, können ab dem Berichtsjahr 2021 noch genauer kodiert werden. Mit der Einführung weiterer ICD-Kodes durch die WHO stehen nun auch Kodes zur Verfügung, die beispielsweise die Spätfolgen von COVID-19 besser abbilden.

4

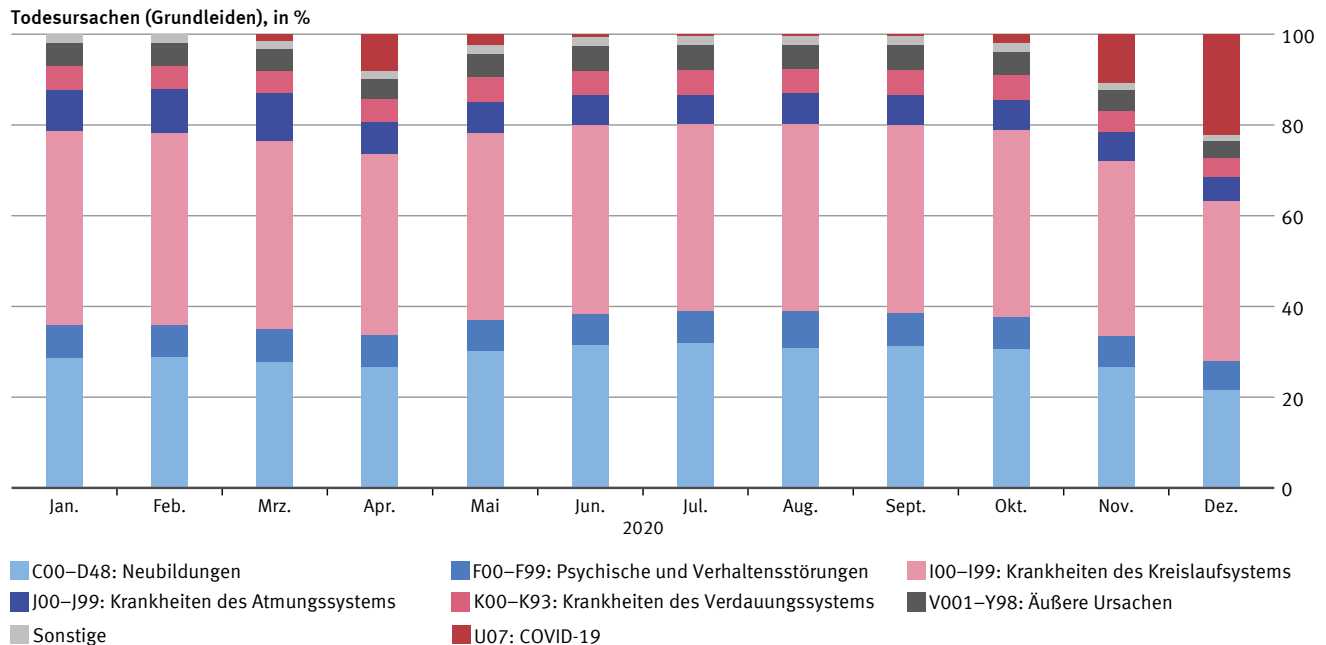
Ausgewählte Monatsergebnisse

4.1 Die wichtigsten Todesursachen- gruppen

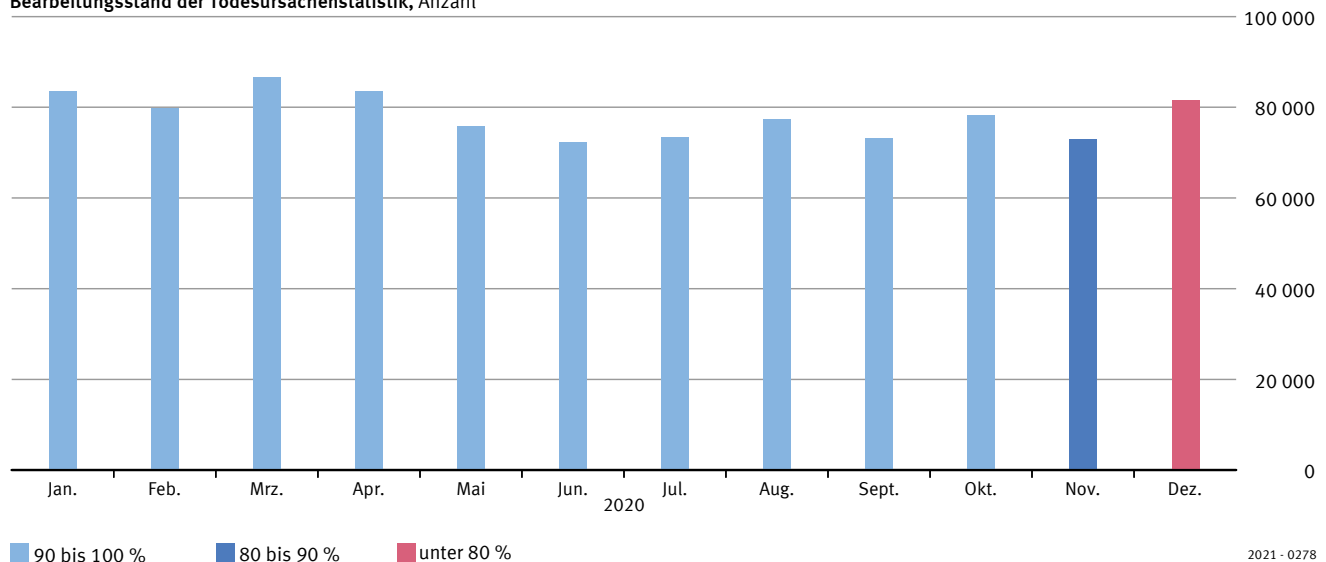
Die vorläufigen Monatsergebnisse der Todesursachenstatistik sind im oberen Teil der [↗ Grafik 1](#) dargestellt. Sie sind in Zusammenhang mit dem Vollständigkeitsgrad der Todesursachenstatistik in den jeweiligen Monaten zu betrachten – zu finden im unteren Teil der Grafik 1. Für die Monate Januar bis Oktober des Jahres 2020 liegt für jeweils über 98 % der Sterbefälle eine plausible Todesursache vor; die Qualität der vorläufigen Todesursachenstatistik ist für diese Monate am höchsten ein-

Grafik 1

Sterbefälle nach Todesursachen und Bearbeitungsstand der Todesursachenstatistik 2020
Stand: 15. Juli 2021



Bearbeitungsstand der Todesursachenstatistik, Anzahl



zustufen. Der Vollständigkeitsgrad der Todesursachenstatistik liegt im November bei 85 %, im Dezember bei 74 %. Demnach ist die Qualität der Todesursachenstatistik für diese Monate als etwas geringer einzuschätzen.

Die Darstellung der Todesursachen erfolgt hier in Anteilen an allen kodierten Sterbefällen, da bei den vorläu-

figen Monatsberichten der Todesursachenstatistik noch nicht alle Sterbefälle enthalten sind.

Nach den vorläufigen Ergebnissen waren im Jahr 2020 etwa 35 % der Sterbefälle auf Krankheiten des Kreislaufsystems (ICD-10-Klassifikation I00 bis I99) zurückzuführen, wobei die Anteile am Ende des Jahres 2020 mit

etwa 33 % im November und 31 % im Dezember etwas geringer waren. Die zweithäufigste Todesursache waren Krebserkrankungen (C00 bis D48) mit 25 %. Manche Todesursachen zeigten Schwankungen im Jahresverlauf. Bei den hier nachgewiesenen Todesursachengruppen war dies vor allem bei den Krankheiten des Atmungssystems (J00 bis J99) der Fall, die am Jahresanfang häufiger auftraten. Im Januar bis März machten diese Erkrankungen 8 bis 9 % der Sterbefälle aus, während sie in den darauffolgenden Monaten des ersten Halbjahres nur noch etwa 6 % umfassten.

Deutlich zu sehen sind die Auswirkungen der Corona-Pandemie: Nach dem ersten Auftreten von COVID-19-Sterbefällen im März 2020 zeigen sich Anteile von etwa 1 % im März, 7 % im April und 2 % im Mai. Die Zahl der Sterbefälle im Zusammenhang mit einer COVID-19-Infektion nahm in den Monaten Juni bis September 2020 wieder deutlich ab (Anteile unter 1 %), bevor im Oktober 2020 ein erneuter Anstieg mit der zweiten Welle erfolgte. Der Anteil der [an](#) COVID-19 Verstorbenen an allen Sterbefällen betrug im Oktober etwa 2 % und im November 9 %; er erreichte im Dezember mit 19 % sein Maximum für das Jahr 2020.¹³ In den Monaten der ersten und zweiten Welle mit einer erhöhten Zahl von Sterbefällen und deutlich erhöhten Anteilen von COVID-19

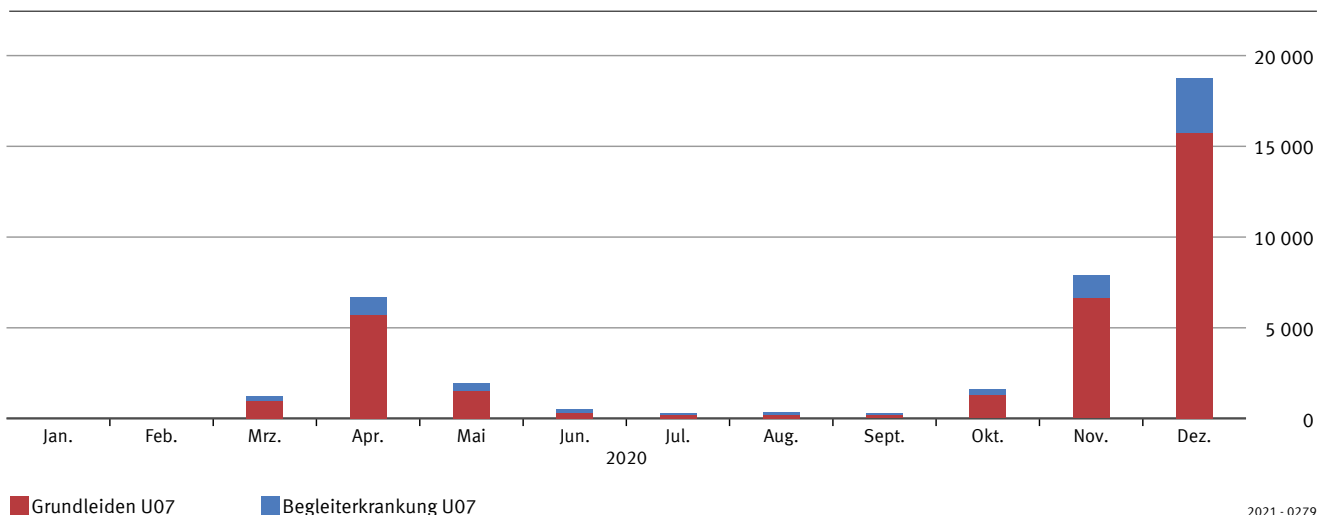
¹³ Es sei darauf hingewiesen, dass trotz des relativ hohen Vollständigkeitsgrads der Daten auf Bundesebene im November (85 %) und Dezember (74 %) manche Länder noch gar nicht oder in sehr geringem Umfang enthalten sind. Dies trifft zum Beispiel auf das stark von der Corona-Pandemie betroffene Land Bayern zu.

als Todesursache gingen in der relativen Betrachtung die Anteile der übrigen Todesursachengruppen zurück, nicht jedoch zwangsweise bei der Betrachtung der absoluten Fallzahlen.

4.2 COVID-19-Sterbefallzahlen in der monatlichen Todesursachenstatistik

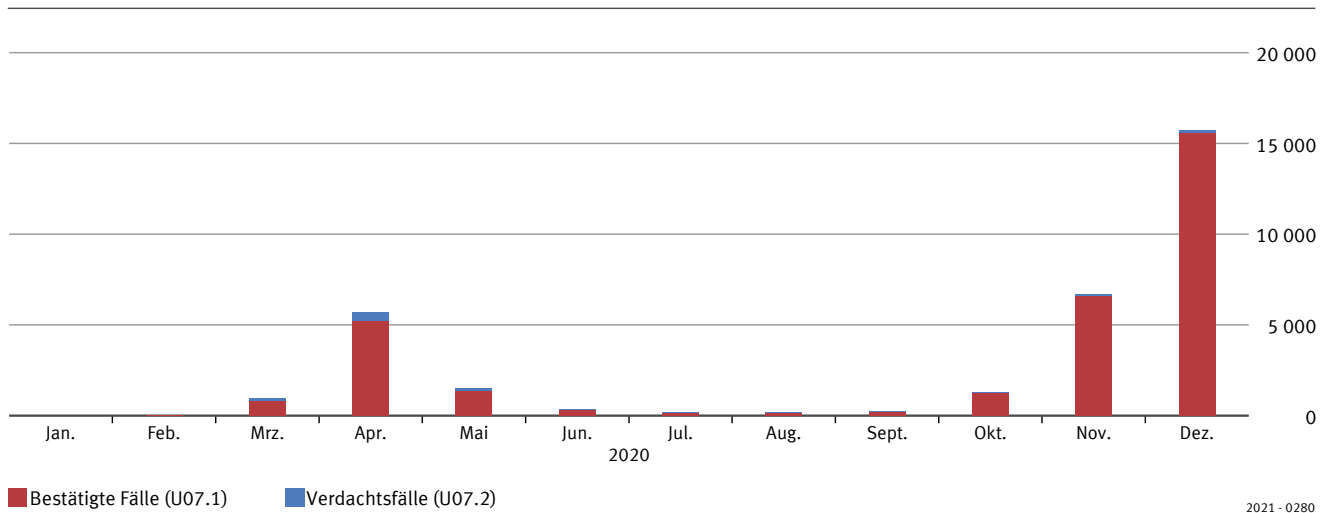
Während Grafik 1 das Auftreten der ersten und zweiten Corona-Welle bereits verdeutlichen konnte, zeigt [Grafik 2](#) COVID-19-Fälle in der Unterscheidung nach Grundleiden beziehungsweise Begleiterkrankung. Insgesamt starben demnach etwa 4 % aller Verstorbenen [an](#) COVID-19, etwa ein weiteres Prozent verstarb [mit](#) COVID-19, das heißt COVID-19 war eine Begleiterkrankung. Im Laufe des Jahres 2020 lässt sich feststellen, dass sich das Verhältnis von COVID-19 als Grundleiden zu jenen Fällen, in denen die Krankheit eine Begleiterkrankung darstellte, unterschiedlich entwickelte. Vor allem in Monaten, in welchen hohe COVID-19-Sterbefallzahlen vorlagen (April, Oktober bis Dezember 2020), war die Erkrankung deutlich häufiger auch Grundleiden (fünf bis sechs Mal so häufig wie Begleiterkrankung). Im Gegensatz dazu machten Sterbefälle mit COVID-19 als Begleiterkrankung in den Sommermonaten – also bei sehr geringen Fallzahlen – einen deutlich höheren Anteil aus.

Grafik 2
Sterbefälle nach COVID-19-Grundleiden und -Begleiterkrankung 2020



Grafik 3

Sterbefälle mit COVID-19 als Grundleiden 2020



Eine COVID-19-Erkrankung wird in der Todesursachenstatistik mit dem Code U07.1 oder U07.2 ausgewiesen, abhängig von den Angaben auf der Todesbescheinigung (siehe Abschnitt 2.3). [Grafik 3](#) zeigt, dass die meisten Verstorbenen mit Grundleiden COVID-19 auf den Code U07.1 entfallen (96 %; 31 679 Personen); Verdachtsfälle (U07.2) kamen nur in 4 % der Sterbefälle mit COVID-19 als Grundleiden vor. Bedingt durch die erhöhte Testkapazität im Jahresverlauf 2020 waren vor allem in den Monaten Oktober bis Dezember geringe Anteile an Verdachtsfällen zu verzeichnen. Während im vierten Quartal zwischen 1 und 4 % der [an](#) COVID-19 Verstorbenen sogenannte Verdachtsfälle (U07.2) waren, betrugen die Anteile der Verdachtsfälle in den ersten drei Quartalen zwischen 9 und 18 %.

durch die monatliche Todesursachenstatistik nachgewiesenen COVID-19-Sterbefälle als Untergrenze anzusehen. Durch die komplette multikausale Todesursachenkodierung aller Todesbescheinigungen mit einem Hinweis auf COVID-19 war es möglich, zusätzlich zum Ausweis des Grundleidens auch Fälle mit COVID-19 als Begleiterkrankung zu identifizieren. Dabei sollten in künftigen Forschungsarbeiten die Schwankungen im Verhältnis von Sterbefallzahlen mit COVID-19 als Grundleiden zu jenen mit COVID-19 als Begleiterkrankung untersucht werden.

Aufgrund der geringeren Fallzahlen und größeren Schwankungen in den Häufigkeiten sind die Monatsberichte der Todesursachenstatistik unter Vorbehalt zu interpretieren. Der Vollzähligkeitsgrad und die Fallzahl für die jeweilige Erkrankung(sgruppe) sind hier unbedingt zu berücksichtigen.

5

Diskussion

Fazit zur vorläufigen monatlichen Todesursachenstatistik

Die monatliche Berichterstattung der Todesursachenstatistik kann die Sterbefallzahlen durch COVID-19 während der Corona-Pandemie monatlich quantifizieren. Da die Bearbeitung der Todesursachenstatistik jedoch noch nicht in allen Fällen abgeschlossen ist, ist die Zahl der

Saisonalität in der Struktur der Todesursachen

Bei der Interpretation der monatlichen Daten ist zu beachten, dass sich die Sterbefallzahlen saisonal unterscheiden. In der Regel sind die Sterbefallzahlen am Anfang eines Jahres am höchsten und um die Mitte der zweiten Jahreshälfte am niedrigsten. Zu Beginn eines Jahres, wenn die Grippewellen ihren Höhepunkt haben, zeigt besonders die Sterblichkeit an Krankheiten des Atmungssystems deutliche Ausschläge nach oben, aber auch die Sterblichkeit an anderen Todesursachen ist erhöht.

Der beobachtete Rückgang an respiratorischer Mortalität im Lauf des Jahres 2020 scheint allerdings nicht allein auf saisonale Muster zurückzuführen sein, sondern ist vermutlich auch ein Nebeneffekt der Maßnahmen zur Bekämpfung der Corona-Pandemie. Durch Kontaktbeschränkungen, Hygienemaßnahmen und Maskenpflicht werden Übertragungswege von Infektionen unterbrochen (Buchholz und andere, 2020).

Verdachtsfälle bei COVID-19-Fällen

Auf den Todesbescheinigungen ist eine COVID-19-Erkrankung in manchen Fällen von den leichenschauenden Ärztinnen und Ärzten mit dem Zusatz versehen, dass es sich um einen Verdachtsfall handelt. Wenn möglich überprüfen die Statistischen Landesämter diesen Verdacht. Dabei erhärtete sich dieser für einen großen Teil der Fälle nicht, es lag keine COVID-19-Infektion vor. Für einen kleineren Teil der Verdachtsfälle bestätigte sich dieser, sie wurden geändert in COVID-19-Erkrankung mit Virusnachweis. Bei einem sehr geringen Teil blieb der Verdachtsfall auch nach Überprüfung als solcher bestehen. In den meisten Ländern war eine Überprüfung der Verdachtsangaben jedoch nicht möglich: Viele Gesundheitsämter sahen sich aufgrund der pandemiebedingten Zusatzbelastungen nicht in der Lage, mögliche Testergebnisse bei Ärztinnen und Ärzten, Kliniken oder Laboren zu recherchieren, Korrekturen vorzunehmen und Rückfragen der Landesämter zu beantworten. Nachfragen bei den Gesundheitsämtern zu den vorhandenen Verdachtsfällen haben in Niedersachsen⁴ ergeben, dass sich dort in etwa einem Viertel der Fälle der Verdacht auf eine COVID-19-Infektion bestätigt hat (24 %). Diese Fälle wurden auf den Kode U07.1 umgeschlüsselt. In 63 % der Verdachtsfälle hat sich eine COVID-19-Infektion nicht durch einen entsprechenden Test bestätigt, weitere 13 % blieben Verdachtsfälle (U07.2), da eine Klärung nicht möglich war. Die Tatsache, dass bei einem relativ hohen Anteil an Verdachtsfällen letztlich kein positiver PCR-Test vorlag, sollte bei der Interpretation der COVID-19-Sterbefallzahlen berücksichtigt werden.

Die monatliche Betrachtung der Todesursachen zeigte im Verlauf der Corona-Pandemie einen Rückgang der Angabe von COVID-19-Verdachtsfällen auf der Todes-

bescheinigung. Dies ist sicher auch auf die erhöhten Testkapazitäten und etablierten Testprozeduren zurückzuführen.

Exkurs: COVID-19-Sterbefälle in der Todesursachenstatistik und im Meldewesen nach dem Infektionsschutzgesetz

COVID-19-Sterbefälle werden über zwei Meldewege erfasst: zum einen über die amtliche Todesursachenstatistik, zum anderen über die Meldepflichten nach dem Infektionsschutzgesetz. Das Robert Koch-Institut (RKI) und die Landesgesundheitsbehörden veröffentlichen COVID-19-Sterbefallzahlen nach dem Infektionsschutzgesetz.

Die Unterschiede zwischen den beiden Dokumentationsformen führen dazu, dass die Fallzahlen der COVID-19-Sterbefälle in beiden Statistiken nicht identisch sind.

Erstens differiert die Datenbasis in beiden Statistiken. In die Todesursachenstatistik gehen alle COVID-19-Fälle ein, die auf der Todesbescheinigung einen entsprechenden Eintrag haben. Die Todesursachenstatistik unterscheidet nach nachgewiesenen (U07.1) und Verdachtsfällen (U07.2) sowie nach Grunderkrankung und Begleiterkrankung. In die unikausale Jahresstatistik der Todesursachenstatistik gehen nur die Fälle mit Grunderkrankung ein, während in den Monatsberichten der Todesursachenstatistik auch jene Sterbefälle nachrichtlich ausgewiesen werden, bei denen COVID-19 als Begleiterkrankung auftrat.

Gesundheitsämter melden an die zuständige Landesbehörde und das RKI COVID-19-Todesfälle gemäß § 6 Absatz 1 des Infektionsschutzgesetzes. Vom RKI werden nur diejenigen COVID-19-Todesfälle publiziert, bei denen ein laborbestätigter Nachweis von SARS-CoV-2 vorliegt und die infolge dieser Infektion verstorben sind (RKI, 2020). Die Zahl der COVID-19-Sterbefälle der Todesursachenstatistik wäre theoretisch dann deckungsgleich mit der Sterbefallzahl des RKI, wenn jedem U07.1-Sterbefall der Todesursachenstatistik ein positiver Labortest zugrunde liegen würde. Da die Todesursachenstatistik auf den Angaben der Ärztin beziehungsweise des Arztes beruht, werden all jene Fälle zu „nachgewiesenen“ Fällen, bei denen durch die Ärztin oder den Arzt eine COVID-19-Erkrankung auf der Todesbescheinigung vermerkt wurde. Ob diese Gewissheit auf Grundlage eines

⁴ Die exemplarische Auswertung basiert auf 71 Verdachtsfällen aus Niedersachsen.

positiven PCR-Tests besteht, ist aus der Todesbescheinigung nicht immer ersichtlich.

Zweitens ist in der Todesursachenstatistik die Unterscheidung zwischen den [an](#) und den [mit](#) COVID-19 Verstorbenen wesentlich. In der Todesursachenstatistik wird das Grundleiden (verstorben [an](#)) anhand aller Angaben auf der Todesbescheinigung auf Basis des Regelwerks der WHO bestimmt. Jedoch kann es insbesondere bei fehlerhaften oder unvollständigen Todesbescheinigungen schwierig sein, beide Gruppen verlässlich voneinander abzugrenzen. Bei den Statistiken nach dem Infektionsschutzgesetz findet eine solche Unterscheidung nicht immer statt.

Drittens können die Datenstände zu einem jeweiligen Stichtag in den beiden Dokumentationen unterschiedlich weit aufgearbeitet sein.

Viertens sollte bei einem Vergleich der Zahlen beachtet werden, ob die Ergebnisse nach Sterbedatum, Berichtsdatum (Todesursachenstatistik) oder Meldedatum (Infektionsschutzgesetz) ausgewiesen werden.

Aufgrund dieser Dokumentationsunterschiede kann es zwischen den beiden Statistiken somit verfahrenstechnisch bedingt zu Diskrepanzen bezüglich der COVID-19-Sterbefälle kommen. Ein Vergleich dieser beiden Statistiken sollte daher immer vor dem Hintergrund dieser differierenden Datengrundlagen und Meldewege erfolgen.

6

Ausblick

Die Monatsberichte der Todesursachenstatistik haben sich im hier vorgelegten ersten Konzept auf eine Ausweisung der Todesursachen in meist größeren Gruppen und COVID-19-Sterbefälle ohne weitere Differenzierung fokussiert. Eine Weiterentwicklung der Monatsberichte könnte eine weitere Differenzierung der Sterbefälle, zum Beispiel nach Alter und Geschlecht, umfassen. Zu berücksichtigen ist dabei ein Geheimhaltungsverfahren, das im Einklang mit der Geheimhaltung in der Jahresstatistik der Todesursachen- und Sterbefallstatistik steht.

Um die Ergebnisse valide einzuordnen, ist zudem ein Vergleich mit den Zahlen der Vorjahre nötig. Da die vorläufigen Monatsberichte der Todesursachenstatistik nicht vollständig sind, ist ein Vergleich der absoluten Zahlen schwierig; auch ein Vergleich von Anteilswerten ist mit Einschränkungen versehen. Für den Zeitvergleich von Anteilswerten bestimmter Krankheiten oder ICD-Gruppen ist somit ein Analysekonzept notwendig, welches dem erstmaligen Vorkommen bestimmter Erkrankungen (zum Beispiel COVID-19) sowie den jährlichen Schwankungen von Krankheiten (zum Beispiel Jahre mit ausgeprägter Grippewelle) Rechnung trägt.

Nach Abschluss des Berichtsjahrs kann die monatliche Berichterstattung mit der Jahresstatistik verglichen werden, das heißt die vorläufigen monatlichen Sterbefälle können denen aus der Todesursachenjahresstatistik, ausgewertet nach Monaten, gegenübergestellt werden. So kann nochmals evaluiert werden, ob unterjährig eine Selektivität besteht.

Die multikausale Analyse der Sterbefälle bietet weiteres Analysepotenzial, das zunächst für die Todesursachenjahresstatistik konzipiert und mit einem passenden Geheimhaltungsverfahren versehen werden muss. Erst in einem weiteren Schritt kann geprüft werden, welche multikausalen Analysen auch auf Monatsebene zu validen Ergebnissen führen. Multikausale Auswertungen dienen zum einen der Analyse, welche Begleiterkrankungen und Kombinationen an Vorerkrankungen bei der Todesursache COVID-19 relevant waren. Zum anderen können sie zeigen, an welcher Krankheit jene Personen verstorben sind, für die COVID-19 als Begleiterkrankung auf der Todesbescheinigung vermerkt war.

Um die Todesursachenstatistik noch aktueller zu veröffentlichen und größtmögliche Vergleichbarkeit im Zeitverlauf und zwischen den Ländern herzustellen ist jedoch eine elektronische Übermittlung aller Todesbescheinigungen, im Idealfall als elektronische Todesbescheinigung erfasst, unerlässlich. [UU](#)

LITERATURVERZEICHNIS

Buchholz, Udo/Buda, Silke/Prahm, Kerstin. *Abrupter Rückgang der Raten an Atemwegserkrankungen in der deutschen Bevölkerung*. In: Epidemiologisches Bulletin Nr. 16/2020, Seite 7 ff. DOI 10.25646/6636.2. [Zugriff am 2. Juli 2021]. Verfügbar unter: edoc.rki.de

DIMDI (Deutsches Institut für Medizinische Dokumentation und Information). *Internationale statistische Klassifikation der Krankheiten und verwandter Gesundheitsprobleme 10. Revision*. Band 2 Regelwerk. Version 2019. [Zugriff am 2. Juli 2021]. Verfügbar unter: www.dimdi.de

Eckert, Olaf. *Elektronische Kodierung von Todesbescheinigungen*. In: Bundesgesundheitsblatt. Ausgabe 62/2019, Seite 1468 ff. [Zugriff am 2. Juli 2021]. Verfügbar unter: link.springer.com

Fückel, Sebastian/Reh, Gerd/Vogel, Jeanette. *Sterbefälle im bisherigen Jahresverlauf. Auswertung vorläufiger Daten aus der Sterbefallstatistik und dem rheinland-pfälzischen Mortalitätsregister*. In: Statistische Monatshefte Rheinland-Pfalz. Ausgabe 10/2020, Seite 669 ff. [Zugriff am 2. Juli 2021]. Verfügbar unter: www.statistik.rlp.de

Möbius, Katrin. *Sterben die Menschen in Berlin und Brandenburg mit oder an Corona?* In: Zeitschrift für amtliche Statistik Berlin Brandenburg. Ausgabe 2/2020, Seite 10 ff. [Zugriff am 5. Juli 2021]. Verfügbar unter: www.statistik-berlin-brandenburg.de

RKI (Robert Koch-Institut). *Antworten auf häufig gestellte Fragen zum Coronavirus SARS-CoV-2 / Krankheit COVID-19: Wie werden Todesfälle erfasst?* 2020. [Zugriff am 2. Juli 2021]. Verfügbar unter: www.rki.de

Statistisches Bundesamt. *Todesursachenstatistik. Todesursachen in Deutschland. Qualitätsbericht*. 2020. [Zugriff am 2. Juli 2021]. Verfügbar unter: www.destatis.de

WHO (World Health Organization). *International Guidelines for Certification and Classification (Coding) of COVID-19 as Cause of Death*. 2020. [Zugriff am 2. Juli 2021]. Verfügbar unter: www.who.int

Zack, F./Kaden, A./Riepenhausen, S./Rentsch, D./Kegler, R./Büttner A. *Fehler bei der Ausstellung der Todesbescheinigung*. In: Rechtsmedizin. Jahrgang 27. Ausgabe 6/2017, Seite 516 ff.

Zur Nieden, Felix/Sommer, Bettina/Lüken, Stephan. *Sonderauswertung der Sterbefallzahlen 2020*. In: WISTA Wirtschaft und Statistik. Ausgabe 4/2020, Seite 38 ff.



Albert Braakmann

ist Diplom-Ökonom und leitet die Abteilung „Volkswirtschaftliche Gesamtrechnungen, Preise“ des Statistischen Bundesamtes. Er befasst sich unter anderem mit der administrativen Nutzung makroökonomischer Indikatoren sowie der Weiterentwicklung der Volkswirtschaftlichen Gesamtrechnungen. Er ist Mitglied der Advisory Expert Group on National Accounts (AEG-NA) bei den Vereinten Nationen.

DIGITALISIERUNG, GLOBALISIERUNG SOWIE WOHLFAHRT UND NACHHALTIGKEIT – KERNTHEMEN DER AKTUELLEN ÜBERARBEITUNG DER INTERNATIONALEN VGR-METHODIK

Albert Braakmann

📌 **Schlüsselwörter:** Volkswirtschaftliche Gesamtrechnungen – Revision – Digitalisierung – Globalisierung – Wohlfahrt – Nachhaltigkeit

ZUSAMMENFASSUNG

Der Startschuss für die Überarbeitung des derzeitigen weltweiten Systems Volkswirtschaftlicher Gesamtrechnungen, dem System of National Accounts (SNA) 2008, aus dem sich das rechtsverbindliche Europäische System Volkswirtschaftlicher Gesamtrechnungen (ESVG) ableitet, ist gefallen. Inzwischen liegen erste Überlegungen zur Weiterentwicklung des bisherigen SNA 2008 vor, die nachfolgend skizziert werden. Im Fokus dieses Beitrags stehen dabei vor allem Punkte im Rahmen der drei Kernthemen Digitalisierung, Globalisierung sowie Wohlfahrt und Nachhaltigkeit.

📌 **Keywords:** *national accounts – revision – digitalisation – globalisation – well-being – sustainability*

ABSTRACT

The revision of the current worldwide System of National Accounts (SNA) 2008, which is the basis of the binding European System of National and Regional Accounts (ESA), has been launched. First ideas regarding the further development of the current SNA 2008 are available and will be outlined below. The article focuses on items connected with the three core topics of digitalisation, globalisation as well as well-being and sustainability.

1

Hintergrund

Die Volkswirtschaftlichen Gesamtrechnungen (VGR) zielen darauf ab, eine quantitative Gesamtschau des wirtschaftlichen Geschehens zu liefern, und zwar für einen bestimmten Zeitraum (Jahre, Quartale) und in einem Wirtschaftsgebiet (Land, Region). Um einen möglichst hohen Grad an Vergleichbarkeit zu erreichen, wurden die Konzepte, Definitionen, Klassifikationen und Bewertungsregeln bereits Anfang der 1950er-Jahre auf internationaler Ebene festgelegt und in gewissen zeitlichen Abständen überarbeitet. Solche Überarbeitungen sind notwendig, um neue wirtschaftliche Entwicklungen aufzunehmen, die Konzepte zu überprüfen und zu aktualisieren, neue Methoden und Klassifikationen einzubauen, aber auch um eine Grundlage für veränderte Analyseanforderungen und eine evidenzbasierte Politikgestaltung zu schaffen.

Im März 2020 fiel der Startschuss für die Überarbeitung des derzeitigen weltweiten Systems Volkswirtschaftlicher Gesamtrechnungen, des System of National Accounts (SNA) 2008. Für die Mitgliedstaaten der Europäischen Union (EU) ist diese Revision höchst relevant, da sich das rechtsverbindliche Europäische System Volkswirtschaftlicher Gesamtrechnungen (ESVG) aus dem weltweiten SNA ableitet, um die internationale Vergleichbarkeit der VGR-Daten zu gewährleisten. Ein Grund für die Entwicklung einer europäischen VGR-Methodik ist der homogenere Stand der Wirtschaftsentwicklung in der EU. Hinzu kommen europaspezifische Anforderungen aufgrund der Verwendung von VGR-Daten für administrative Zwecke der EU. Dazu zählen insbesondere Informationen zur Finanzierung des EU-Haushalts, für die europäische Haushaltsüberwachung im Rahmen des Stabilitäts- und Wachstumspakts mit dem Verfahren bei einem übermäßigen Defizit oder für die Abgrenzung der Förderregionen im Zuge der europäischen Strukturförderung.

Der folgende Beitrag thematisiert die ersten Überlegungen der internationalen Diskussion, die im Fokus der aktuellen Überarbeitung des SNA 2008 stehen: Kapitel 2 beschreibt den Revisionsprozess und den auf internationaler Ebene vereinbarten Zeitplan für die Überarbeitung. Wie künftig die zunehmende Digitalisierung in den VGR

besser erfasst und sichtbar gemacht werden können, zeigt Kapitel 3 anhand von drei ausgewählten Beispielen. Die wachsenden Herausforderungen durch die wirtschaftliche Globalisierung behandelt Kapitel 4. Wie die Themen Wohlfahrt und Nachhaltigkeit künftig besser in den VGR berücksichtigt werden könnten, erläutert Kapitel 5. Abschließend wird ein Fazit zu den skizzierten Revisionspunkten aus den drei Themenbündeln gezogen.

2

Revisionsprozess und Zeitplan auf internationaler Ebene

Die Statistische Kommission der Vereinten Nationen (UNSC) hat auf ihrer Sitzung im März 2020 die Intersekretariatsarbeitsgruppe Volkswirtschaftliche Gesamtrechnungen (ISWGNA)¹ mit der Überarbeitung der weltweiten Methodik der VGR, dem SNA 2008, beauftragt. Gleichzeitig hat die Statistische Kommission darum gebeten, einen Zeitplan für die Überarbeitung bis 2025 vorzulegen. Ein Jahr später, im März 2021, wurde eine vorläufige Liste mit zu bearbeitenden Revisionspunkten angenommen. Diese Liste umfasst derzeit 55 Themen.² Für die einzelnen Revisionspunkte werden dabei zunächst konzeptionelle Leitlinien (Guidance Note) von hierzu eingerichteten Arbeitsteams (Task Teams) erstellt und diese, vor Änderung des SNA, von Mitgliedstaaten auf Machbarkeit getestet. Die drängendsten inhaltlichen Revisionspunkte, also die prioritären Themen der SNA-Forschungsagenda, wurden den drei Schwerpunktthemen Digitalisierung, Globalisierung sowie Wohlfahrt und Nachhaltigkeit zugeordnet. Außerdem wurde eine Arbeitsgruppe zur Verbesserung der Kommunikation der SNA-Methodik eingerichtet. Die weltweite Expertengruppe Volkswirtschaftliche Gesamtrechnungen (Advisory Expert Group on National Accounts – AEG-NA), in die 15 erfahrene Gesamtrechner aus verschiedenen UN-Mitgliedstaaten berufen sind, unterstützt die ISWGNA inhaltlich. Aus dem SNA 2025 wird sich das künftige ESGV ableiten, das für die EU-Mitgliedstaaten rechtsverbindlich sein wird.

1 Der Intersekretariatsarbeitsgruppe VGR gehören fünf große internationale Organisationen an: Vereinte Nationen (UN), Internationaler Währungsfonds, Organisation für wirtschaftliche Zusammenarbeit und Entwicklung (OECD), Weltbank sowie Europäische Kommission.

2 Sie ist als Anhang in diesem Beitrag enthalten.

Parallel zur Revision des SNA 2008 wird ebenfalls bis 2025 die Methodik der Zahlungsbilanzstatistik, das Balance of Payments Manual (BPM) Version 6, überarbeitet. International liegt hier die Zuständigkeit beim Zahlungsbilanzausschuss (BOPCOM) des Internationalen Währungsfonds, für Deutschland erstellt die Deutsche Bundesbank die Zahlungsbilanz. Um einen hohen Grad an Übereinstimmung bei diesen beiden wichtigen makroökonomischen Statistiken sicherzustellen, bearbeiten gemeinsame Arbeitsteams (Joint Task Teams) überlappende Themen. Möglicherweise schließen sich weitere Statistikbereiche dem Revisionszyklus 2020 bis 2025 an, wie etwa die Staatsfinanzstatistik.

Der von der ISWGNA vorgelegte und von der Statistischen Kommission der Vereinten Nationen angenommene Zeitplan erstreckt sich über den Zeitraum 2020 bis 2025. [↗ Grafik 1](#)

- › Beratung zu den einzelnen Themen durch (gemeinsame) Arbeitsteams, mit Erstellung von vorläufigen Leitlinien und Empfehlungen zur Änderung beziehungsweise Überarbeitung des SNA im Zeitraum 2020 bis 2023.¹³

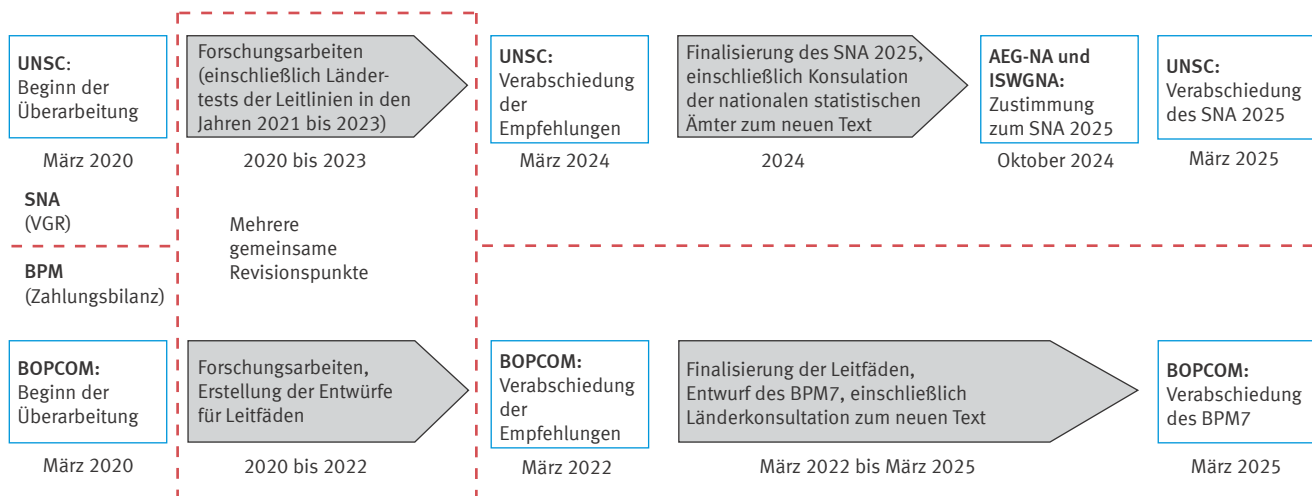
¹³ Die Liste der bei der Revision des SNA 2008 zu bearbeitenden Themen umfasst 55 einzelne Themen (Stand: März 2021), von denen allein 33 Punkte auf die drei Kernthemen entfallen (siehe Anhang).

- › Überprüfung der vorgeschlagenen Empfehlungen durch die weltweite Expertengruppe AEG-NA; sukzessiv, ebenfalls 2020 bis 2023.
- › Weltweite Konsultation der statistischen Ämter der Mitgliedstaaten zu den einzelnen Empfehlungen und der Möglichkeit, die Machbarkeit zu testen; sukzessiv, ebenfalls 2020 bis 2023.
- › Zusammenstellung und Konsolidierung aller Empfehlungen zu Änderungen beziehungsweise Klarstellung des SNA 2008 bis Mitte 2023 und anschließende weltweite Konsultation. Danach Billigung der Empfehlungen durch die Statistische Kommission im März 2024.
- › Erstellung des SNA 2025 bis Mai 2024 und nochmalige weltweite Konsultation. Zustimmung der weltweiten Expertengruppe AEG-NA und der ISWGNA bis Ende 2024.
- › Annahme des SNA 2025 durch die Statistische Kommission der Vereinten Nationen auf ihrer Sitzung Anfang 2025. Anschließend soll die nationale Implementierung erfolgen.

Die Nutzerinnen und Nutzer der VGR werden außerdem über Nutzerkonferenzen auf globaler und regionaler Ebene einbezogen. Sie können dort die vorgeschla-

Grafik 1

Zeitplan für die Revision des SNA 2008 und des BPM6



Erläuterung der Abkürzungen: SNA: System of National Accounts (System Volkswirtschaftlicher Gesamtrechnungen); VGR: Volkswirtschaftliche Gesamtrechnungen; BPM: Balance of Payments Manual (Methodik der Zahlungsbilanzstatistik); UNSC: Statistische Kommission der Vereinten Nationen; AEG-NA: weltweite Expertengruppe Volkswirtschaftliche Gesamtrechnungen; ISWGNA: Intersekretariatsarbeitsgruppe Volkswirtschaftliche Gesamtrechnungen; BOPCOM: Arbeitsgruppe Zahlungsbilanzstatistik

2021 - 0289

nen Empfehlungen beziehungsweise Änderungen kommentieren und beeinflussen, das heißt insbesondere auch zu den drei Kernthemen Digitalisierung, Globalisierung sowie Wohlfahrt und Nachhaltigkeit.

3

Digitalisierung

Warum steigt die Arbeitsproduktivität in entwickelten Volkswirtschaften trotz der zunehmenden Digitalisierung kaum an? Seit einigen Jahren gibt es zur Beantwortung dieser Frage lediglich Vermutungen. Eine für Statistiker wenig komfortable These ist dabei, dass ein Messfehler vorliegen könnte (Rürup, 2021; Hoffmann, 2017). Somit kann es kaum überraschen, dass Digitalisierungsthemen weit oben auf der SNA-Forschungsagenda stehen und zu einem der drei Kernthemen der Überarbeitung der internationalen VGR-Methodik gebündelt werden. Nachfolgend wird hier auf die folgenden ausgewählten Punkte näher eingegangen: Daten als Investitions- und Vermögensgut, kostenlose digitale Dienste, Visibilität der Digitalwirtschaft.⁴

3.1 Daten als Investitions- und Vermögensgut

Im derzeitigen SNA 2008 werden die Erstellung von Software und Datenbanken als Produktion angesehen. Die dabei erzeugte Wertschöpfung geht entstehungsseitig in das Bruttoinlandsprodukt (BIP) ein. Verwendungsseitig sind sie grundsätzlich als Investitionsgut zu buchen, sofern sie länger als ein Jahr für Produktionszwecke genutzt werden. Unklar bleibt, ob auch die Erstellung von Daten eine Produktion im Sinne des SNA darstellt. Somit besteht theoretisch die Möglichkeit, dass Erwerbstätige in der Datenproduktion arbeiten, deren Wertschöpfung aber nach den derzeitigen Regeln nicht beziehungsweise nicht zur Gänze erfasst wird. In der internationalen Diskussion wird vorgeschlagen, Daten explizit als produziertes Investitions- und Vermögensgut zu erfassen.

Die Einbeziehung von digitalen Daten in das BIP wirft zunächst einige konzeptionelle Fragen auf. So ist einmal eine Klarstellung erforderlich, was der Unterschied zwischen Daten einerseits und Datenbanken andererseits ist. Vor allem ist zu fragen, ab wann Daten als produziertes Gut angesehen werden können. In der internationalen Diskussion wird dabei zwischen „Daten“ und „beobachtbaren Erscheinungen“ unterschieden. Beobachtbare Erscheinungen werden als singuläre Ereignisse angesehen, sie werden nicht produziert und fallen somit nicht in die SNA-Produktionsgrenze. Der Kassenbeleg, den ein Kunde erhält, wird beispielsweise nicht als Produktion betrachtet. Erst der Zugriff auf verschiedene beobachtbare Erscheinungen, das heißt ihre Sammlung, Aufzeichnung oder Speicherung, könnte als Produktion angesehen werden. Mit einer solchen Definition würde auch vermieden, dass private Haushalte zu Datenproduzenten werden.

Eine andere konzeptionelle Frage betrifft die Bewertung des produzierten Investitions- und Vermögensgutes Daten. Relativ einfach könnte die Bewertung sein, wenn die Sammlung, Aufzeichnung und Speicherung von Daten an Dritte gegen Entgelt übertragen wurde. Werden dagegen die Daten intern im Unternehmen gesammelt, aufgezeichnet und gespeichert, bleibt häufig als Ersatzlösung die Bewertung über die Summe der Produktionskosten, also Vorleistungen, Arbeitnehmerentgelt, Abschreibungen, Gewinnzuschlag und Nettoproduktionssteuern. Eine schwierige Frage ist dabei, wie explizite und implizite Entgelte für den Zugriff auf beobachtbare Erscheinungen behandelt werden sollen. Dieser Zugriff kann über eine explizite Zahlung erfolgen, etwa bei Studierenden für medizinische Tests, oder durch Vergünstigungen, etwa bei Kunden beziehungsweise Rabattkarten von Händlern. Hier stellt sich die Frage, ob damit verbundene explizite beziehungsweise implizite Entgelte in die Bewertung des Investitions- und Vermögensgutes Daten eingehen sollten. Während bei expliziten Entgelten für den Zugriff derzeit eine Einbeziehung in den Wert der Daten favorisiert wird, ist dies bei impliziten Vergünstigungen strittig. Auch bleibt offen, wie solche Entgelte für das Bereitstellen der beobachtbaren Erscheinungen in den VGR gebucht werden sollten. Wenn die Bereitstellung von beobachtbaren Erscheinungen keine Produktion darstellt, scheidet eine Behandlung als Produktionsart aus. Derzeit in der Diskussion ist eine Buchung als Transfer oder als Pacht.

⁴ Weitere Punkte betreffen die Behandlung von Krypto-Geld oder künstlicher Intelligenz in den VGR.

3.2 Kostenlose digitale Produkte

In den letzten beiden Jahrzehnten wurden immer mehr digitale Produkte angeboten, die kostenlos genutzt werden können. Diese reichen von kostenlosen Informationen (zum Beispiel Wikipedia, Karten, Wetter oder Verkehrsverbindungen) über kostenlose Nachrichtendienste und soziale Medien bis hin zu Open-Source-Software. Da ihre Nutzung kostenfrei ist, fließen sie einerseits nicht ins Inlandsprodukt ein, ähnlich wie die Haushaltsproduktion. Andererseits stiften sie durchaus Nutzen, sodass gefordert werden könnte, solche kostenfreien Produkte schätzungsweise in das BIP einzubeziehen. Hier ist die internationale Diskussion noch gespalten:

- › Nach dem Bündelungsansatz sind die kostenfreien digitalen Produkte Teil eines Bündels von Produkten, die über Werbung oder Querfinanzierung bei kostengedeckten Produkten finanziert werden, sodass keine Zuschätzung zum BIP erfolgen muss.
- › Nach dem Naturaltauschansatz können die privaten Haushalte die kostenfreien digitalen Dienste nutzen als Ausgleich für den Zugriff auf ihre Daten. Hier wäre eine (modellmäßige) Schätzung des Werts der kostenfreien digitalen Dienste nötig. Diese könnte einerseits für eine Zuschätzung zum BIP herangezogen werden oder andererseits für eine Schätzung über den Umfang von kostenfreien digitalen Diensten in einer getrennten Zusatzrechnung.

Da es sich vor allem um private Haushalte handelt, ist es nicht zwingend, die Schätzung des Werts der kostenfreien digitalen Dienste zum Inlandsprodukt zu addieren. Ähnlich wie bei der Haushaltsproduktion könnte der geschätzte Wert der kostenfreien digitalen Dienste in einer Zusatzrechnung sichtbar gemacht werden. Einen solchen Ansatz stellen Brümmerhoff und Grömling (2017) vor und verweisen dabei unter anderem auf die Bewertung mit Preisen des Substitutionsprodukts.

3.3 Bessere Visibilität der Digitalwirtschaft

Aus gesamtwirtschaftlicher Sicht von Interesse ist der Ansatz des United States Bureau of Economic Analysis (BEA) zur Visibilität der Digitalwirtschaft. Grundlagen des BEA-Ansatzes sind Aufkommens- und Verwendungs-

tabellen, die im Hinblick auf die Digitalwirtschaft weiter untergliedert werden. So wurden spaltenweise zusätzliche Wirtschaftsbereiche gezeigt, die die Digitalwirtschaft repräsentieren. Zeilenweise wurden zusätzliche Digitalprodukte nachgewiesen, und zwar getrennt für solche innerhalb der SNA-Produktionsgrenze und solche außerhalb (wie kostenfreie digitale Dienste). Außerdem wird vorgeschlagen, für jede Produktseite die Art der Transaktion (digital bestellt, geliefert oder beides) zu unterteilen. Ein kritischer Punkt ist, dass weder die derzeitigen internationalen Klassifikationen der Wirtschaftszweige noch die der Güter eine solche Unterteilung vorsehen, was tendenziell die internationale Vergleichbarkeit limitiert. Möglicherweise ist der BEA-Ansatz ein erster wichtiger Schritt, der noch weiter zu entwickeln ist. So wird zum BEA-Ansatz angemerkt: *«Diese vom BEA verwendete Klassifikation führt jedoch zu einer Konzentration der digitalen Wertschöpfung in bestimmten Wirtschaftsbereichen. Die digitalen Wertschöpfungsanteile weiterer Produkte, Dienstleistungen und besonders von Prozessen bleiben in dieser Definition der digitalen Wirtschaft unberücksichtigt.»* (Fritsch/Lichtblau, 2021)

4

Globalisierung

Wirtschaftliche Globalisierung und multinationale Unternehmensgruppen sind zwar kein ganz neues Phänomen, stellen aber die VGR vor wachsende Herausforderungen. So liegt der Fokus der VGR generell auf der Erfassung von wirtschaftlichen Vorgängen in einem Wirtschaftsgebiet. Dagegen führt die zunehmende Bedeutung von multinationalen Unternehmensgruppen (MUG) zu steigenden außenwirtschaftlichen Transaktionen zwischen verbundenen Unternehmen. Um den Anteilseignern möglichst hohe Dividenden auszuschütten, können beispielsweise Wertschöpfungsanteile konzernintern zwischen Ländern verschoben werden, sodass die Steuerlast minimiert wird. Eine solche Gewinnverschiebung, etwa über Transferpreise, Darlehen oder Patentverlagerungen, hätte auch Auswirkungen auf das Inlandsprodukt, wie der sogenannte Irland-Fall von 2016 schlagartig verdeutlicht hat (Stapel-Weber/Verrinder, 2016). Aufgrund von Verlagerung von Patenten beziehungsweise geistigem Eigentum musste die Veränderungsrate des preisbereinigten irischen Bruttoinlandsprodukts 2015/2014

im Sommer 2016 von knapp 8 % auf gut 26 % korrigiert werden. Damit wird die Glaubwürdigkeit des Bruttoinlandsprodukts als zentralem Wirtschaftsindikator in Frage gestellt.

4.1 Produkte geistigen Eigentums und Zweckgesellschaften

Vor dem Hintergrund des erwähnten Irland-Falls wird international diskutiert, ob zusätzliche Regeln in den VGR für die Verortung von (selbsterstellten) Produkten des geistigen Eigentums die Aussagefähigkeit und Glaubwürdigkeit volkswirtschaftlicher Kenngrößen wie dem Inlandsprodukt stützen können. Eine erste Überlegung ist dabei, Produkte des geistigen Eigentums grundsätzlich bei der Muttergesellschaft zu buchen. Da multinationale Unternehmensgruppen zur Gewinnverschiebung häufig sogenannte Zweckgesellschaften einrichten, wird auch überlegt, (auslandskontrollierte) Zweckgesellschaften getrennt in jedem Land zu identifizieren und darzustellen. Darüber hinaus könnten Zweckgesellschaften auch mit ihrer jeweiligen ausländischen Muttergesellschaft konsolidiert werden, was aber eine Änderung des Inlandsprinzips bedeuten würde. Alternativ könnte das Inländerkonzept erweitert werden, zum Beispiel indem der inländischen Muttergesellschaft sämtliche wirtschaftliche Aktivitäten aller von ihr kontrollierten Einheiten zugerechnet werden. In diesem Zusammenhang wäre Kontrolle zu definieren. Hierbei wird zwischen direkter Kontrolle, das heißt über 50 % der Stimmrechte, und indirekter Kontrolle, das heißt direkt zwar weniger als 50 % der Stimmrechte, die sich aber etwa über verschachtelte Beteiligungen auf eine Mehrheit summieren, unterschieden.

4.2 Bewertung von Exporten und Importen

Sowohl in den VGR als auch in der Zahlungsbilanzstatistik sind die Exporte und Importe von Waren und Dienstleistungen jeweils frei Grenze des Ausfuhrlands zu bewerten (FOB: free-on-board). Dies ist aufgrund der Zollformalitäten bei der Ausfuhr bei Exporten generell wenig problematisch. Allerdings erfasst der inländische Zoll bei Einfuhren den Wert an der deutschen Grenze (CIF: cost, insurance, freight), sodass von der Statis-

tik auf den FOB-Wert an der Grenze des Ausfuhrlands zurückzurechnen ist (sogenannte CIF-FOB-Anpassung). Dieser aufwendige Rechenschritt könnte entfallen, würde stattdessen der Transaktions- beziehungsweise Rechnungspreis für die Bewertung von Exporten und Importen verwendet. Außerdem sei konzeptionell die FOB-FOB-Bewertung bei Exporten und Importen, so wird angeführt, nicht ganz vergleichbar mit der Bewertung der Produktion zu Herstellungspreisen. Allerdings hätte eine Bewertung von Einfuhr und Ausfuhr zu Rechnungspreisen auch Nachteile: Zunächst ist ein Bruch in der Zeitreihe zu befürchten, da für zurückliegende Jahre solche Daten nicht mehr erzeugt werden können. Außerdem kann der Rechnungspreis – je nach Vereinbarung – unterschiedliche Dienstleistungen und sogar Steuerzahlungen enthalten (zum Beispiel bei delivered duty paid) oder ebenso Gewinnverschiebungen durch Transferpreise. Damit wäre auch die Preisbereinigung mit größeren Risiken behaftet – mit direkter Auswirkung auf das preisbereinigte Bruttoinlandsprodukt. Vor diesem Hintergrund wird auch eine Bewertung ohne jegliche Transportleistung, das heißt auch ohne inländische Transportleistung wie im SNA 1968, diskutiert.

4.3 Getrennter Nachweis von multinationalen Unternehmensgruppen

Um die Bedeutung und den Einfluss von multinationalen Unternehmensgruppen für eine Volkswirtschaft besser beleuchten zu können, bietet sich eine getrennte Darstellung dieser Einheiten im Inland an. Hierzu werden international folgende Möglichkeiten diskutiert:

- › Separater Nachweis der Bruttowertschöpfung für inländische und auslandskontrollierte Produktionseinheiten auf makroökonomischer Ebene.
- › Getrennter Nachweis der auslandskontrollierten Produktionseinheiten sowohl in der detaillierten Aufkommens- als auch Verwendungstabelle.
- › Separater Nachweis der auslandskontrollierten Wirtschaftseinheiten über die gesamte Sequenz des Kontensystems, das heißt eine Unterteilung vor allem der Sektoren nichtfinanzielle und finanzielle Kapitalgesellschaften.

Dabei ist zu beachten, dass generell die amtliche Statistik einen nationalen Fokus hat, da meistens nur Wirt-

schaftseinheiten im Inland mit Auskunftspflicht belegt werden können. In der EU könnten Auskunftspflichten ebenfalls durch einen europäischen Rechtsakt auferlegt werden. Außerhalb der EU können Informationen zu multinationalen Unternehmensgruppen bisher nur auf freiwilliger Basis ausgetauscht werden, sofern überhaupt zulässig. Für einen internationalen Datenaustausch außerdem hilfreich wäre ein internationaler Wirtschaftsidentifikator.

5

Wohlfahrt/Nachhaltigkeit

Sowohl das Thema Wohlfahrt als auch das Thema Nachhaltigkeit haben in den letzten Jahren stark an Bedeutung gewonnen. Initiiert durch den sogenannten Stiglitz-Sen-Fitoussi-Bericht (2009) haben sowohl die OECD als auch das Statistische Amt der Europäischen Union (Eurostat) das Thema Wohlfahrt ganz nach oben auf ihre Agenda gesetzt. In Deutschland hat unter anderem die Enquete-Kommission Wachstum, Wohlfahrt, Lebensqualität des Deutschen Bundestags dazu ein Dashboard mit zehn Leitindikatoren vorgeschlagen (Enquete-Kommission Wachstum, Wohlfahrt, Lebensqualität, 2013, hier: Seite 237 f.). Das Thema Nachhaltigkeit und Umwelt erhielt vor allem durch den Bericht der sogenannten Brundtland-Kommission große Aufmerksamkeit. Dieser Bericht thematisierte die Verbindung von wirtschaftlicher, gesellschaftlicher und ökologischer Entwicklung (World Commission on Environment and Development, 1987). Die andauernde Diskussion über den globalen Klimawandel und die Erderwärmung dürfte dazu beigetragen haben, dass die Vereinten Nationen die Agenda für nachhaltige Entwicklung mit 17 Nachhaltigkeitszielen auf den Weg gebracht haben (United Nations, 2016). Daten zur deutschen Nachhaltigkeitsstrategie veröffentlicht das Statistische Bundesamt regelmäßig (Statistisches Bundesamt, 2021).

5.1 Wohlfahrt in den Volkswirtschaftlichen Gesamtrechnungen

Aus dem Blickwinkel der Wohlfahrt betrachtet blendet die VGR wichtige Aspekte aus, nämlich die Verteilung von Einkommen und Vermögen, die unbezahlte Haus-

arbeit (auch Ehrenamt), aber auch Bildung und Humankapital sowie Gesundheit und soziale Bedingungen.

- › Zur statistischen Erfassung der unbezahlten Hausarbeit verfügt das Statistische Bundesamt bereits über eine langjährige Erfahrung, sowohl was die etwa alle zehn Jahre durchgeführte Zeitverwendungserhebung als auch die Schätzung eines Werts für Hausarbeit durch die Volkswirtschaftlichen Gesamtrechnungen betrifft (Schwarz/Schwahn, 2016). Die konzeptionellen Vorschläge, etwa zur Abgrenzung der unbezahlten Arbeit anhand des Drittpersonenkriteriums oder zur Bewertung der Hausarbeit, sind größtenteils umsetzbar. Allerdings ist derzeit in der internationalen Diskussion offen, ob ein erweitertes BIP einschließlich unbezahlter Hausarbeit ermittelt werden soll. Damit könnten einerseits Verschiebungen über die Produktionsgrenze (zum Beispiel Do-it-yourself) und andererseits die Wohlfahrt besser beleuchtet werden.
- › Hinsichtlich der Ermittlung einer Einkommensverteilung im Rahmen der VGR ist zunächst darauf hinzuweisen, dass in den Konten der VGR der Fokus auf der funktionalen Einkommensverteilung liegt. Die international zusätzlich angestrebte personelle Einkommensverteilung nach Haushaltsgruppen erfordert zwingend individuelle Mikrodaten für einzelne Haushalte auf repräsentativer Basis. Aus dem Blickwinkel des Haushaltssektors sind sowohl die Privaten Organisationen als auch Personen in sogenannten institutionellen Haushalten (zum Beispiel Altersheimen) üblicherweise nicht in den Mikrodaten enthalten. Außerdem dürften bei privaten Haushalten erhobene Mikrodaten kaum Informationen zu bestimmten Einkommensgrößen der VGR bieten, wie Einkommen aus illegalen Aktivitäten, Schattenwirtschaft, unterstellte Einkommen aus selbstgenutztem Wohneigentum oder Unterstützungsleistungen zwischen Haushalten. Für diese Punkte sind Anpassungen erforderlich, sei es bei den VGR-Daten oder den Ergebnissen der Mikroerhebungen, damit es nicht zu einer Restdifferenz zwischen diesen beiden Datensätzen kommt. Verbleibt eine Restdifferenz zwischen den angepassten Einkommen laut VGR und der Summe der Einkommen aus Mikrodaten, ist ein modellmäßiges Abstimmverfahren nötig, das die Verteilungskenngrößen tendenziell umso stärker beeinflusst je größer sie ausfällt. Aus deutscher Sicht ist derzeit die Restdifferenz bei den Einkommen aus Unterneh-

menstätigkeit und Vermögen so groß, dass übliche Kenngrößen der Einkommensverteilung (80/20-Relation oder Gini-Koeffizient) davon merklich beeinflusst werden (Schwahn/Schwarz, 2012). Bei internationalen Vergleichen der Einkommensverteilung stellt sich außerdem die Frage, wie soziale Sachleistungen (Gesundheit, Bildung), die ja in den verschiedenen Ländern sehr unterschiedlich finanziert werden, in vergleichbarer Form berücksichtigt werden können (Schwahn/Schwarz, 2015).

Bei den Themen Gesundheit und soziale Bedingungen sowie Bildung und Humankapital steht die internationale Diskussion im Zuge der Überarbeitung des System of National Accounts 2008 noch am Anfang.

5.2 Nachhaltigkeit und Volkswirtschaftliche Gesamtrechnungen

Im Bruttoinlandsprodukt schlagen sich ökologische Kosten nach den derzeitigen Regeln nicht nieder. Vielmehr wirkt die Reparatur von Umweltschäden tendenziell sogar BIP-steigernd. Um die Beziehung zwischen wirtschaftlicher und ökologischer Entwicklung besser zu beleuchten, wurde 2012 das Handbuch zu den Umweltökonomischen Gesamtrechnungen als internationaler Statistikstandard angenommen. Vereinfacht gesagt wird hier die Entnahme natürlicher Inputs, wie Bodenschätze, Holz, Wasser oder Energie, aus der Umwelt sowie die Abgabe von „Resten“ in die Umwelt in den Blick genommen. Internationale Überlegungen gehen jetzt dahin, ökologische Aspekte stärker in die VGR zu integrieren und gegebenenfalls vom BIP abzuziehen. Dabei werden derzeit beispielsweise folgende Punkte diskutiert:

- › Den Abbau natürlicher Ressourcen als Produktionskosten behandeln. Damit würden diese Kosten vom Vermögensänderungskonto in das Produktionskonto umgebucht und vom BIP abgezogen. Unklar ist bisher, ob dies nur den Abbau selbst oder auch die Schädigung der Umwelt im Zusammenhang mit dem Abbau einschließen soll. Das traditionelle Argument, dass es in der statistischen Praxis schwierig sei, hier vergleichbare Schätzungen vorzunehmen, wird mit Hinweis auf die Vielzahl von Studien und Anleitungen in der Literatur als nicht mehr gegeben postuliert.
- › Einbeziehen auch von unkultivierten biologischen Ressourcen in das Volksvermögen, und zwar zumin-

dest in einer Zusatztafel. Dabei wird unter anderem vorgeschlagen, die Atmosphäre als Vermögensgut in den VGR zu behandeln. Damit ist auch beabsichtigt, die Buchung von Emissionszertifikaten in den VGR zu vereinfachen. Unklar bleibt allerdings beispielsweise, wie das Vermögensgut Atmosphäre erfasst und bewertet werden kann.

6

Fazit

Im Fokus der Überarbeitung der internationalen VGR-Methodik, dem SNA 2008, stehen die drei Themenbündel Digitalisierung, Globalisierung sowie Wohlfahrt und Nachhaltigkeit. Insgesamt umfasst die Liste der Revisionspunkte derzeit 55 Einzelthemen (Stand: März 2021), die in den nächsten beiden Jahren erforscht und bearbeitet sein sollen. Ziel ist, hierzu international gültige Empfehlungen für die Behandlung dieser Punkte im künftigen SNA 2025 zu entwickeln und festzuschreiben. Im vorliegenden Beitrag wurden erste Überlegungen der internationalen Diskussion zu ausgewählten Revisionspunkten aus den drei Themenbündeln skizziert. Beim Kernthema Digitalisierung ist die Ausweitung der Produktions- und Vermögensgrenze durch Einbeziehung von Daten als produziertem Gut angedacht. Außerdem wird die Behandlung von kostenlosen digitalen Diensten in den VGR sowie die Möglichkeit die Digitalwirtschaft besser sichtbar zu machen diskutiert. Beim Kernthema Globalisierung geht es zum einen um die Behandlung von Produkten geistigen Eigentums (zum Beispiel Patente) und von (ausländischen) Zweckgesellschaften in den VGR. Zum anderen wird die Bewertung von Exporten und Importen zu Rechnungspreisen (statt der bisherigen FOB-Bewertung) ebenso diskutiert wie die Möglichkeiten eines getrennten Nachweises von multinationalen Unternehmensgruppen in den VGR. Beim Themenbündel Wohlfahrt und Nachhaltigkeit sind die Einbeziehung von Hausarbeit sowie der Einkommensverteilung nach Haushaltsgruppen in das SNA 2025 wichtige Themen. Aus dem Blickwinkel der Nachhaltigkeit wird außerdem diskutiert, den Abbau von natürlichen Ressourcen, gegebenenfalls mit Einbeziehung der Umweltverschmutzung durch den Abbau, vom Inlandsprodukt als Produktionskosten abzuziehen. 

Liste der Revisionspunkte (Stand: März 2021)

Globalization

1. Treatment of multinational enterprise groups and intra-multinational enterprise group flows
2. Definition of special purpose entities, economic presence and residency
3. Valuation of imports and exports
4. Economic ownership of intellectual property products
5. Merchanting
6. Factoryless goods producers and recording of their transactions
7. Nationality concept/extension of intellectual property products on a nationality basis
8. Typology of global production arrangements
9. Analysing global value chains and trade in value added

Digitalization

10. Framework for a satellite account on the digital economy
11. Role of data and the System of National Accounts (SNA) asset boundary
12. Free digital products
13. Price and volume measurement of goods and services affected by digitalization
14. Cryptoassets

Wellbeing/Sustainability

15. Broader framework for well-being and sustainability
16. Distribution of household income, consumption, saving and wealth
17. Education and human capital, labour and productivity
18. Health and social conditions
19. Unpaid household service work
20. Refining economic ownership of natural resources
21. Accounting for biological resources
22. Accounting for a broader range of renewable natural resources
23. Stronger guidance on valuation of natural resources
24. Recording of losses
25. System of Environmental-Economic Accounting (SEEA) classifications
26. Recording of environmental taxes and subsidies
27. Distinction between recording a tax or a services transaction
28. Recording of emissions permits
29. Recording of provisions
30. Recording of depletion
31. Accounting for renewable energy resources

Liste der Revisionspunkte (Stand: März 2021)

noch: Wellbeing/Sustainability

- 32. Defining elements of the SEEA which should enter the broader well-being sustainability information system
- 33. Explaining the borderline between SNA and the SEEA

Cross-cutting and other issues

- 34. Relationship of SNA and International Accounting Standards Board
- 35. Statistical units
- 36. Trusts
- 37. Broadening the fixed asset boundary to include other intellectual property assets
- 38. Treatment of private-public partnerships
- 39. Harmonization of SNA and balance of payments
- 40. Islamic finance
- 41. Informal economy

Communication

- 42. Develop the design, content, functionality, administration and roll-out of the collaborative hub – Collaborative hub task team
- 43. Develop interactive and digitally enabled statistical standards for economic statistics and classifications – Digitized statistical standards task team
- 44. Review of terminology and branding – Terminology and branding task team
- 45. Development of a compilation taxonomy – Compilation taxonomy task team
- 46. Review of the assessment framework used to measure alignment with statistical standards – Assessment frameworks task team

Financial issues

- 47. More disaggregated definition of the financial sector and financial instruments
- 48. Asymmetric treatment of retained earnings between direct and portfolio investment and potential extension to domestic relationships
- 49. Reverse transactions
- 50. Financial derivatives by type
- 51. Treatment of credit default swaps
- 52. Non-bank financial intermediation
- 53. Impact of financial technology and other financial innovations
- 54. Valuation of debt securities at both market and nominal value
- 55. Valuation of loans (fair value)

LITERATURVERZEICHNIS

Brümmerhoff, Dieter/Grömling, Michael. *Haushaltssatellitensysteme – ein Baustein auch zur Erfassung der Digitalökonomie?* In: Wirtschaftswissenschaftliches Studium (WiSt). Heft 9/2017, Seite 26 ff.

Enquete-Kommission „Wachstum, Wohlfahrt, Lebensqualität – Wege zu nachhaltigem Wirtschaften und gesellschaftlichem Fortschritt in der Sozialen Marktwirtschaft“. *Schlussbericht*. Bundestags-Drucksache 17/13300 vom 3. Mai 2013. [Zugriff am 29. Juli 2021]. Verfügbar unter: dejure.org

Fritsch, Manuel/Lichtblau, Karl. *Die digitale Wirtschaft in Deutschland: Grenzen der Datenverfügbarkeit und erste Schätzungen*. In: Institut der deutschen Wirtschaft, Köln. IW-Trends 1/2021.

Hoffmann, Catherine. *Falsch gerechnet*. In: Süddeutsche Zeitung. 22. Juni 2017. [Zugriff am 13. Juli 2021]. Verfügbar unter: www.sueddeutsche.de

Rürup, Bert. *Wir können Wohlfahrtsveränderungen nicht adäquat messen*. In: Handelsblatt vom 6. Mai 2021. [Zugriff am 13. Juli 2021]. Verfügbar unter: www.handelsblatt.com

Stapel-Weber, Silke/Verrinder, John. *Globalisation at work in statistics – Questions arising from the ‘Irish case’*. In: Eurostat. Eureka, Heft 2/2016.

Statistisches Bundesamt. *Nachhaltige Entwicklung in Deutschland*. Indikatorenbericht 2021. Wiesbaden 2021.

Stiglitz, Joseph E./Sen, Amartya/Fitoussi, Jean-Paul. *Report by the Commission on the Measurement of Economic Performance and Social Progress*. Paris 2009.

Schwahn, Florian/Schwarz, Norbert. *Einkommensverteilung als Baustein der Wohlfahrtsmessung*. In: Wirtschaft und Statistik. Ausgabe 10/2012, Seite 829 ff.

Schwahn, Florian/Schwarz, Norbert. *Einkommenskonzepte zur Wohlfahrtsmessung: soziale Sachleistungen – ein Einkommensbestandteil?* In: WISTA Wirtschaft und Statistik. Ausgabe 3/2015, Seite 25 ff.

Schwarz, Norbert/Schwahn, Florian. *Entwicklung der unbezahlten Arbeit privater Haushalte*. In: WISTA Wirtschaft und Statistik. Ausgabe 2/2016, Seite 35 ff.

World Commission on Environment and Development. *Our Common Future*. (Brundtland Report). 1987.

United Nations. *Transforming our world: the 2030 Agenda for sustainable development*. Document A/RES/70/1 of 21 October 2015, 2016.

MACHINE LEARNING IN DER AMTLICHEN STATISTIK – ERGEBNISSE UND BEWERTUNG EINES INTERNATIONALEN PROJEKTS

Florian Dumpert

🔗 **Schlüsselwörter:** maschinelles Lernen – Klassifikation – Plausibilisierung – Imputation – Bildanalyse – Qualität

ZUSAMMENFASSUNG

Die High-Level-Group for the Modernization of Official Statistics der Wirtschaftskommission der Vereinten Nationen für Europa hatte für die Jahre 2019 und 2020 ein Machine-Learning-Projekt initiiert. Das Projekt ermöglichte die Durchführung von Pilotstudien, um den Mehrwert maschinellen Lernens in der amtlichen Statistik zu zeigen, sowie Arbeiten an einem Qualitätsrahmenwerk. Die Potenziale von maschinellem Lernen wurden dabei ebenso deutlich wie die Fallstricke und Hindernisse für dessen Einführung. Der Aufsatz beleuchtet die verschiedenen Arbeitsgebiete des Projekts und ordnet die Ergebnisse für die deutsche amtliche Statistik ein.

🔗 **Keywords:** machine learning – classification – data editing – imputation – imagery – quality

ABSTRACT

The machine learning project of the High-Level Group for the Modernization of Official Statistics of the United Nations Economic Commission for Europe facilitated pilot studies in 2019 and 2020 to demonstrate the added value of machine learning in official statistics, as well as work on a quality framework. The potential of machine learning became clear, as did the pitfalls and barriers to its adoption. The article highlights the different work areas of the project and discusses the results for German official statistics.



Dr. Florian Dumpert

ist als Referent im Statistischen Bundesamt mit Verfahren des maschinellen Lernens und der Imputation im Referat „Künstliche Intelligenz, Big Data“ befasst. Der Diplom-Mathematiker beschäftigt sich unter anderem mit methodischen Fragestellungen beim Einsatz dieser Verfahren und vertritt das Themengebiet in internationalen Projekten.

1

Einleitung

Das maschinelle Lernen (Machine Learning – ML) ist ein Teilgebiet des Forschungsfelds „Künstliche Intelligenz“ und gehört unbestritten zu den Zukunftstechnologien unserer Zeit. Davor kann und will sich die amtliche Statistik weltweit nicht verschließen. Einen frühen umfassenden Schritt dazu machte der Proof of Concept Machine Learning (Beck und andere, 2018). Methodisch am Puls der Zeit zu sein, ist ein Qualitätsaspekt amtlicher Statistik, der so auch Eingang in die Qualitätshandbücher fand. Beispielsweise schreibt das Qualitätshandbuch der Statistischen Ämter des Bundes und der Länder, dass „[d]ie statistischen Prozesse zur Erhebung, Aufbereitung und Verbreitung von Statistiken [...] internationalen Standards und Leitlinien in vollem Umfang genügen und zugleich dem aktuellen Stand der wissenschaftlichen Forschung entsprechen [sollen]. Dies gilt sowohl für die eingesetzte Methodik als auch für die angewendeten statistischen Verfahren.“ (Statistische Ämter des Bundes und der Länder, 2021, hier: Seite 19)

Um diesem Anspruch gerecht zu werden, sind Verfahren des maschinellen Lernens in den Werkzeugkasten der amtlichen Statistik als „neue“ statistische Methoden aufzunehmen. Wie immer bei der Einführung neuer Ansätze und Ideen ist – neben der internen Validierung – eine Orientierung erforderlich, einerseits an der Wissenschaft und dem aktuellen Forschungsstand, andererseits an vergleichbaren Einrichtungen, im vorliegenden Fall also an anderen nationalen statistischen Ämtern. Zum Austausch mit letzteren hat die High-Level-Group for the Modernization of Official Statistics (HLG-MOS) der Wirtschaftskommission der Vereinten Nationen für Europa (United Nations Economic Commission for Europe – UNECE) im Jahr 2019 ein Machine-Learning-Projekt ins Leben gerufen, das im folgenden Kapitel in seinen Grundzügen vorgestellt wird. Die weiteren Kapitel gehen auf verschiedene Spezialaspekte des Projekts ein; der Aufsatz schließt mit einer Zusammenfassung und einer Einordnung für das Statistische Bundesamt.

2

Das UNECE-HLG-MOS-Machine-Learning-Projekt

Die HLG-MOS beschreibt sich als „Gruppe engagierter Leitungen von nationalen statistischen Ämtern, die die Modernisierung statistischer Organisationen aktiv steuern“ (HLG-MOS, 2021); diese Gruppe hat das Machine-Learning-Projekt 2019 ins Leben gerufen. Die Herausforderung des Projekts bestand im Vorantreiben der Forschung, Entwicklung und Anwendung von Machine Learning in der amtlichen Statistik mit dem Ziel, dort einen Mehrwert zu schaffen.

Drei inhaltliche Arbeitspakete waren schließlich vorgesehen: Pilotstudien, Qualität und Integration. Das Arbeitspaket Pilotstudien umfasste dabei die Unterpunkte Klassifizierung und Kodierung, Plausibilisierung und Imputation sowie die Nutzung von Bilddaten (Imagery). Hinsichtlich seiner Durchführung gliederte sich das Projekt in zwei Arbeitsjahre (2019 und 2020). Während der Arbeitsjahre gab es gemeinsame (Präsenz-)Tagungen und Arbeiten in den Arbeitspaketen über Distanz. Insgesamt waren 124 Vertreterinnen und Vertreter aus 23 Staaten, 33 nationalen statistischen Ämtern und vier internationalen Organisationen vertreten.

Die Resultate des Projekts wurden in einer öffentlichen Online-Veranstaltung vorgestellt und diskutiert (UNECE, 2020a).

3

Klassifikation und Kodierung

Der Abschnitt „Classification and Coding“ (C&C) behandelte Fragestellungen zum Mehrwert von Machine Learning im Teilprozess 5.2 „Daten klassifizieren und kodieren“ des Geschäftsprozessmodells Amtliche Statistik (GMAS; Statistische Ämter des Bundes und der Länder, 2021). Beispielsweise werden in diesem Teilprozess Textantworten einem numerischen Kode zugeordnet. Die im Zuge des Projekts durchgeführten Pilotstudien befassten sich auch sämtlich mit dieser Fragestellung, allerdings in unterschiedlichen Kontexten: berufliche Tätigkeit (von Personen), wirtschaftliche Tätigkeit (von

Firmen), Unfallbeschreibung (bei Arbeitsunfällen), Produktbezeichnungen (bei Preisen), Stimmungserkennung (bei Twittereinträgen). Die zugehörigen Klassifikations-schemata führen in ihrer großen Mehrheit zu sogenannten Multi-Class-Problemen, das heißt zu Klassifikationen mit (meist deutlich) mehr als zwei möglichen Klassen. Lediglich die Klassifikation der Twittereinträge wurde als binäres Klassifikationsproblem behandelt (positive gegenüber negative Grundstimmung des Eintrags).

Im Unterschied zu bereits tabellarisch vorliegenden Daten bedürfen Textdaten vorbereitender Arbeiten, sogenanntem Preprocessing. Die Pilotstudien zeigten, dass ein nicht zu unterschätzender zeitlicher wie auch fachlicher Aufwand für diese vorbereitenden Arbeiten zu erbringen ist, um schließlich gute Klassifikationsergebnisse zu erzielen. Im Preprocessing wurden verschiedene Techniken eingesetzt und untersucht: Entfernung von Stop-Wörtern (zum Beispiel Artikel und Präpositionen), Stemming (das heißt die Reduktion von Wörtern auf ihren Wortstamm), Lemmatisierung (das heißt das Ersetzen von Wörtern durch über definierte Wörterbücher festgelegte Basisformen), das Ersetzen von Groß- durch Kleinbuchstaben, das Ersetzen von Umlauten und allgemein diakritischen Zeichen sowie die Bildung von n-Grammen auf Wort- und Zeichen-ebene, Bag-of-Words- und Bag-of-Features-Ansätze.¹

Diese vorbereitenden Arbeiten bereiten die textlichen Eingaben unter anderem derart auf, dass sie in Tabellenform vorliegen. Prinzipiell kann nun jedes Klassifikationsverfahren eingesetzt werden, um die Zuordnung der Texte zu den Klassen des Klassifikationssystems vorzunehmen. In den Pilotstudien wurden hierfür Random Forests, Logistische Regression, Nächste Nachbarn, Naive Bayes, XGBoost, neuronale Netze (Multi-Layer-Perceptron, Convolutional Neural Network, Recurrent Neural Network) und Support Vector Machines (SVMs)² sowie direkt das Programm FastText eingesetzt.

Als Gütemaße zur Evaluation wurden die für die Klassifikation üblichen Größen (Gesamtgenauigkeit, Sensitivität und Vorhersagewert sowie Kombinationen daraus) herangezogen. Der Umstand, dass einige Klassen deut-

lich schwächer in den Daten vertreten waren als andere, dass also ein Imbalanced-Data-Problem vorlag, wurde dabei berücksichtigt.

Die Details zu den einzelnen Pilotstudien sind in Sthamer (2020) und den dort zitierten Pilotstudienberichten zu finden.

Der Mehrwert für die amtliche Statistik im Falle von Textklassifikation besteht vorrangig darin, langwierige, zeit- und arbeitsintensive, bislang manuell durchzuführende Prozessschritte zu automatisieren. Bei der Bemessung des Mehrwerts ist neben der schnelleren Bearbeitung auch die Qualität der Ergebnisse zu beurteilen. Die Klassifikationsgüten sollten dabei mit den Güten verglichen werden, die auf Basis der bisherigen manuellen oder algorithmischen Arbeiten erzielt werden. Dabei stellt sich die Frage nach der wahren Klassenzugehörigkeit einer Texteingabe. Untersuchungen im Rahmen einer Pilotstudie haben ergeben, dass auch händische Klassifikationen durch Fachleute nicht einheitlich und mithin fehleranfällig sind. Die Ergebnisse des ML-Verfahrens nur mit den händisch erzielten Ergebnissen zu vergleichen, kann die Güteschätzung daher verzerren. In der genannten Pilotstudie wurden dabei zunächst viel Zeit und Arbeitskraft dafür investiert, einen fehlerfreien Trainings- und (vor allem) einen fehlerfreien Testdatensatz zu erzeugen. Gegen diesen konnten dann die automatisierten und die händisch erzeugten Ergebnisse gehalten werden. Dieser Vergleich zeigte den Mehrwert des automatisierten Verfahrens auch in der Dimension Genauigkeit auf (Measure, 2020, und darin genannte Quellen).

Statt einer Vollautomatisierung wurden in einer anderen Pilotstudie die Ergebnisse des ML-Verfahrens nur als Vorschlag unterbreitet, der oder die zuständige Beschäftigte konnte diesen Vorschlag annehmen oder überschreiben. Eine Mischung aus beiden Ansätzen ist ebenfalls denkbar und auch wünschenswert: Ist das ML-Verfahren hinreichend sicher in der Klassifikation, so wird diese übernommen. Unterhalb eines Schwellenwerts jedoch werden die Texteingaben einer Sachbearbeiterin oder einem Sachbearbeiter zur gegebenenfalls maschinell assistierten manuellen Kodierung übergeben.

1 Für einen ersten, praxisorientierten Überblick zu Methoden der Textklassifikation siehe Lane und andere (2019).

2 Für Details zu den hier und im Weiteren beschriebenen Methoden siehe James und andere (2013) sowie Goodfellow und andere (2016).

Einordnung für die deutsche amtliche Statistik

Auch in der deutschen amtlichen Statistik stellt sich immer wieder – so auch in aktuellen Arbeiten – die Frage nach einer effizienten Bearbeitung von Freitexteingaben. Dabei können die Erkenntnisse aus den C&C-Pilotstudien des Machine-Learning-Projekts bestätigt und ergänzt werden:

- › Gutes Datenmaterial zum Trainieren und Testen der ML-Verfahren ist essenziell, einerseits für die erfolgreiche Modellbildung und -testung, andererseits aber auch für den fairen Vergleich mit dem bisher eingesetzten (gegebenenfalls manuellen) Verfahren. Was „gutes Datenmaterial“ bedeutet, ist dabei problemabhängig, bezieht sich jedoch grundsätzlich auf die in den Daten enthaltene Information. Je mehr unterschiedliche Klassen beispielsweise voneinander unterschieden werden sollen, desto mehr sich angemessen auf die Klassen verteilende Trainingsdaten werden benötigt.
- › Zwar ist es häufig möglich, schnelle Ergebnisse auch ohne intensives Preprocessing zu erreichen, den Qualitätsanforderungen der amtlichen Statistik genügende Ergebnisse sind jedoch nur mit einem solchen zu erzielen.
- › Neuronale Netze erscheinen im Hinblick auf die Textklassifikation anderen ML-Verfahren überlegen, stellen aber auch erhöhte Anforderungen an die Hardware, insbesondere in Form spezieller Grafikkarten. Steht diese nicht zur Verfügung, ist – qualitätseinschränkend – auf andere ML-Verfahren auszuweichen.
- › Gerade bei der Arbeit mit internationalen Klassifikationssystemen ergeben sich drei bislang nicht geklärte methodische Fragen für die amtliche Statistik:
 - › Ist es zielführend, die hierarchische Gliederung vieler dieser Systeme zu nutzen, beispielsweise um mehrstufig Modelle zu trainieren? Falls ja: Wie sollte dies idealerweise geschehen?
 - › Wie ist sicherzustellen, dass auch Klassen, deren Anteil im Trainingsmaterial oder in der Grundgesamtheit im Verhältnis zu anderen Klassen sehr gering ist, adäquat vom ML-Verfahren berücksichtigt werden?
 - › Wie kann es gelingen, einen zukunftssicheren Bestand an Trainingsdaten auch in deutscher Sprache zu konzipieren, aufzubauen und aktuell zu halten?

Hinsichtlich der – kontextabhängig zu beantwortenden – Frage, ob ein ML-Verfahren eher Vorschläge unterbreiten oder eher vollautomatisieren sollte, erscheint auch eine Dreiteilung denkbar: 1. Sicher klassifizierte Texteingaben werden vollautomatisiert verarbeitet. 2. „Mittelsicher“ verarbeitete Eingaben werden automatisiert verarbeitet, verpflichtend jedoch stichprobenartigen Prüfungen größeren Ausmaßes durch Sachbearbeiterinnen und Sachbearbeiter unterzogen. 3. Unsicher klassifizierte Eingaben werden vollständig manuell klassifiziert. Für eine solche Dreiteilung ist es erforderlich, ein Maß für die Unsicherheit einer Klassifikation zu vereinbaren. Unabhängig davon sind im Nachgang und über die Perioden der Fachstatistik hinweg immer wieder stichprobenartige Überprüfungen aller (auch der sicher klassifizierten) Eingaben erforderlich, um

- › die Qualität der Klassifikation zu sichern,
- › das für künftiges Training herangezogene Datenmaterial zu validieren und
- › Veränderungen im Meldeverhalten oder der Struktur der Eingaben im Allgemeinen (zum Beispiel bei Einzug neuer, noch nicht standardisierter Berufsbezeichnungen in den Sprachgebrauch) wahrzunehmen.

Die hier dargelegten Einordnungen lassen sich analog auch auf die Kapitel 4 und 5 übertragen.

4

Plausibilisierung und Imputation

Der zweite Untersuchungsbereich „Editing and Imputation“ (E&I) beschäftigte sich mit dem Mehrwert von Machine Learning im Bereich von Plausibilisierung (PL) und Imputation, also vornehmlich mit dem GMAS-Teilprozess 5.4 „Daten plausibilisieren und imputieren“, jedoch mit engem Bezug zum GMAS-Teilprozess 5.3 „Daten prüfen und validieren“. Dabei waren verschiedene Konstellationen denkbar: Machine Learning könnte bestehende Verfahren ersetzen, verbessern oder ergänzen.

Die am Projekt Beteiligten verständigten sich darauf, die Untersuchungen auf zwei Fälle zu fokussieren: einerseits das (nicht rein regelbasierte) Erkennen fehlerhafter oder als solche verdächtiger Werte, andererseits das Abändern dieser und das Ersetzen fehlender Werte mit durchaus unterschiedlichen Zielen. Wie schon bei Klassifikation und Kodierung ist ein Ziel von Machine Learning im E&I-Prozess, bisher manuell durchgeführte Arbeiten zu unterstützen oder teilweise zu automatisieren und damit zu genaueren oder schnelleren Ergebnissen beizutragen.

Im Folgenden werden die Ergebnisse der Pilotstudien zusammengefasst, Details finden sich bei Dumpert (2020) und den dort zitierten Pilotstudienberichten. Traditionell werden regelbasierte Vergleiche von beobachteten Werten mit (schwachen oder starken) Plausibilitätsbedingungen, Verteilungsuntersuchungen (zum Beispiel zur Ausreißererkennung) und Vergleiche mit externen und/oder früheren Datensätzen angewendet. Dabei kommen mehrere Schritte zum Tragen, bei denen sowohl die Automatisierung (durch Plausibilitätsregeln) als auch die Fachleute (durch interaktives Bearbeiten) eine wichtige Rolle bei der Erkennung problematischer Daten spielen. Eine vollständige Automatisierung sollte nicht das unbedingte Ziel des Einsatzes von Machine Learning bei der Plausibilisierung sein. Ebenso wie beim Klassifizieren und Kodieren ist eine zumindest stichprobenartige Überprüfung (auch von vermeintlich sehr zuverlässig erkannten Fehlern im Datenbestand) angeraten. Die Pilotstudien zur Plausibilisierung ergaben, dass ein Lernen von Mustern aus früheren Fehlererkennungsrunden möglich ist. Liegen also Trainingsdaten vor, die plausible und unplausible Werte, die als solche auch gekennzeichnet sind, enthalten, so kann daraus ein ML-Verfahren ein Modell erlernen und neu ankommende Daten hinsichtlich der Plausibilität klassifizieren. Eine daraus prinzipiell mögliche Ableitung von PL-Regeln leidet aber daran, dass eine gute Performanz im Sinne einer guten Detektion fehlerhafter Werte nur möglich ist, wenn die Regeln sehr komplex sind, also sehr viele Fallunterscheidungen enthalten. Solche Regeln wiederum bieten keinen unbedingten Mehrwert. Die Durchführung der Fehlererkennung wurde in der Pilotstudie durch Einsatz des ML-Verfahrens deutlich beschleunigt und lieferte im Vergleich zur manuellen Fehlerdetektion konsistentere Resultate.

Hinsichtlich der Imputation lieferten die Pilotstudien die Einsicht, dass Machine Learning häufig, aber nicht immer plausible Einsetzungen vornimmt. Auch kann es durch das Überspringen einiger Preprocessing-Schritte (zum Beispiel Berücksichtigung von Korrelationen, statistische Transformationen von Variablen und so weiter) zu einer schnelleren Statistikproduktion beitragen. Dennoch erfordert maschinelles Lernen viele einschlägige Simulationen und mithin eine genaue Beschäftigung mit den Daten. Außerdem erlaubt Machine Learning, die Anzahl menschlicher Eingriffe zu reduzieren (wenn es beispielsweise die Variablenauswahl für das Imputationsmodell automatisch vornimmt).

Zum Einsatz kamen beim Plausibilisieren und Imputieren von Daten die folgenden ML-Verfahren: Klassifikations- und Regressionsbäume, Random Forests, neuronale Netze, (regularisierte) lineare Modelle, Nächste Nachbarn, bayesianische Netze, SVMs und Kombinationen daraus.

Einordnung für die deutsche amtliche Statistik

Es gibt kaum Möglichkeiten, schnelle Erfolge mit dem Machine Learning im Bereich Plausibilisierung und Imputation zu erzielen; die sorgfältige methodische Arbeit ist hier von besonders großer Bedeutung. Ein anderer, bei der Diskussion des Projekts aufgeworfener, aus Zeitgründen jedoch bislang nicht näher untersuchter Aspekt der Plausibilisierung ist die Frage, wie die Sachbearbeiterinnen und Sachbearbeiter bei manueller Plausibilisierung zu ihren Einsetzungen kommen. Das mag in manchen Fällen offensichtlich sein, in anderen Fällen wiederum stammt das nötige Wissen aus Zeitungsartikeln, Branchenreports, von Unternehmenswebsites und ähnlichen Quellen. Die Unterstützung der manuellen Plausibilisierung mithilfe solcher Informationen setzt einerseits ihre Beschaffung, andererseits ihre Verarbeitung voraus. Hier kann Machine Learning gegebenenfalls hilfreich sein.

5

Bildanalyse

Mit zunehmender Verfügbarkeit von Fernerkundungsdaten stellt sich auch für die amtliche Statistik die Frage nach deren Nutzung. In diesem Zusammenhang auftretende Fragen und Probleme wurden in den Pilotstudien zur Nutzung von Bilddaten (Imagery) erörtert. Ähnlich wie bei den Untersuchungsbereichen Klassifikation und Kodierung sowie Plausibilisierung und Imputation ist ein Ziel von Machine Learning im Kontext von Fernerkundungsdaten eine automatisierte Verarbeitung der in den Daten enthaltenen Informationen. Um welche Informationen es sich handelt, ist kontextabhängig. In den Pilotstudien handelte es sich um die Nutzung landwirtschaftlicher Flächen, um die Erfassung von Armut oder von Siedlungswachstum oder auch um ganz konkrete Fragestellungen im Zuge der Aktualisierung von Adressregistern. Die Arbeiten sind den GMAS-Teilprozessen 5.2 „Daten klassifizieren und kodieren“ und 5.5 „Neue Merkmale und Einheiten ableiten“ zuzuordnen.

Die in den verschiedenen Pilotstudien genutzten Fernerkundungsdaten hatten Auflösungen von 23 cm bis 30 m. Der Prozess, die Daten zu labeln, das heißt dem Bild die entsprechende Klassenzugehörigkeit anzufügen und somit Trainings- und Testdaten erst zu generieren, war jeweils zeit- und arbeitsintensiv. Es standen in der Regel auch noch keine vorgelabelten Bilddaten zur Verfügung, sodass nicht auf bereits vorhandene Bestände aufgebaut werden konnte. Bei der Erstellung der Trainings- und Testdaten war zu beachten, dass verschiedene Situationen (Stadt/Land, gutes Wetter/schlechtes Wetter, Regionalspezifika in Bau und landwirtschaftlicher Nutzung und Ähnliches) hinreichend repräsentiert sind. Die eingesetzte Anzahl der gelabelten Bilder betrug dabei das 1 000- bis 20 000-Fache der unterschiedenen Klassen. Zusätzlich zu den reinen unmittelbar im Bild enthaltenen Daten wurden noch weitere Informationen als erklärende Variablen eingesetzt, beispielsweise die Geokoordinaten des aufgenommenen Objekts oder dessen Lage über dem Meeresspiegel.

Wichtige Schritte in der Datenaufbereitung umfassen neben der technischen Anpassung zur weiteren Bearbeitung auch die Variation des vorhandenen Materials (Drehungen, Spiegelungen, Skalierungen), um einerseits

mehr Trainingsmaterial zu erhalten, andererseits aber auch der Gefahr einer Überanpassung an das Originalmaterial entgegenzuwirken.

Außerdem spielt die Variablenselektion bei der Bildanalyse eine herausgehobene Rolle, um die Bilder mit der vorhandenen Hardware in vertretbarer Laufzeit im Lernprozess auswertbar zu machen.

Als ML-Verfahren kamen Convolutional Neural Networks sowie Random Forests, SVMs und Extreme Randomized Trees auf gewöhnlichen Rechenkernen und Grafikkarten zum Einsatz. Sofern auf der vorhandenen Hardware berechenbar, erwiesen sich meist die Convolutional Neural Networks als überlegene Verfahren.

Alle durchgeführten Pilotstudien wurden positiv evaluiert und sind zum Teil bereits in die Produktion überführt. Der Mehrwert von Machine Learning bestand letztlich darin, dass die Fernerkundungsdaten überhaupt nutzbar gemacht wurden und somit die Statistikproduktion bereichern konnten, sei es durch zusätzliche Informationen oder die Entlastung der Sachbearbeiterinnen und Sachbearbeiter in sicher klassifizierten Fällen (Coronado/Juárez, 2020).

Einordnung für die deutsche amtliche Statistik

Die deutsche amtliche Statistik greift die Idee, Fernerkundungsdaten für die amtliche Statistikproduktion zu nutzen, immer wieder auf (Arnold/Kleine, 2017; Statistisches Bundesamt, 2019). Auch im Zuge der Einführung des Registerzensus (Modul Gebäude und Wohnungen) sollen gemeinsam mit Kooperationspartnern Verfahren entwickelt werden, mit denen Fernerkundungsdaten zur Validierung von Daten aus Registern eingesetzt werden können. Gemeinsam mit den im Projekt durchgeführten Pilotstudien werden Potenzial und praktischer Nutzen von Fernerkundungsdaten offensichtlich. Neben hinreichend verlässlichen und möglichst alle auch künftigen Aufnahmen abdeckenden Trainings- und Testdaten ist aber insbesondere für die Bildverarbeitung eine ausreichend dimensionierte IT-Ausstattung erforderlich.

6

Qualität

Neben Pilotstudien, die den Mehrwert von Machine Learning in der amtlichen Statistik prinzipiell untersuchen sollten, befasste sich ein weiteres Arbeitspaket des Machine-Learning-Projekts mit Fragen der Qualität. Die Arbeiten in diesem Arbeitspaket orientierten sich einerseits an bestehenden Qualitätsrahmen der amtlichen Statistik im Allgemeinen, beispielsweise dem European Statistics Code of Practice (Eurostat, 2017), präzisierten und ergänzten diese. Andererseits wurde nicht ausschließlich Machine Learning betrachtet, sondern ein Qualitätsrahmen für alle statistischen Algorithmen, alle statistischen (maschinellen Lern)Verfahren, entworfen, eine Mischung aus Handlungsanweisung und Rahmengogebung. Der Qualitätsrahmen „Quality Framework for Statistical Algorithms (QF4SA)“ (Yung und andere, 2020) umfasst fünf Dimensionen.

Ein statistischer Algorithmus (als Umschreibung für statistische Verfahren, seien sie „klassisch“ oder im Sinne des maschinellen Lernens eingesetzt oder verstanden) hat in diesen fünf Dimensionen die Tauglichkeit für den Einsatz in der amtlichen Statistik unter Beweis zu stellen:

1. **Erklärbarkeit** ist definiert als die Möglichkeit zu verstehen, welche Zusammenhänge der Algorithmus nutzt, um Vorhersagen zu treffen oder Analysen auszuweisen. Es geht also darum, den gegebenenfalls nur lokalen Zusammenhang zwischen Eingabe- und Ausgabevariablen (sofern vorhanden) darlegen zu können.
2. **Genauigkeit** ist definiert als der Grad eines statistischen Outputs, zu dem dieser in der Lage ist, das zu messende Phänomen korrekt zu beschreiben. Es geht also letztlich um den geeignet zu messenden Abstand zwischen Schätzung und wahren Wert.
3. **Reproduzierbarkeit** ist im Basisniveau definiert als die Fähigkeit, gleiche Ergebnisse zu erzielen, sofern mit den gleichen Daten und dem gleichen Algorithmus gearbeitet wird. Im höheren Niveau ist darunter zu verstehen, dass verschiedene Zufallsstichproben aus einer Grundgesamtheit bei gleichem Algorithmus regelmäßig zu im Wesentlichen gleichen Ergebnissen führen sollen.

4. **Rechtzeitigkeit** ist definiert als die Fähigkeit, innerhalb der geforderten Zeit den Algorithmus konzipieren, trainieren und anwenden zu können. Dabei können gegebenenfalls die Rechtzeitigkeit von Konzeption und Training und die Rechtzeitigkeit der Anwendung getrennt voneinander betrachtet werden.
5. **Wirtschaftlichkeit** schließlich ist definiert als Genauigkeit oder Schnelligkeit oder Erklärbarkeit je Kosteneinheit. Dabei enthalten die Kosten eines Algorithmus fixe Kosten (zum Beispiel für Personal und IT-Infrastruktur) und variable Kosten (zum Beispiel für Schulungsmaßnahmen, das Monitoring des Algorithmus oder gegebenenfalls notwendiges Nachtrainieren).

Einordnung für die deutsche amtliche Statistik

Die qualitätsgesicherte Einführung von Machine Learning ist eine wichtige Aufgabe für das Statistische Bundesamt und den Statistischen Verbund³. Der Begriff der Qualität von Machine Learning ist valide zu fassen und weiterzuentwickeln. Dazu werden – basierend auf Vorwissen und Vorerfahrungen – im Austausch im Statistischen Verbund und im Kontakt mit der Wissenschaft bereits verschiedene Anstrengungen unternommen. Der Qualitätsrahmen „Quality Framework for Statistical Algorithms“ wird zu dieser Debatte in der deutschen amtlichen Statistik einen wichtigen Beitrag leisten.

7

Integration

In diesem Arbeitspaket ging es um die Frage der Integration von Machine Learning in die amtliche Statistik. Schwierigkeiten konnten in allen beteiligten nationalen statistischen Ämtern beobachtet werden. Es wurde herausgearbeitet, dass auf organisatorischer Seite die Koordination zwischen den internen Stakeholdern sowie Widerstände interner Stakeholder die größten Hindernisse für die Einführung von Machine Learning darstellen, gefolgt von Unklarheit über Verantwortlich-

3 Den Statistischen Verbund bilden die Statistischen Ämter des Bundes und der Länder.

keiten in einem Projekt. Auf technischer Seite bestehen Schwierigkeiten darin, geeignetes Personal für diese Aufgaben zu finden, geeignete Hardware nutzen zu können sowie geeignetes Trainings- und Testmaterial zur Verfügung zu haben. Als förderlich wurden hingegen die Zusammenarbeit mit anderen statistischen Ämtern, die Kooperation mit der Wissenschaft, interne und externe Schulungsmaßnahmen sowie die klare Festlegung von Rollen und Verantwortlichkeiten in den Ämtern genannt.

Auf die Frage, wo Machine Learning organisatorisch verortet ist, gab es keine einheitliche Antwort. Einige statistische Ämter sehen Machine Learning im Bereich der Methodik, andere in eigenen Exzellenzzentren. In weiteren Fällen tritt Machine Learning rein dezentral oder explizit multidisziplinär in Erscheinung.

Bei der Frage schließlich, wie sich der Mehrwert von Machine Learning in der amtlichen Statistik zeigen kann, wurden drei Aspekte herausgearbeitet:

1. Es bedarf eines fairen Vergleichs zwischen bisherigem Vorgehen und Machine Learning – fair hinsichtlich der Testdaten, des Arbeits- und Zeitaufwands und der statistischen Qualitätsindikatoren.
2. Machine Learning muss Bestehendes nicht vollständig ersetzen. Auch die Unterstützung oder Ergänzung der bisherigen Vorgehensweise stellt einen Mehrwert dar.
3. Der Mehrwert von Machine Learning zeigt sich nicht zuletzt beim Einsatz für Aufgaben, die sonst gar nicht zu bearbeiten wären. Darunter fallen beispielsweise Aufgaben im Bereich der Bilderkennung.

Einordnung für die deutsche amtliche Statistik

Die aufgezeigten Schwierigkeiten sind auch in Deutschland bekannt. Engpässe zeigen sich regelmäßig bei gutem Trainings- und Testdatenmaterial sowie bei der vorhandenen IT-Infrastruktur. Zwar stehen für die Programmiersprachen R und Python Serverlösungen zur Verfügung, ein gewinnbringender Einsatz von neuronalen Netzen ist auf diesen aber aktuell nicht möglich. Darüber hinaus ist bereits jetzt zu beobachten, dass die Kapazitäten der vorhandenen Hardware häufig nicht ausreichen, um mit – für Machine Learning üblichen –

großen Datensätzen valide methodische Untersuchungen durchzuführen. Diese sind jedoch unbedingte Voraussetzung für einen qualitätsgesicherten Einsatz von Machine Learning im späteren Produktivbetrieb.

Hinsichtlich der im Projekt genannten begünstigenden Faktoren ist das Statistische Bundesamt auf einem guten Weg. Es gibt Angebote für Schulungsveranstaltungen zu Machine Learning und die Kooperation mit der Wissenschaft in diesem Bereich wird etabliert. Auch der Austausch mit anderen statistischen Ämtern ist über internationale Projekte (wie diesem) und über den Arbeitskreis Maschinelles Lernen (für den Statistischen Verbund) gewährleistet.

8

Zusammenfassung des Projekts

Als konsolidierte Fassung und Bewertung der Teilergebnisse legt der Abschlussbericht (Julien, 2020) die wesentlichen Aspekte dar, die für die Akzeptanz von Machine Learning in der amtlichen Statistik erforderlich sind: Der Einsatz von Machine Learning muss

- › sich an den Bedürfnissen der Ämter als Datenproduzenten ausrichten,
- › durch ein Qualitätsrahmenwerk begleitet werden,
- › faktischen Mehrwert bieten,
- › über die Zeit stabile Resultate liefern,
- › im Einklang mit ethischen und rechtlichen Vorgaben eingesetzt und
- › auf Basis wissenschaftlicher Erkenntnisse ausgewählt und weiterentwickelt werden.

Um die Einführung von Machine Learning in die Statistikproduktion zu ermöglichen, bedarf es der engen Zusammenarbeit von Fachstatistikerinnen und Fachstatistikern, Methodikerinnen und Methodikern und (spätestens beim Übergang in den Produktivbetrieb) IT-Expertinnen und -Experten. Notwendig sind darüber hinaus eine leistungsfähige IT-Infrastruktur, Forschung und Entwicklung, der Austausch mit anderen statistischen Ämtern, die Unterstützung der Führungsebenen sowie letztlich die Unterstützung durch alle Beschäftigten.

Die im dargestellten Machine-Learning-Projekt durchgeführten Arbeiten warfen weitere Fragen auf und verdeutlichten die Notwendigkeit des weiteren internationalen Austauschs. Das Nachfolgeprojekt „Machine Learning Group 2021“ (UNECE, 2021) setzt sich mit den an den aufgekommenen Fragen orientierenden Arbeitspaketen „Von der Idee zur Lösung“, „Von der Lösung zur Produktion“, „Datenethik und -governance“, „Qualität von Trainingsdaten“ und „QF4SA – Qualitätsrahmen für statistische Algorithmen“ auseinander.


9

Einordnung für die Bundesstatistik

In der Gesamtschau lässt sich festhalten, dass die Bundesstatistik den internationalen Vergleich nicht scheuen muss. Digitalabteilung und Fachbereiche arbeiten zusammen, um Bedarfe für den Einsatz von Machine Learning zu identifizieren, konkrete Projekte werden gemeinsam geprüft, durchgeführt und evaluiert. Einschlägige Fortbildungen finden statt, die Zusammenarbeit mit der Wissenschaft wird ausgebaut. Hinsichtlich der eingesetzten Verfahren folgt die Bundesstatistik – im Rahmen der technischen Möglichkeiten – der aktuellen Forschung.

Die IT-Infrastruktur stellt – hier wie in Ämtern anderer Staaten – sowohl in der Entwicklungs- und Testphase als auch im späteren Produktivbetrieb regelmäßig einen Engpass dar. Das bezieht sich einerseits auf die reine Quantität (Arbeitsspeicher, Rechenkerne), andererseits auch auf die Qualität (für neuronale Netze sind Grafikkarten das Mittel der Wahl). Dass auch bei diesem Engpass der Vergleich nicht gescheut werden muss, liegt weniger an der überdurchschnittlichen Ausstattung im Statistischen Verbund als vielmehr an einem weltweiten Niveau, das unter seinen Möglichkeiten bleibt. Zuletzt steht und fällt der Erfolg von Machine Learning mit dem Vorhandensein qualitativ hochwertiger Trainings- und Testdaten. Diese stehen in der benötigten Form nicht immer direkt aus den bestehenden Produktionsprozessen zur Verfügung, sondern müssen in der Regel noch aufbereitet werden. Besonders bei Fragen der Zuordnung von statistischen Einheiten zu einer Klasse (Unternehmen zu wirtschaftlicher Tätigkeit, Luftbild zu Bodennutzung und so weiter) ist es entscheidend, dass

die Trainings- und Testdaten korrekt gelabelt, also richtig klassifiziert sind. Dies geschieht idealerweise nicht nur durch eine einzige Beschäftigte oder einen einzigen Beschäftigten. Außerdem muss sichergestellt sein, dass die Trainings- und Testdaten die künftig zu klassifizierenden Daten gut widerspiegeln. Diese Aufbereitung und Qualitätssicherung benötigt Zeit und Arbeitskraft; ohne sie ist ein Projekt nicht zu realisieren. Gleiches gilt für den Fall, dass notwendige methodische Untersuchungen oder Weiterentwicklungen nicht durchgeführt oder vorangetrieben werden können.

Das qualitätsgesicherte Etablieren maschinellen Lernens in den Prozessen der amtlichen Statistikproduktion ist somit keine einmalige, sondern vielmehr eine fortlaufende Aufgabe. Sie ist stets an die aktuellen wissenschaftlichen Entwicklungen, die vorhandenen IT-Kapazitäten und die individuellen Bedürfnisse der Fachbereiche anzupassen. 

LITERATURVERZEICHNIS

Arnold, Stephan/Kleine, Sarah. *Neue Wege der Geodatennutzung: Perspektiven der Fernerkundung für die Statistik*. In: WISTA Wirtschaft und Statistik. Ausgabe 5/2017, Seite 31 ff.

Beck, Martin/Dumpert, Florian/Feuerhake, Jörg. *Proof of Concept Machine Learning: Abschlussbericht*. 2018. [Zugriff am 30. Juni 2021]. Verfügbar unter: www.statistischebibliothek.de

Coronado, Abel/Juárez, Jimena. *UNECE-HLG-MOS Machine Learning Project Imagery Theme Report*. 2020. [Zugriff am 30. Juni 2021]. Verfügbar unter: statswiki.unece.org

Dumpert, Florian. *UNECE-HLG-MOS Machine Learning Project Edit and Imputation Theme Report*. 2020. [Zugriff am 30. Juni 2021]. Verfügbar unter: statswiki.unece.org

Eurostat. *European statistics code of practice*. 2017. [Zugriff am 30. Juni 2021]. Verfügbar unter: ec.europa.eu

Goodfellow, Ian/Bengio, Yoshua/Courville, Aaron. *Deep Learning*. Cambridge 2016.

HLG MOS. *HLG-MOS Strategy*. 2021. [Zugriff am 30. Juni 2021]. Verfügbar unter: statswiki.unece.org

James, Gareth/Witten, Daniela/Hastie, Trevor/Tibshirani, Robert. *An Introduction to Statistical Learning*. New York 2013.

Julien, Claude. *Machine Learning Project Report*. 2020. [Zugriff am 30. Juni 2021]. Verfügbar unter: statswiki.unece.org

Lane, Hobson/Howard, Cole/Hapke, Hannes Max. *Natural Language Processing in Action*. Shelter Island 2019.

Measure, Alexander. *Automatic classification of work-related injury and illness narratives*. 2020. [Zugriff am 30. Juni 2021]. Verfügbar unter: statswiki.unece.org

Statistische Ämter des Bundes und der Länder. *Qualitätshandbuch der Statistischen Ämter des Bundes und der Länder*. 2021. [Zugriff am 30. Juni 2021]. Verfügbar unter: www.destatis.de

Statistisches Bundesamt. *Nutzen von Satellitendaten für die amtliche Statistik*. 2019. [Zugriff am 30. Juni 2021]. Verfügbar unter: www.destatis.de

Sthamer, Claus. *UNECE-HLG-MOS Machine Learning Project Classification and Coding Theme Report*. 2020. [Zugriff am 30. Juni 2021]. Verfügbar unter: statswiki.unece.org

UNECE. *HLG-MOS ML Project webinar*. 2020a. [Zugriff am 30. Juni 2021]. Verfügbar unter: statswiki.unece.org

UNECE. *HLG-MOS Machine Learning Project*. 2020b. [Zugriff am 30. Juni 2021]. Verfügbar unter: statswiki.unece.org

LITERATURVERZEICHNIS

UNECE. *Machine Learning Group 2021*. 2021. [Zugriff am 30. Juni 2021]. Verfügbar unter: statswiki.unece.org

Yung, Wesley/Tam, Siu-Ming/Buelens, Bart/Dumpert, Florian/Ascari, Gabriele/Rocci, Fabiana/Burger, Joep/Chipman, Hugh/Choi, InKyung. *A Quality Framework for Statistical Algorithms*. 2020. [Zugriff am 30. Juni 2021]. Verfügbar unter: statswiki.unece.org



Peter Kuntze

ist Diplom-Volkswirt und leitet das Referat „Entstehung des Inlandsprodukts“ des Statistischen Bundesamtes. Die Tätigkeitsschwerpunkte des Referats liegen bei der Berechnung der Bruttowertschöpfung nach Wirtschaftsbereichen sowie der Nettogütersteuern.



Benedikt Kuckelkorn

ist Ökonom und war bis Ende Juni 2021 wissenschaftlicher Mitarbeiter im Referat „Entstehung des Inlandsprodukts“ des Statistischen Bundesamtes. Seit Juli 2021 ist er wissenschaftlicher Mitarbeiter im Referat „Vermögensrechnung, Internationale VGR Methodik“.

MULTIFAKTORPRODUKTIVITÄT IN DEN VOLKSWIRTSCHAFTLICHEN GESAMTRECHNUNGEN

Peter Kuntze, Benedikt Kuckelkorn

➤ **Schlüsselwörter:** Arbeitsproduktivität — Multifaktorproduktivität — Volkswirtschaftliche Gesamtrechnungen — Kapaldienstleistungen — Wirtschaftswachstum

ZUSAMMENFASSUNG

Die Produktivität einer Volkswirtschaft ist ein Kennzeichen ihrer Wettbewerbsfähigkeit und zugleich ein entscheidender Faktor für den künftigen materiellen Wohlstand. Dabei steht meist die Arbeitsproduktivität im Fokus. Dieser Artikel beleuchtet ein alternatives Produktivitätsmaß, die Multifaktorproduktivität. Diese ist bislang kein Bestandteil des Veröffentlichungsprogramms der amtlichen Statistik in Deutschland, kann jedoch mit den Daten der Volkswirtschaftlichen Gesamtrechnungen berechnet werden. Ausgehend von der theoretischen Herleitung werden zwei verschiedene Indikatoren für die Multifaktorproduktivität vorgestellt und deren Ergebnisse analysiert, sowohl für die Gesamtwirtschaft als auch für wichtige Teilbereiche.

➤ **Keywords:** labour productivity — multifactor productivity — national accounts — capital services — economic growth

ABSTRACT

The productivity of an economy is an indicator of its performance and at the same time a decisive factor for future material prosperity. Usually, the focus is on labour productivity. This article looks at an alternative productivity measure, multifactor productivity. It is not yet part of the publication programme of official statistics in Germany but can be calculated with national accounts data. Starting from the theoretical derivation, we present two different indicators for multifactor productivity and analyse their results both for the economy as a whole and for important sub-sectors.

1

Einleitung

Die Produktivitätsentwicklung ist ein wichtiger Indikator für die Veränderung der Leistungs- und Wettbewerbsfähigkeit. Dies gilt sowohl für im Wettbewerb stehende Unternehmen als auch für die Wirtschaft als Ganzes. Die Entwicklung der Produktivität hat darüber hinaus Einfluss auf den lohnpolitischen Verteilungsspielraum und damit letztlich auf den materiellen Wohlstand einer Gesellschaft.

Für die makroökonomische Beurteilung der Produktivitätsentwicklung stellen die Volkswirtschaftlichen Gesamtrechnungen (VGR) derzeit zwei Indikatoren zur Verfügung: die Arbeitsproduktivität sowie die Kapitalproduktivität. Veröffentlicht werden diese für zehn zusammengefasste Wirtschaftsbereiche sowie die Gesamtwirtschaft.¹ Die Arbeitsproduktivität wird zudem vierteljährlich sowie auf Stunden- und Personenbasis publiziert.

Die veröffentlichten Indikatoren für die Arbeits- und die Kapitalproduktivität haben den Vorteil, dass sie ohne besondere Annahmen zu berechnen und intuitiv verständlich sind. Die Multifaktorproduktivität dagegen misst die Effizienz des kombinierten Einsatzes der Produktionsfaktoren und stellt in beiderlei Hinsicht höhere Anforderungen. Das Statistische Bundesamt arbeitet derzeit an einem von der Europäischen Kommission kofinanzierten Projekt, das die Weiterentwicklung der Produktivitätsmessung in den VGR zum Ziel hat. Hierzu zählt die erstmalige Berechnung der Multifaktorproduktivität durch die amtliche Statistik. Indikatoren sollen möglichst vollständig auf Basis veröffentlichter VGR-Daten und mit wenigen zusätzlichen Annahmen zu berechnen sein, damit die Ergebnisse bei einer möglichen Veröffentlichung für die Nutzerinnen und Nutzer leicht nachvollziehbar sind. Im vorliegenden Beitrag werden diese Indikatoren vorgestellt.

Kapitel 2 erläutert zunächst das theoretische Fundament der Produktivitätsmessung sowie die Unterschiede zwischen der Multifaktorproduktivität und den zuvor

genannten partiellen Produktivitätsmaßen. Kapitel 3 stellt die Berechnung und die verwendeten Daten vor. Dabei liegt besonderes Augenmerk auf der Messung des Faktors Kapital, für die es konzeptionell stark voneinander abweichende Optionen gibt. In Kapitel 4 werden die Ergebnisse analysiert und durch den Vergleich mit vorhandenen Berechnungen anderer Institutionen in einen größeren Kontext gestellt. Der Beitrag schließt mit einer Zusammenfassung der wichtigsten Erkenntnisse und einem Ausblick auf die weiteren Arbeiten.

2

Grundlagen der Produktivitätsmessung

Produktivität wird im Allgemeinen definiert als das Verhältnis des Outputs zu einem oder mehreren Inputs. Sie ist ein Maß für die Effizienz der Produktion, das auf der Ebene einzelner Unternehmen, aber auch gesamtwirtschaftlich von großer Bedeutung ist. Maßgeblich für den vorliegenden Beitrag ist die gesamtwirtschaftliche Perspektive.

Aus der Kombination von Maßen für den Output und für den Input resultiert eine Vielzahl möglicher Produktivitätsindikatoren. Grundsätzlich wird zwischen der (partiellen) Einzelfaktor- und der Multifaktorproduktivität (auch Totale Faktorproduktivität genannt) unterschieden. Im ersten Fall wird die Ausbringungsmenge auf einen einzelnen Produktionsfaktor – wie Arbeit, Kapital oder Energie – bezogen. Das bekannteste Einzelfaktor-Produktivitätsmaß ist die Arbeitsproduktivität. Sie ist definiert als Output (in der Regel gemessen durch das Bruttoinlandsprodukt für die Gesamtwirtschaft oder die Bruttowertschöpfung für einzelne Wirtschaftsbereiche) geteilt durch den Arbeitseinsatz:

$$(1) \quad \text{Arbeitsproduktivität} = \frac{\text{Output}}{\text{Arbeitseinsatz}}$$

Der Arbeitseinsatz kann über die Anzahl der Erwerbstätigen oder die Anzahl der geleisteten Arbeitsstunden der Erwerbstätigen gemessen werden. Die letztgenannte Größe ist zielgenauer, vor allem aufgrund der angestiegenen Bedeutung der Teilzeitarbeit in Deutschland (Chalupa/Mai, 2018) oder der aktuell stark erhöhten Kurzarbeit aufgrund der Corona-Pandemie.

¹ Die Ausgangsdaten zur Berechnung der Produktivität werden auf deutlich tieferer Ebene (64 Wirtschaftsbereiche der Klassifikation der Wirtschaftszweige, Ausgabe 2008 [WZ 2008]) veröffentlicht. Daher können Nutzerinnen und Nutzer Produktivitätsentwicklungen für eine Vielzahl weiterer Wirtschaftszweige selbst berechnen.

Eine Vielzahl an Faktoren beeinflusst die Arbeitsproduktivität. Dazu zählen der technologische Fortschritt, die Ausbildung der Arbeitnehmerinnen und Arbeitnehmer, die Kapazitätsauslastung, organisatorische Aspekte und im Besonderen der Einsatz der anderen Produktionsfaktoren Kapital und Vorleistungen. Indikatoren der Multifaktorproduktivität setzen den Output nicht in Relation zu einem einzelnen Produktionsfaktor, sondern zum kombinierten Einsatz der Produktionsfaktoren Arbeit und Kapital.¹² Der resultierende Indikator reflektiert – abseits von Kapital, Arbeitseinsatz und Vorleistungen – den kombinierten Einfluss der übrigen zuvor genannten Faktoren auf den Output. Seine langfristige Entwicklung wird häufig als Maß für den technologischen Fortschritt interpretiert.

Um den kombinierten Faktoreinsatz zu bestimmen, sind jedoch zusätzliche Annahmen zu treffen. Zentral ist das Konzept einer Produktionsfunktion, die den Zusammenhang zwischen dem Output und den Produktionsfaktoren beschreibt. Im Folgenden wird eine Cobb-Douglas-Produktionsfunktion betrachtet. Der Output Y ergibt sich als multiplikative Funktion des Arbeitseinsatzes L , des Kapitaleinsatzes K sowie der Multifaktorproduktivität A .

$$(2) \quad Y = A \cdot L^\alpha \cdot K^{(1-\alpha)}$$

Die Koeffizienten α und $1-\alpha$ bezeichnen die Produktionselastizitäten der Faktoren Arbeit und Kapital. Das bedeutet, dass der Output um α % ansteigt, wenn der Arbeitseinsatz um 1 % erhöht wird, während der Kapitaleinsatz konstant bleibt. Die Annahme, dass die Produktionselastizitäten von Arbeit und Kapital sich zu eins addieren, impliziert konstante Skalenerträge. Dies bedeutet, dass sich der Output verdoppelt, wenn alle Inputs verdoppelt werden. Weiterhin wird angenommen, dass die Unternehmen kostenminimierend und profitmaximierend auf kompetitiven Güter- und Arbeitsmärkten agieren.

Obwohl diese Bedingungen in der Praxis nicht zwangsläufig erfüllt sind, bieten sie eine vernünftige Annäherung für viele Märkte. Kritisch sind die Annahmen in Wirtschaftsbereichen, in denen der Anteil der Sektoren Staat und Private Organisationen ohne Erwerbszweck an der Produktion besonders hoch ist. Die disaggregierte Analyse in Kapitel 4 ist daher auf das Produzie-

rende Gewerbe und ausgewählte Dienstleistungsbereiche begrenzt.¹³ Werden die genannten Annahmen als erfüllt angesehen, erfolgt die Entlohnung der Produktionsfaktoren Kapital und Arbeit mit ihrem jeweiligen Grenzprodukt¹⁴. Als Folge dessen lässt sich die Produktionselastizität des Faktors Arbeit als Anteil des Arbeitnehmerentgelts an der Bruttowertschöpfung messen.

Anders als die Arbeitsproduktivität, deren absolute Größe eine recht verständliche Maßeinheit hat, eignet sich das Niveau der Multifaktorproduktivität nicht für eine Interpretation. Relevant ist vielmehr deren Veränderung. Daher wird die Wachstumsrate der Multifaktorproduktivität direkt berechnet als:

$$(3) \quad WR_{MFP} = \frac{\Delta A}{A} = \frac{\Delta Y}{Y} - \alpha \frac{\Delta L}{L} - (1 - \alpha) \frac{\Delta K}{K}$$

Die Wachstumsrate der Multifaktorproduktivität ergibt sich als Residuum, also den Teil des Output-Wachstums, der nicht durch das Wachstum des Arbeitseinsatzes und das des Kapitaleinsatzes erklärt werden kann. Zwischen Multifaktorproduktivität und Arbeitsproduktivität besteht ein direkter Zusammenhang. Durch Umformen der Produktionsfunktion ergibt sich folgende Gleichung für die Arbeitsproduktivität:

$$(4) \quad \frac{Y}{L} = A \cdot \left(\frac{K}{L}\right)^{1-\alpha}$$

Die Entwicklung der Arbeitsproduktivität ist abhängig vom Wachstum der Multifaktorproduktivität und dem der Kapitalintensität. Diese Vorgehensweise, die sogenannte Wachstumszerlegung (englisch: growth accounting), geht zurück auf Arbeiten von Jan Tinbergen (1942) und Robert Solow (1957).

¹² Vorleistungen müssen als weiterer Produktionsfaktor berücksichtigt werden, falls der Output durch den Produktionswert gemessen wird.

¹³ Die Analyse des Dienstleistungssektors beschränkt sich auf die Abschnitte G bis N der WZ 2008. Diese umfassen Handel, Verkehr, Gastgewerbe, Information und Kommunikation, Finanz- und Versicherungsdienstleister, das Grundstücks- und Wohnungswesen sowie Unternehmensdienstleister. Das Produzierende Gewerbe umfasst die Abschnitte B bis F (Bergbau, Verarbeitendes Gewerbe, Energie- und Wasserversorgung sowie das Baugewerbe).

¹⁴ Grenzprodukt bezeichnet den Zuwachs der Produktion, der durch den Einsatz einer weiteren Einheit des Produktionsfaktors erzielt wird.

3

Berechnung der Multifaktorproduktivität

Das folgende Kapitel erläutert die gewählte Vorgehensweise zur Berechnung der Multifaktorproduktivität und beschreibt die verwendeten Daten. Ziel des Projekts ist es, Indikatoren für die Multifaktorproduktivität zu entwickeln, die so weit wie möglich auf veröffentlichten Daten der VGR basieren.¹⁵

Der Output Y wird gemessen durch das Bruttoinlandsprodukt für die Gesamtwirtschaft und die Bruttowertschöpfung, also dem Saldo aus Produktionswert und Vorleistungen, für einzelne Wirtschaftsbereiche. Die Produktionselastizität des Faktors Arbeit α wird gemessen als Anteil des Arbeitseinkommens an der Bruttowertschöpfung abzüglich sonstiger Nettoproduktionsabgaben. Da allerdings auch Selbstständige Arbeitseinkommen beziehen, muss das in den VGR ausgewiesene Arbeitnehmerentgelt angepasst werden. Wir treffen dazu die – in der Literatur übliche – Annahme, dass Selbstständige die gleiche durchschnittliche Entlohnung wie Arbeitnehmerinnen und Arbeitnehmer im selben Wirtschaftsbereich erhalten (OECD, 2001). Der Einkommensanteil des Faktors Arbeit ergibt sich dann wie folgt:

$$(5) \quad \alpha = \frac{\text{Arbeitnehmerentgelt} \cdot \frac{\text{Anzahl Erwerbstätige}}{\text{Anzahl Arbeitnehmer}}}{\text{Bruttowertschöpfung} - \text{Sonstige Nettoproduktionsabgaben}}$$

Die gewählte Methode liefert einen relativ stabilen Einkommensanteil für den Faktor Arbeit. Im Beobachtungszeitraum von 1993 bis 2018 liegt dieser gesamtwirtschaftlich zwischen 60 und 65 %. Alternative Berechnungen haben gezeigt, dass der Einkommensanteil α zwar auf die getroffenen Annahmen reagiert, die Auswirkungen auf die Entwicklung der Multifaktorproduktivität aber vernachlässigbar sind.

Der Arbeitseinsatz L wird gemessen durch die tatsächlich geleisteten Arbeitsstunden der Erwerbstätigen, die

¹⁵ Mit Ausnahme des Nettoanlagevermögens ist dies, gegliedert nach Wirtschaftsbereichen und Vermögensarten, auch möglich. Dieses wird zur Berechnung des Kapitaleinsatzes benötigt. Das Nettoanlagevermögen ist bei Interesse ebenfalls für Nutzerinnen und Nutzer verfügbar, da es Teil des Lieferprogramms des Europäischen Systems Volkswirtschaftlicher Gesamtrechnungen ist und vom Statistischen Amt der Europäischen Union (Eurostat) veröffentlicht wird.

das Institut für Arbeitsmarkt- und Berufsforschung (IAB) im Rahmen der Arbeitszeitrechnung berechnet. Hierbei wird implizit die Annahme getroffen, dass jede Stunde Arbeit den gleichen Beitrag zur Produktion liefert, unabhängig vom Qualifikationsniveau. Um dieses berücksichtigen zu können, müsste der Arbeitseinsatz um die Qualität der Arbeitsstunden korrigiert werden. Hierfür wäre die Verlinkung von Mikrodaten der Arbeitskräfteerhebung mit VGR-Daten erforderlich. Eurostat veröffentlicht einen Indikator für diesen qualitätsangepassten Arbeitseinsatz (englisch: Quality Adjusted Labour Input oder QALI) als experimentelle Statistik.¹⁶ Unter der Maßgabe, dass möglichst nur veröffentlichte VGR-Angaben Verwendung finden sollen, wird vorerst auf diese Anpassung verzichtet.

Der Kapitaleinsatz K kann mit zwei sehr unterschiedlichen Methoden gemessen werden. Der erste Ansatz nutzt das Nettoanlagevermögen. Dies hat den Vorteil, dass veröffentlichte Zahlen der VGR ohne weitere Berechnungen verwendet werden können. Konzeptionell ist das Nettoanlagevermögen allerdings nicht uneingeschränkt zur Produktivitätsmessung geeignet. Das Nettoanlagevermögen wird berechnet, indem verschiedene Vermögensgüter mit ihren Wiederbeschaffungspreisen aggregiert werden und ist daher ein Maß für den Marktwert des vorhandenen Kapitals eines Wirtschaftsbereiches oder einer Volkswirtschaft. Der Marktwert eines Vermögensguts hängt ab von allen künftigen Beiträgen zur Produktion. Für die Produktivitätsmessung ist allerdings der Beitrag des Kapitals innerhalb einer Beobachtungsperiode besser geeignet. Relativ kurzlebige Vermögensgüter, wie Computerhardware, können im Vergleich zu langlebigen Anlagen, wie Gebäuden, nur für einen geringeren Zeitraum genutzt werden. Daher müssen die jährlichen Beiträge zur Produktion relativ zum Marktwert größer sein.

Der zweite Ansatz folgt diesem Gedanken. Er misst den Produktionsbeitrag des Kapitals als die von Unternehmen bezogenen Kapitaldienstleistungen (englisch: capital services). Deren Wert und Volumen können jedoch in der Regel nicht direkt beobachtet werden, da Nutzer und Besitzer der Assets häufig identisch sind. Es gibt also keine Zahlungsströme, die erfasst werden könnten. Daher müssen die Kapitaldienstleistungen geschätzt werden. Dies erfolgt über den Ansatz der Nutzungs-

¹⁶ Mehr Informationen und Daten zu QALI unter ec.europa.eu

kosten des Kapitals (Jorgenson, 1963). Verschiedene Vermögensarten werden aggregiert zum gesamten Kapitaleinsatz, indem sie mit dem jeweiligen Anteil ihrer Nutzungskosten an den Gesamtnutzungskosten gewichtet werden (siehe den Exkurs). Dazu werden für jede Vermögensart in jedem Wirtschaftsbereich die Nutzungskosten u wie folgt geschätzt:

$$(6) \quad u_t^{i,j} = q_t^{i,j} \cdot (r_t^i + d_t^j) - (q_t^{i,j} - q_{t-1}^{i,j})$$

Hierbei sind die Nutzungskosten $u_t^{i,j}$ die Kosten, die für die Nutzung der Vermögensart j im Wirtschaftsbereich i je Periode t anfallen. $q_t^{i,j}$ ist der Marktwert des Nettoanlagevermögens der Vermögensart j im Wirtschaftsbereich i , d_t^j ist die Abschreibungsrate der Vermögensart j und r_t^i ist ein Maß für die Finanzierungskosten im Wirtschaftsbereich i .

Exkurs

Den Unterschied zwischen beiden Ansätzen soll ein einfaches Beispiel verdeutlichen, das zwei unterschiedliche Investitionsgüter betrachtet: zum einen einen Laptop im Wert von 3 000 Euro und einer Nutzungsdauer von 3 Jahren, zum anderen eine Produktionsanlage im Wert von 20 000 Euro und einer Nutzungsdauer von 20 Jahren. Finanzierungs- und Opportunitätskosten sowie Preisänderungen seien hier vernachlässigt. Bei Nutzung des Nettoanlagevermögens – hier vereinfacht angesehen als Summe der Marktwerte in Höhe von 23 000 Euro – spielt der Laptop eine vergleichsweise geringe Rolle. Beim Konzept der Kapitaldienstleistungen dagegen stehen die Nutzungskosten im Vordergrund. Bei der unterstellten Nutzungsdauer und linearer Abschreibung betragen diese für beide Güter 1 000 Euro. Die relative Bedeutung des Investitionsgutes Laptop ist also um ein Vielfaches höher, wenn Güter nicht mit dem Wert, sondern dem Anteil an den Nutzungskosten gewichtet werden.

Die Nutzungskosten setzen sich also zusammen aus

- › den Finanzierungskosten (falls die Assets fremdfinanziert sind) oder den Opportunitätskosten (die entstehen, weil Kapital nicht anderweitig eingesetzt werden kann),
- › dem durch Abschreibungen gemessenen Wertverlust und
- › Kapitalgewinnen oder -verlusten (die sich durch Preisänderungen der Assets ergeben).

Die Nutzungskosten werden berechnet für die sieben Vermögensarten Wohnbauten, Nichtwohnbauten, Ausrüstungen und militärische Waffensysteme, Fahrzeuge, Nutztiere und Nutzpflanzungen, Forschung und Entwicklung sowie Software und Datenbanken.

Die Abschreibungsrate ergibt sich aus den VGR-Veröffentlichungen der Abschreibungen und des Nettoanlagevermögens. Da die Abschreibungen für die genannten Vermögensarten nur für die Gesamtwirtschaft veröffentlicht werden, wird vereinfachend angenommen, dass die Abschreibungsraten der verschiedenen Vermögensarten über die Wirtschaftsbereiche identisch sind. Kapitalgewinne und -verluste ergeben sich aus Preisänderungen der jeweiligen Klasse und dem Volumen des Nettoanlagevermögens.

Für die Bestimmung der Zinsrate gibt es zwei grundsätzlich unterschiedliche Ansätze. Erstens kann eine exogene Zinsrate angenommen werden. Diese Methode wird beispielsweise vom Sachverständigenrat zur Begutachtung der gesamtwirtschaftlichen Entwicklung (2017) angewandt. Die Zinsrate wird in diesem Fall gemessen durch die Umlaufrendite von Inhaberschuldverschreibungen abzüglich der von Consensus Economics⁷ erhobenen Inflationserwartungen (Deutsche Bundesbank, 2012). Der Zins ist in diesem Fall für alle Vermögensarten und Wirtschaftsbereiche identisch.

Die zweite Methode nimmt eine endogene Zinsrate an, die sich aus VGR-Daten berechnen lässt. Dazu wird angenommen, dass sich die gesamte Bruttowertschöpfung aufteilt in Arbeits- und Kapitaleinkommen. Das Kapitaleinkommen entspricht dem gesamten Wert der Kapitaldienstleistungen und lässt sich bestimmen aus der Differenz der Bruttowertschöpfung (abzüglich sonstiger Nettoproduktionsabgaben) und dem angepassten Arbeitseinkommen (siehe oben). Mithilfe der Abschreibungsrate d , den Kapitalgewinnen oder -verlusten π , und dem Nettoanlagevermögen KN lässt sich die Zinsrate r im Wirtschaftsbereich i dann residual bestimmen als:

$$(7) \quad r_t^i = \frac{(\text{Kapitaleinkommen}_t^i - \sum_j d_t^j \cdot KN_t^{i,j} + \sum_j \pi_t^{i,j})}{\sum_j KN_t^{i,j}}$$

⁷ Consensus Economics erhebt monatliche Prognosen internationaler Experten bezüglich einer Vielzahl makroökonomischer Variablen, zum Beispiel der Inflationsrate.

Die Zinsrate ist identisch für die verschiedenen Vermögensarten, unterscheidet sich aber zwischen den Wirtschaftsbereichen. Aus diesem Grund und weil diese Vorgehensweise keine externen Finanzmarktdaten benötigt, wird diese Methode bevorzugt. Vergleichsrechnungen zeigen außerdem nur minimale Unterschiede im resultierenden Wachstum der Multifaktorproduktivität.

Nachdem Zinsrate, Abschreibungsrate sowie Kapitalgewinne und -verluste bestimmt sind, können die Kapitalnutzungskosten der einzelnen Vermögensarten berechnet werden. Die Wachstumsrate der gesamten bezogenen Kapitaldienstleistungen KD im Wirtschaftsbereich i ergibt sich als Törnqvist-Index der mit ihrem Anteil an den gesamten Kapitalnutzungskosten gewichteten realen Wachstumsraten des Nettoanlagevermögens der Vermögensarten:

$$(8) \quad WR_{KD_t^i} = \frac{1}{2} \sum_j \left(\frac{u_{t-1}^{i,j}}{u_{t-1}^i} + \frac{u_t^{i,j}}{u_t^i} \right) \cdot WR_{KN_t^{i,j}}$$

Berechnet werden die Kapitaldienstleistungen für 21 Wirtschaftsbereiche.

4

Ergebnisse

Im Folgenden werden die Ergebnisse für die Berechnung des Kapitaleinsatzes und der Multifaktorproduktivität in beiden Varianten präsentiert und mit bereits veröffentlichten Indikatoren verglichen.

Über die gesamtwirtschaftlichen Ergebnisse hinaus werden das Produzierende Gewerbe (Abschnitte B bis F der Wirtschaftszweigklassifikation) und ausgewählte Dienstleistungsbereiche (Abschnitte G bis N) für den Zeitraum 1993 bis 2018 separat analysiert. Die Abschnitte O bis T werden nicht näher betrachtet, da aufgrund des relativ hohen Anteils der Sektoren Staat und Private Organisationen ohne Erwerbszweck die Annahmen von kompetitiven Märkten und kostenminimierender beziehungsweise profitmaximierender Unternehmen nicht erfüllt sind (siehe Kapitel 2). Der Abschnitt A – Land- und Forstwirtschaft, Fischerei – wird in dieser Analyse unter anderem deshalb nicht betrachtet, weil Output und in der Folge die Produktivitätsentwicklung zu einem großen Teil durch Umweltbedingungen bestimmt werden.

Die betrachteten Abschnitte B bis N erzeugen zusammen durchschnittlich 77,4 % der gesamten Bruttowertschöpfung⁸.

4.1 Kapitaleinsatz

➤ **Grafik 1** auf Seite 70 zeigt die Entwicklung des Kapitaleinsatzes für die Gesamtwirtschaft, gemessen als Veränderungsrate des Nettoanlagevermögens beziehungsweise der Kapitaldienstleistungen. Zu Beginn des Beobachtungszeitraums sind die Kapitaldienstleistungen mit einer etwas niedrigeren Rate als das Nettoanlagevermögen gewachsen. Ab dem Jahr 1998 stiegen die Kapitaldienstleistungen deutlich stärker an als das Nettoanlagevermögen. Dies lässt sich auf die steigende Bedeutung von relativ kurzlebigen immateriellen Vermögensgütern und Ausrüstungen der Informations- und Kommunikationstechnik zurückführen. Aufgrund der höheren Abschreibungsrate werden sie bei der Aggregation mit Nutzungskosten stärker gewichtet als bei der Gewichtung mit dem Wiederbeschaffungspreis, um den relativ höheren jährlichen Beitrag zur Produktion zu reflektieren.

In beiden Fällen ist ein Einbruch der Wachstumsrate zu Beginn des 21. Jahrhunderts zu beobachten, der sich im Fall der Kapitaldienstleistungen gegen Ende des Betrachtungszeitraums wieder erholt. Dies geht auf die stark steigenden Investitionen in Forschung und Entwicklung, Software und Datenbanken und das ebenfalls wieder höhere Wachstum der Ausrüstungsinvestitionen zurück.

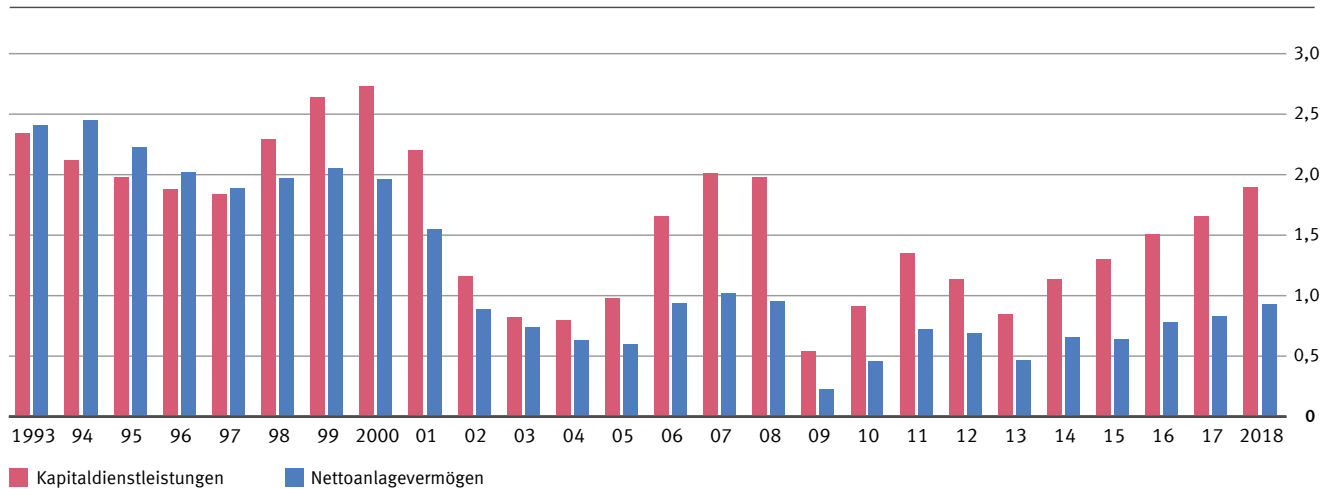
➤ **Grafik 2** auf Seite 71 zeigt die Entwicklung des Kapitaleinsatzes im Produzierenden Gewerbe und in den Dienstleistungsbereichen. Die Trends, die für die Gesamtwirtschaft beobachtet werden können, gelten für beide zusammengefasste Wirtschaftsbereiche in vergleichbarer Weise. Nettoanlagevermögen und Kapitaldienstleistungen sind im Produzierenden Gewerbe zunächst mit ähnlichen Raten gewachsen. Mit steigender Bedeutung neuer Technologien ist allerdings ein deutlich stärkeres Wachstum der Kapitaldienstleistungen zu

8 Der durchschnittliche Anteil an der gesamten Bruttowertschöpfung der Marktproduktion beträgt im Beobachtungszeitraum 87 %. Der Anteil des Produzierenden Gewerbes an der gesamten Bruttowertschöpfung beträgt 30,5 %, der der gewählten Dienstleistungsbereiche 47 %.

Grafik 1

Kapitaleinsatz der Gesamtwirtschaft

Veränderungsrate in %



Quelle: Eigene Berechnungen

2021 - 0274

beobachten. Deren durchschnittliche jährliche Wachstumsrate ist mit 0,93 % deutlich höher als die des Nettoanlagevermögens (0,15 %). Ebenfalls ist ein Einbruch der Wachstumsrate beider Indikatoren um die Jahrtausendwende mit anschließender Erholung speziell der Kapitaldienstleistungen festzustellen.

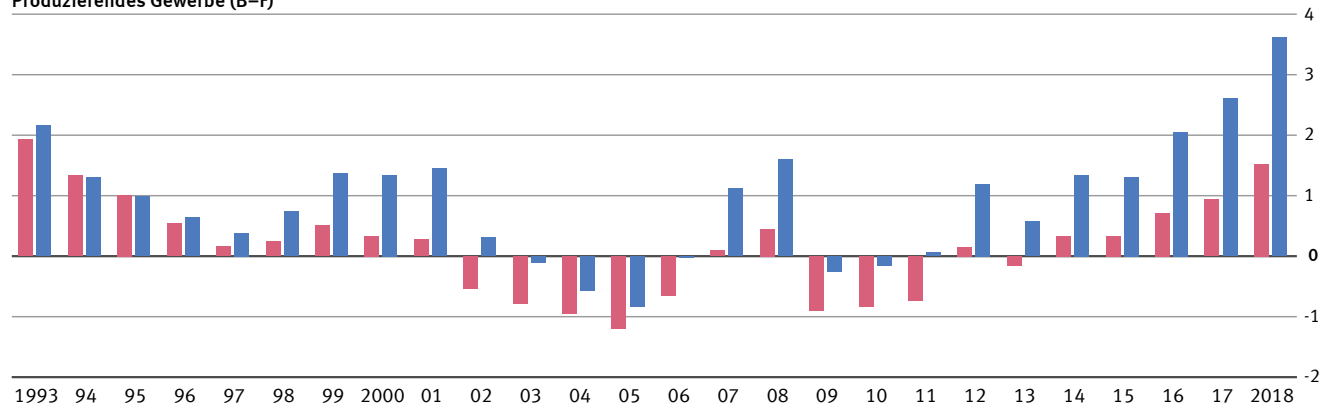
Im Dienstleistungsbereich zeigen sich ebenfalls Unterschiede zwischen der Wachstumsrate des Nettoanlagevermögens und der Kapitaldienstleistungen. Die Kapitaldienstleistungen haben jahresdurchschnittlich um 2 % zugenommen, das Nettoanlagevermögen dagegen lediglich um 1,6 %. Beide Indikatoren für den Kapitaleinsatz sind seit der Jahrtausendwende nicht mehr so stark gestiegen, im Gegensatz zum Produzierenden Gewerbe und der Gesamtwirtschaft ist allerdings kein nennenswerter Aufholprozess zu beobachten.

Zusammenfassend lassen sich deutlich sichtbare Unterschiede zwischen den Entwicklungen des Nettoanlagevermögens und den Kapitaldienstleistungen erkennen. Da das Nettoanlagevermögen den Marktwert des Kapitals und nicht den jährlichen Beitrag zur Produktion misst, unterschätzt es mit steigender Bedeutung kurzlebiger neuer Technologien die Zunahme des Kapitaleinsatzes in der Produktion systematisch. Weiterhin sind deutliche Unterschiede in der Dynamik zwischen sekundärem und tertiärem Sektor zu sehen.

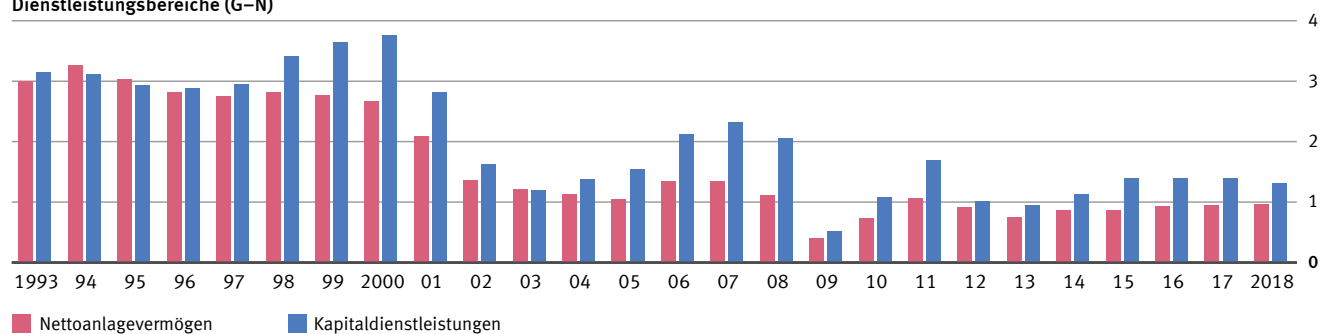
Grafik 2

Kapitaleinsatz des Produzierenden Gewerbes und der Dienstleistungsbereiche
Veränderungsrate in %

Produzierendes Gewerbe (B–F)



Dienstleistungsbereiche (G–N)



Wirtschaftsabteilungen der Klassifikation der Wirtschaftszweige, Ausgabe 2008 (WZ 2008).
Quelle: Eigene Berechnungen

2021 - 0275

4.2 Multifaktorproduktivität

➤ Grafik 3 auf Seite 72 zeigt die jährlichen Veränderungsrate der „Basis“-Multifaktorproduktivität, deren Berechnung das Nettoanlagevermögen als Maß für den Kapitaleinsatz zugrunde liegt, und der Multifaktorproduktivität mit Kapitaldienstleistungen. Zudem wird die Entwicklung der Arbeitsproduktivität gegenübergestellt. Gezeigt werden wiederum Ergebnisse für die Gesamtwirtschaft sowie für das Produzierende Gewerbe und ausgewählte Dienstleistungsbereiche.

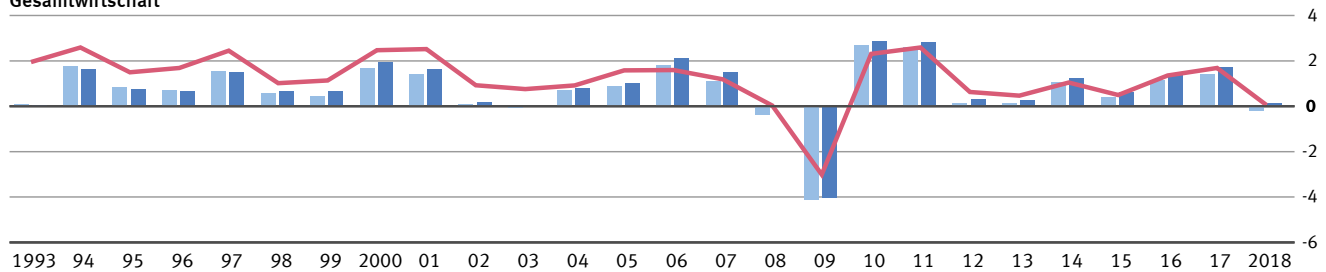
Im Beobachtungszeitraum beträgt das jahresdurchschnittliche Wachstum der Multifaktorproduktivität basierend auf Kapitaldienstleistungen 0,7 % für die gesamte Wirtschaft. Der Basis-Indikator wächst hingegen

mit einer durchschnittlichen Rate von 0,9 %, die Arbeitsproduktivität wächst durchschnittlich um 1,2 %. Im Produzierenden Gewerbe verhält sich die Produktivität dynamischer, die Wachstumsraten sind deutlich größer. Der Zuwachs der Multifaktorproduktivität beträgt durchschnittlich 1,5 % (Basis-Indikator: 1,7 %), das Wachstum der Arbeitsproduktivität 2,0 %. In den Dienstleistungsbereichen ist das Wachstum geringer, die jährliche Wachstumsrate der Multifaktorproduktivität beträgt 0,6 % (Basis-Indikator: 0,8 %), die der Arbeitsproduktivität 1,2 %. Der Basis-Indikator überschätzt das Wachstum der Multifaktorproduktivität, da das Nettoanlagevermögen, wie oben ausgeführt, den Kapitaleinsatz unterschätzt.

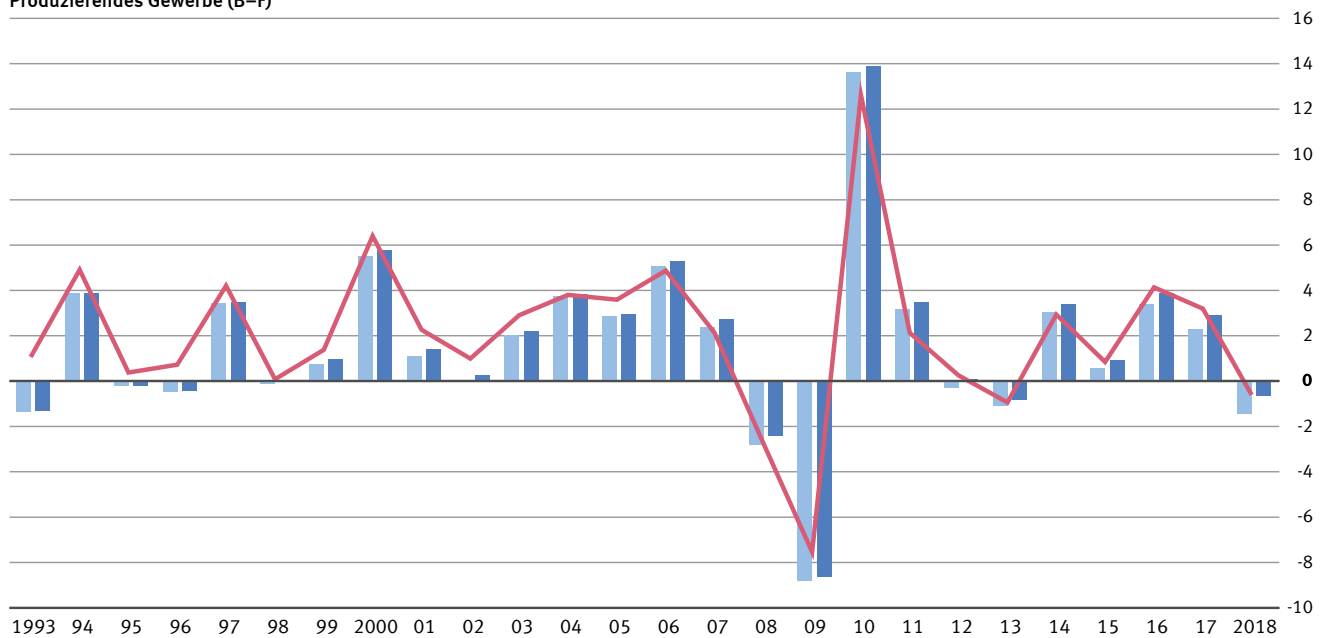
Grafik 3

Arbeits- und Multifaktorproduktivität
Veränderungsrate in %

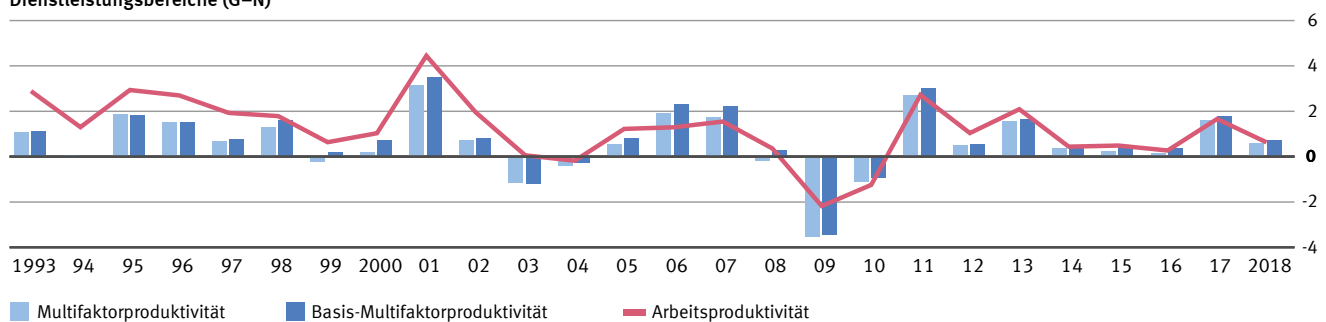
Gesamtwirtschaft



Produzierendes Gewerbe (B–F)



Dienstleistungsbereiche (G–N)



Wirtschaftsabteilungen der Klassifikation der Wirtschaftszweige, Ausgabe 2008 (WZ 2008).
Quelle: IAB, eigene Berechnungen

2021 - 0276

Im Gegensatz zur Arbeitsproduktivität, deren Wachstum sich zunehmend abschwächt, ist dies bei der Multifaktorproduktivität weder für die Gesamtwirtschaft noch für das Produzierende Gewerbe der Fall. Im Gegensatz dazu sinkt das Wachstum der Multifaktorproduktivität im Dienstleistungsbereich im Beobachtungszeitraum leicht. Der Rückgang ist allerdings nicht so stark ausgeprägt wie bei der Arbeitsproduktivität. Die Entwicklung der beiden Indikatoren ist auf allen drei Aggregations Ebenen dennoch sehr ähnlich, dies zeigt sich auch in einer sehr starken Korrelation zwischen den Maßen. Aus diesem Grund eignet sich der Basis-Indikator trotz der konzeptionellen Schwächen zur Analyse der Entwicklung der Multifaktorproduktivität.

Da sich die Arbeitsproduktivität direkt aus Multifaktorproduktivität und Kapitalintensität ableiten lässt, erlauben diese Ergebnisse zwei Schlussfolgerungen: Zum einen ist die Multifaktorproduktivität der Haupttreiber des Wachstums der Arbeitsproduktivität. Gesamtwirtschaftlich liegt ihr Beitrag zu deren Wachstum bei durchschnittlich 60 %, im Produzierenden Gewerbe sogar bei 68 %. Zum anderen lässt sich das nachlassende Wachstum der Arbeitsproduktivität nicht durch die Entwicklung der Multifak-

torproduktivität erklären, sondern wird durch Rückgänge des Wachstums der Kapitalintensität verursacht.

4.3 Vergleich mit anderen Ergebnissen

Um die Qualität der vorliegenden Ergebnisse zu überprüfen, werden diese mit zwei anderen Berechnungen zur Multifaktorproduktivität verglichen.

Der Sachverständigenrat zur Begutachtung der gesamtwirtschaftlichen Entwicklung berechnet im Rahmen der Mittelfristprojektion als Teil seines Jahresgutachtens Indikatoren der Multifaktorproduktivität für die deutsche Volkswirtschaft. In der Regel wird lediglich deren Trendwachstum veröffentlicht, was einen direkten Vergleich mit den hier vorliegenden Ergebnissen ausschließt. Die letzte Veröffentlichung unbereinigter Schätzungen des jährlichen Wachstums der Multifaktorproduktivität stammt aus dem Jahresgutachten 2015/16 (SVR, 2015). Der Kapitaleinsatz wurde mit dem Bruttoanlagevermögen gemessen. Diese Ergebnisse werden hier zum Vergleich genutzt.

Grafik 4

Multifaktorproduktivität der Gesamtwirtschaft

Veränderungsrate in %



1 EU KLEMS ist ein Forschungsprojekt zur Produktivitätsmessung innerhalb der Europäischen Union.

Quelle: Sachverständigenrat zur Begutachtung der gesamtwirtschaftlichen Entwicklung (2015), Stehrer und andere (2019), eigene Berechnungen

Finanziert durch die Generaldirektion Wirtschaft und Finanzen der Europäischen Kommission, veröffentlicht das Wiener Institut für Internationale Wirtschaftsvergleiche (wiiw) im Rahmen des EU KLEMS-Projekts Schätzungen zur Multifaktorproduktivität für die 27 EU-Mitgliedstaaten, das Vereinigte Königreich, die Vereinigten Staaten und Japan (wiiw, 2019).¹⁹ Methodische Unterschiede finden sich bei der Schätzung des Kapitaleinsatzes sowie bei der Wahl des Maßes für den Arbeitseinsatz. Die aktuelle Veröffentlichung deckt den Zeitraum 1996 bis 2017 ab.

➤ **Grafik 4** stellt die Ergebnisse des Sachverständigenrats und von EU KLEMS den hier vorgelegten gegenüber.

Trotz der methodischen Differenzen liefern die betrachteten Ansätze sehr ähnliche Ergebnisse. Die durchschnittliche Differenz unserer Ergebnisse zu denen des Sachverständigenrats beträgt 0,3 Prozentpunkte und die Differenz zu EU KLEMS 0,4 Prozentpunkte. Der Korrelationskoeffizient liegt bei 0,93 und 0,96.

Wachstumsraten der Kapitaldienstleistungen sind ebenfalls Teil der EU KLEMS-Veröffentlichung. Die durchschnittliche Differenz zu unserer Schätzung beträgt 0,3 Prozentpunkte. Die Indikatoren sind mit einem Korrelationskoeffizienten von 0,96 ebenfalls in der Aussage sehr vergleichbar.

5


Fazit und Ausblick

Der Beitrag untersucht die Multifaktorproduktivität als mögliche Erweiterung des statistischen Datenangebots der Volkswirtschaftlichen Gesamtrechnungen. Den im Vergleich zu den Einzelfaktor-Produktivitätsmaßen größeren Anforderungen in Bezug auf Berechnung und Interpretation steht ein erheblicher zusätzlicher Erkenntnisgewinn gegenüber. Die Multifaktorproduktivität misst jenen Teil des Wirtschaftswachstums, der nicht durch Änderungen der Produktionsfaktoren erklärt werden kann. Sie ist dadurch ein wichtiges Maß für langfristige

Effizienzsteigerungen und technologischen Fortschritt in der Produktion.

Ausgehend von den zwei vorgestellten Optionen zur Messung des Kapitaleinsatzes (Nettoanlagevermögen und Kapitaldienstleistungen) wurden zwei Maße für die Multifaktorproduktivität präsentiert und deren Ergebnisse analysiert. Beide Varianten zeigen, dass das Wachstum der Multifaktorproduktivität – im Gegensatz zu dem der Arbeitsproduktivität – nicht spürbar zurückgeht. Produzierendes Gewerbe und Dienstleister unterscheiden sich in der Entwicklung nicht wesentlich, jedoch liegen die Veränderungsraten im Produzierenden Gewerbe höher. Die Nutzung der Kapitaldienstleistungen in der Berechnung führt zu systematisch niedrigeren Veränderungsraten der Multifaktorproduktivität.

Die vorgestellten Berechnungen basieren ausschließlich auf veröffentlichten VGR-Daten. Sie folgen einer anerkannten Methodik und erfordern relativ wenige Annahmen. Der Abgleich mit bisherigen Veröffentlichungen bestätigt die Ergebnisse. Damit sind die Grundlagen für eine Erweiterung des Veröffentlichungsprogramms des Statistischen Bundesamtes um diesen wichtigen Produktivitätsindikator gelegt.

Bei der Entscheidung für eine der beiden Varianten muss eine Abwägung erfolgen: Die Basisvariante unter Verwendung des Nettoanlagevermögens ist einfacher konstruiert und damit für Nutzerinnen und Nutzer leichter nachvollziehbar. Die Variante mit den Kapitaldienstleistungen ist anspruchsvoller und enthält mehr Annahmen, hat aber gewisse konzeptionelle Vorteile. Die Darstellung der Multifaktorproduktivität auf Grundlage der VGR-Daten wird auch auf europäischer Ebene forciert: So plant Eurostat die Veröffentlichung der Basisvariante für die 27 Mitgliedstaaten der Europäischen Union als experimentelle Statistik zum Jahresende 2021. Vor diesem Hintergrund bietet sich als Ergänzung eine weitergehende nationale Veröffentlichung auf Basis der Kapitaldienstleistungen an. 

¹⁹ EU KLEMS ist ein Forschungsprojekt zur Produktivitätsmessung innerhalb der Europäischen Union. KLEMS steht für die Produktionsfaktoren Kapital (K), Arbeit (L), Energie (E), Roh-, Hilfs- und Betriebsstoffe (M) und Dienstleistungen (S). Daten und zusätzliche Informationen unter euklems.eu

LITERATURVERZEICHNIS

Chalupa, Johannes/Mai, Christoph-Martin. *Entwicklungen am Arbeitsmarkt in Österreich und Deutschland – zwischen Jobwunder und Produktivitätsparadoxon*. In: WISTA Wirtschaft und Statistik. Ausgabe 6/2018, Seite 48 ff.

Deutsche Bundesbank. *Potenzialwachstum der deutschen Wirtschaft – Mittelfristige Perspektiven vor dem Hintergrund demographischer Belastungen*. In: Monatsbericht. Ausgabe April 2012, Seite 13 ff.

Jorgenson, Dale W. *Capital Theory and Investment Behaviour*. In: American Economic Association. 1963. Jahrgang 53, Seite 247 ff.

OECD (Organisation für wirtschaftliche Zusammenarbeit und Entwicklung). *Measurement of aggregate and industry-level productivity growth*. 2001. OECD Manual.

Solow, Robert M. *Technical Change and the Aggregate Production Function*. In: Review of Economics and Statistics. Jahrgang 39. Ausgabe 3/1957, Seite 312 ff.

SVR (Sachverständigenrat zur Begutachtung der gesamtwirtschaftlichen Entwicklung). *Zukunftsfähigkeit in den Mittelpunkt*. Jahresgutachten 2015/16. Wiesbaden 2015.

SVR (Sachverständigenrat zur Begutachtung der gesamtwirtschaftlichen Entwicklung). *Für eine zukunftsorientierte Wirtschaftspolitik*. Jahresgutachten 2017/18. Wiesbaden 2017.

Tinbergen, Jan. *Zur Theorie der langfristigen Wirtschaftsentwicklung*. In: Zeitschrift des Instituts für Weltwirtschaft an der Universität Kiel. Jahrgang 55. Ausgabe 1942, Seite 511 f.

wiiw (Wiener Institut für Internationale Wirtschaftsvergleiche). *Growth and Productivity Accounts*. Release 2019. 2019. [Zugriff am 12. Juli 2021]. Verfügbar unter: euklems.eu



Jannik Schaller

ist seit seinem Masterabschluss in „Survey-Statistik“ wissenschaftlicher Mitarbeiter im Referat „Forschungsdatenzentrum, Methoden der Datenanalyse“ des Statistischen Bundesamtes. Er betreut das Thema Mikrosimulation und erforscht Verfahren der Datenfusion/Statistical Matching zur Erstellung einer synthetischen Mikrosimulationsdatengrundlage. Für seine an der Otto-Friedrich-Universität Bamberg bei Professor Dr. Martin Messingschlager verfasste Masterarbeit zum Thema „Datenfusion von EU-SILC und HBS: Vergleich zwischen Random Hot-Deck und Predictive Mean Matching im Rahmen einer Simulationsstudie“ wurde er mit dem Gerhard-Fürst-Preis 2020 des Statistischen Bundesamtes ausgezeichnet. Mit dem folgenden Beitrag stellt er diese Masterarbeit vor.

DATENFUSION VON EU-SILC UND HOUSEHOLD BUDGET SURVEY – EIN VERGLEICH ZWEIER FUSIONS-METHODEN

Jannik Schaller

➤ **Schlüsselwörter:** Statistisches Matching – fehlende Daten – Imputationsverfahren – Missing-By-Design-Pattern – Predictive Mean Matching

ZUSAMMENFASSUNG

Zur Beurteilung des sozialen und wirtschaftlichen Lebensstandards in der Europäischen Union (EU) strebt die amtliche Statistik an, Einkommen und Konsumausgaben privater Haushalte gemeinsam zu betrachten. Hierfür existiert derzeit keine einheitliche Datengrundlage. Die Einkommensangaben sind detailliert in der europaweiten EU-SILC-Erhebung, die Konsuminformationen wiederum im Household Budget Survey erfasst. Beide Datenbestände sollen fusioniert werden, um eine gemeinsame Analyse von Einkommen und Konsumausgaben privater Haushalte zu ermöglichen. Der vorliegende Beitrag vergleicht den von Eurostat vorgeschlagenen Random Hot-Deck-Ansatz mit einer alternativen Datenfusionsmethode, dem Predictive Mean Matching. Er evaluiert die Performance beider Verfahren in einer Simulationsstudie.

➤ **Keywords:** statistical matching – missing data – imputation techniques – missing-by-design pattern – predictive mean matching

ABSTRACT

In order to assess the social and economic living standards in the European Union (EU), official statistics seek to provide joint information on income and consumption expenditures of private households. Currently, no uniform data basis exists for this purpose. While income is measured across the EU in great detail by EU-SILC, information on consumption is reported in the household budget survey. Both stocks of data are to be fused in order to enable a joint analysis of household income and consumption expenditure. This article compares the random hot deck approach proposed by Eurostat with predictive mean matching, an alternative data fusion method. It evaluates the performance of both methods in a simulation study.

1

Einleitung

Zur umfassenderen Beurteilung des sozialen und wirtschaftlichen Lebensstandards privater Haushalte in der Europäischen Union (EU) strebt die amtliche Statistik seit einigen Jahren an, die gemeinsame Verteilung von Einkommen, Konsumausgaben und Vermögen privater Haushalte zu messen. Diese Bestrebungen basieren auf Empfehlungen der Stiglitz-Sen-Fitoussi-Kommission von 2009 (Stiglitz und andere, 2009). Momentan existiert in der amtlichen Statistik jedoch keine befriedigende Datenquelle, die alle drei Komponenten gemeinsam erfasst. Daher sollen entsprechende Datenquellen, die die jeweiligen Determinanten umfassend widerspiegeln, fusioniert werden. Dabei stehen drei Datenquellen zur Verfügung: für Einkommen die Erhebung „European Union Statistics on Income and Living Conditions“ (kurz: EU-SILC), für die Konsumausgaben der „Household Budget Survey“ (kurz: HBS)¹ sowie für das Vermögen der „Household Finance and Consumption Survey“ (kurz: HFCS). Der erste Schritt zu einer einheitlichen Datenquelle ist die initiale Fusionierung von EU-SILC und HBS, also die gemeinsame Betrachtung von Einkommen und Konsumausgaben der privaten Haushalte (Lamarche, 2017).² Damit verbindet die amtliche Statistik zusätzlich das Ziel, Armutsrisiken und Armutsindikatoren präziser abbilden zu können als mit dem bisherigen Messinstrument, das sich weitgehend nur auf das Einkommen selbst beschränkt und die Konsumausgaben außer Acht lässt (Serafino/Tonkin, 2017).

Um den primären und initialen Fokus der amtlichen Statistik aufzugreifen, der sich auf die gemeinsame Verteilung von Einkommen und Konsumausgaben bezieht, untersucht der vorliegende Beitrag die mögliche Datenfusion von EU-SILC und HBS. Er baut auf dem aktuellen Forschungsstand des Statistischen Amtes der Europäischen Union (Eurostat) auf. Um EU-SILC und HBS zu fusionieren, verfolgt Eurostat die Ergänzung des vorhandenen EU-SILC-Datensatzes um die im HBS enthaltenen Konsumvariablen. Daraus soll ein vollständiger,

fusionierter Mikrodatenfile aus Einkommens- und Konsuminformationen resultieren. Daher stellt EU-SILC den Empfänger- beziehungsweise Rezipientendatenfile dar, dem Informationen zu den Konsumausgaben hinzugefügt werden sollen. Der HBS ist wiederum der Spender- beziehungsweise Donorendatenfile, aus dem die entsprechenden Konsuminformationen stammen.

Die von Eurostat vorgeschlagene Datenfusionsmethode entspricht dem Random Hot-Deck-Verfahren, ein nicht-parametrischer Ansatz, welcher über Ähnlichkeitsmaße in den gemeinsamen Variablen maximal ähnliche Beobachtungen zu identifizieren versucht (D’Orazio und andere, 2006, hier: Kapitel 2.4.1; Lamarche, 2017). Predictive Mean Matching (PMM) nach Rubin (1986) und Little (1988) könnte hingegen eine vielversprechende Fusionsalternative darstellen, da PMM ein semi-parametrisches Verfahren ist und Beobachtungen des Rezipienten- und Donorendatensatzes anhand modellbasierter Distanzen verknüpft. Ziel dieses Artikels ist es, die Performance von Random Hot-Deck und PMM hinsichtlich des Erhalts gemeinsamer Verteilungen zwischen den nicht gemeinsam beobachteten Variablen Einkommen und Konsum im Zuge einer Simulationsstudie zu evaluieren.

Der vorliegende Artikel gliedert sich dabei wie folgt: Kapitel 2 bietet zunächst eine kurze Einführung zu Datenfusionen und stellt die relevanten Fusionsalgorithmen, Random Hot-Deck und PMM, allgemein vor. Kapitel 3 skizziert das Simulationsdesign und die Evaluationskriterien, während in Kapitel 4 die Ergebnisse der Simulationsstudie diskutiert werden. Kapitel 5 fasst die Ergebnisse und Implikationen für die amtliche Statistik zusammen.

1 In Deutschland entspricht der HBS der Einkommens- und Verbrauchsstichprobe (kurz: EVS).

2 Beim Artikel von Lamarche (2017) handelt es sich um eine vorläufige Version.

2

Methodischer Überblick und Fusionsalgorithmen

2.1 Datenfusion: eine Einführung

Datenfusionen bezeichnen allgemein gesprochen ein spezifisches Datenausfallmuster, konkret ein Missing-By-Design-Pattern, welches durch „Übereinanderstapeln“ zweier oder mehrerer Datenquellen entsteht (Rässler, 2002; Meinfelder, 2013). Während in Datenfile A und B die X-Variablen gemeinsam beobachtet sind, enthält Datenfile A lediglich Informationen zu Y (jedoch nicht zu Z), dagegen sind in Datenfile B lediglich Informationen zu Z vorhanden (jedoch nicht zu Y). Dementsprechend wird fortlaufend die Menge an gemeinsamen Variablen als X, die Menge an spezifischen, nicht gemeinsam beobachteten Variablen hingegen als Y und Z bezeichnet. Das Analyseziel einer Datenfusion bezieht sich stets auf die gemeinsame Verteilung der ursprünglich nicht gemeinsam beobachteten Merkmale Y und Z (Kiesl/Rässler, 2005). ➔ Grafik 1

Grafik 1

Datenfusion als spezifisches Datenausfallmuster

	X	Y	Z
A	beobachtet	beobachtet	fehlend
B	beobachtet	fehlend	beobachtet

Quelle: Meinfelder (2013), hier: Seite 85

2021 - 0283

Zwar wäre es möglich, wie Grafik 1 suggeriert, beide fehlenden Datenblöcke, Z in A und Y in B, zu ergänzen. Allerdings wird in der Praxis meist lediglich einer der beiden Variablenblöcke mittels Datenfusionsverfahren ergänzt. Hinsichtlich der Datenfusion von EU-SILC und HBS entspricht Datenfile A EU-SILC mit den beobachteten Einkommensinformationen (Y), während Datenfile B den HBS widerspiegelt, welcher die Konsuminformationen (Z) liefert, die in der Rezipientenstichprobe EU-SILC imputiert werden sollen.

Datenfusionen werden mithilfe der Informationen der gemeinsamen Variablen (X) durchgeführt, häufig durch einfache, nicht-parametrische Distanzberechnung (Koschnick, 1995; van der Putten, 2002), wobei ebenfalls semi-parametrische Verfahren (wie etwa PMM) oder sogar rein parametrische Ansätze (etwa basierend auf Regressionen) zum Einsatz kommen können (D’Orazio und andere, 2006, hier: Seite 14 ff.; Gilula und andere, 2006).¹³ Herkömmliche Fusionsverfahren, auch Random Hot-Deck und PMM, unterstellen jedoch implizit die Conditional Independence Assumption, also die Unabhängigkeit von Y und Z gegeben X, woraus $\text{corr}(Y, Z|X) = 0$ resultiert. Diese Annahme der bedingten Unabhängigkeit wurde zuerst von Sims (1972) in einem Kommentar zu Okner (1972) aufgezeigt. Die Problematiken dieser Annahme wurden ausführlich in Rodgers (1984) thematisiert, während in den letzten Jahren auch vermehrt mittels Hilfsinformationen versucht wird, die Conditional Independence Assumption abzuschwächen oder zu umgehen (Singh und andere, 1993; Fosdick und andere, 2016). Kiesl und Rässler (2006) diskutieren hingegen die Möglichkeit, mittels Fréchet-Hoeffding-Grenzen für jedes Variablenpaar von Y und Z eine Ober- und Untergrenze für die marginale, gemeinsame kumulative Verteilungsfunktion zu berechnen. Im Rahmen dieses Beitrags werden Annahmeverletzungen nicht explizit thematisiert. Hinsichtlich des angestrebten Performancevergleichs von Random Hot-Deck und PMM wird demnach die Conditional Independence Assumption unterstellt, da beide Algorithmen bedingte Unabhängigkeit im fusionierten Datenfile herstellen.

2.2 Random Hot-Deck

Nach der kurzen methodischen Einführung in das Datenfusionsthema folgt nun die Vorstellung der beiden relevanten Fusionsalgorithmen, Random Hot-Deck und PMM. Der Random Hot-Deck-Ansatz basiert auf einer zufälligen Zuweisung von Rezipienten- und Donorenbeobachtungen, welche in der Regel innerhalb homogener Teilgruppen durchgeführt wird (D’Orazio und andere, 2006, hier: Kapitel 2.4.1).

3 Einen ausführlichen Überblick klassischer Fusionsalgorithmen liefern Rässler (2002) sowie D’Orazio und andere (2006). Rässler (2002) thematisiert zusätzlich alternative, bayesianische Fusionsalgorithmen, während D’Orazio und andere (2006) auch makrobasierte Ansätze aufgreift.

Konkret stellt sich die von Eurostat vorgeschlagene Random Hot-Deck-Methode nach Lamarche (2017) für jede spezifische, zu fusionierende Variable Z_r wie folgt dar: Zunächst erfolgt eine Kategorisierung aller metrischen Variablen. Anschließend werden mithilfe von Backward-Selection aus einer Regression von Z_r auf X für den Fusionsprozess relevante gemeinsame X -Variablen ausgewählt. Hieraus werden Schichten bestehend aus Zellkombinationen der ausgewählten X -Variablen für jede Beobachtung des Rezipienten- und Donorendatensatzes gebildet. Sofern beispielsweise Alter und Geschlecht als gemeinsame Variablen ausgewählt werden, fasst die Schichtung deren Ausprägungen, also die Zellkombination zusammen (Geschlecht „1“ und Altersgruppe „3“ werden etwa zu „13“). Das Random Hot-Deck, also die zufällige Zuweisung der Rezipienten zu potenziellen Donoren, wird nun innerhalb derselben Schichtungs- ausprägung durchgeführt. Um einen ausreichend großen Donorenpool sicherzustellen, wird ein Schwellwert gesetzt, gemäß dem auf einen Donoren höchstens drei Rezipienten fallen dürfen. Dabei wird für 10 % der Rezipienten eine Abweichung dahingehend toleriert, dass ein Donor auf mehr als drei Rezipienten fällt (Lamarche, 2017).

Sollten Rezipienten übrig bleiben, für die entlang der ausgewählten X -Variablen noch kein merkmalsgleicher Donor zur Verfügung steht, wird die Variablenauswahl so lange weiter iteriert, bis entlang der gebildeten Schichtungs- ausprägungen alle verbliebenen Rezipienten mit mindestens einem Donor übereinstimmen. Dabei wird auf den oben genannten Schwellwert nun verzichtet und die Variablenauswahl in jedem Iterationsschritt um eins reduziert. Sofern alle Rezipienten einem Donor zugewiesen werden konnten, bekommt jede Rezipientenbeobachtung den realen Wert der jeweiligen Z_r -Ausprägung des zugewiesenen Donors. Dieser Fusionsansatz basiert lediglich auf Distanzen innerhalb der X -Variablen, ohne modellbasierte Einbeziehung weiterer Informationen. Daher kann das Random Hot-Deck als nicht-parametrisches Verfahren betrachtet werden (Lamarche, 2017).

2.3 Predictive Mean Matching

Predictive Mean Matching (PMM) ist hingegen ein semi-parametrisches Verfahren mit der Grundidee, für jeden fehlenden Wert einer Variablen ihren prädiktiven Mittelwert, der sich mithilfe einer OLS-Regression berech-

net, mit den prädiktiven Mittelwerten der beobachteten Werte zu vergleichen. Die Beobachtung mit dem ähnlichsten prädiktiven Mittelwert fungiert als Donor, dessen realer Wert sodann den fehlenden Wert ersetzt (Rubin, 1986; Little, 1988).

Konkret kann das PMM-Verfahren für jede spezifische, zu fusionierende Variable Z_r wie folgt beschrieben werden: Zunächst werden, wie beim Random Hot-Deck-Ansatz, für die Fusion relevante X -Variablen mithilfe von Backward-Selection ausgewählt. Hierbei entfällt jedoch die Kategorisierung metrischer Variablen und alle Merkmale können auf ihrem ursprünglichen Skalenniveau verbleiben. Anschließend wird, basierend auf einer Regression von Z_r auf X im Donorendatensatz, für jede Beobachtung im Rezipienten- und Donorendatensatz der prädiktive Mittelwert berechnet. Die Suche nach entsprechenden Donoren erfolgt nun unter Verwendung der von Little (1988) vorgeschlagenen Mahalanobis-Distanz:

$$D_{i,j} = (\hat{z}_i - \hat{z}_j)^T S_{Z|X}^{-1} (\hat{z}_i - \hat{z}_j)$$

mit $i = 1, \dots, n_{\text{silc}}$ und $j = 1, \dots, n_{\text{hbs}}$, wobei \hat{z}_i dem prädiktiven Mittelwert der i -ten Rezipientenbeobachtung und \hat{z}_j dem prädiktiven Mittelwert der j -ten Donorenbeobachtung entspricht (Meinfielder, 2013, hier: Seite 89). $S_{Z|X}^{-1}$ stellt die inverse Varianz-Kovarianzmatrix der Residuen der Regression von Z_r auf X dar, die als Gewichtungsmatrix fungiert. Dabei kommt den gemeinsamen X -Variablen, welche eine gute Erklärungskraft bezüglich der zu fusionierenden Variable Z_r aufweisen, ein stärkerer Einfluss auf die Gesamtdistanz zu als jenen X -Variablen mit geringerer Erklärungskraft. Distanzen werden also bei guter Erklärungskraft stärker, bei schwacher Erklärungskraft weniger stark bestraft (Koller-Meinfielder, 2009, hier: Seite 33 f.; Meinfielder, 2013).

Nach Berechnung der Distanzen $D_{i,j}$ wird der real beobachtete Z_r -Wert des maximal ähnlichen Donors im Rezipientendatensatz imputiert. Sofern ein Rezipient zu mehreren Donoren die minimalste Distanz aufweist, erfolgt eine zufällige Auswahl dieser Donoren. Da PMM Distanzen aufgrund modellbasierter Regressionen in parametrischer Weise bildet, jedoch anschließend den realen Wert in nicht-parametrischer Weise imputiert, kann PMM als semi-parametrisches Verfahren angesehen werden.

2.4 Diskussion beider Fusionsverfahren

Beim methodischen Vergleich beider Ansätze, Random Hot-Deck und PMM, sind insbesondere zwei elementare Unterschiede zu erkennen: Erstens müssen bei Random Hot-Deck alle gemeinsamen X -Variablen kategorisiert werden, was für metrische Variablen einen Informationsverlust induziert, während PMM auch nicht kategoriale X -Variablen verarbeiten kann. Zweitens besteht ein entscheidender Unterschied darin, inwiefern die X -Variablen gemäß ihrer Erklärungskraft auf die zu fusio-nierenden, spezifischen Z -Variablen abgestuft beziehungsweise gewichtet werden. Bei Random Hot-Deck fließt jede gemeinsame, ausgewählte X -Variable mit dem gleichen Gewicht in den Fusionsprozess mit ein, ungeachtet dessen, ob diese X -Variablen untereinander eine unterschiedliche Erklärungskraft auf die zu fusio-nierende Z -Variable aufweisen. Wenn etwa eine durch die Backward-Selection ausgewählte X -Variable im Vergleich zu den anderen, ebenfalls ausgewählten gemeinsamen Variablen die in diesem Artikel thematisierten Konsuminformationen Z schlechter erklären kann, so spielt dies bei der Distanzberechnung beziehungsweise bei der Suche nach Übereinstimmungen keine Rolle. PMM hingegen löst dieses Optimierungsproblem mittels Distanzgewichtung anhand der Inversen der Varianz-Kovarianzmatrix der Residuen einer Regression von Z auf X . Je besser beziehungsweise schlechter dabei die gemeinsamen Variablen die Konsumangaben Z erklären können, desto stärker beziehungsweise schwächer ist ihr Einfluss in der Distanzberechnung (Koller-Meinfelder, 2009, hier: Seite 33 f.; Meinfelder, 2013). Dementsprechend wird aufgrund der nicht notwendigen Kategorisierung gemeinsamer Merkmale sowie der Gewichtung der X -Variablen entlang ihres Erklärungsbeitrags als Hypothese unterstellt, dass PMM ein besseres Fusionsergebnis produziert als Random Hot-Deck.

3

Simulationsdesign

Um die unterstellte Arbeitshypothese zu überprüfen und zu evaluieren, wird eine Simulationsstudie durchgeführt. Grund dafür ist, dass andernfalls keine validen Benchmarks über die wahre Verteilung der jeweils nicht gemeinsam beobachteten Merkmale, Einkommen (Y)

und Konsum (Z) im untersuchten Fall, vorliegen. Dabei soll das Simulationsdesign so realitätsbezogen wie möglich bezüglich der begehrten Datenfusion von EU-SILC und HBS konzipiert werden. Dieses Kapitel stellt die dafür relevante Datengrundlage vor und skizziert Details der Monte-Carlo-Studie.

3.1 Datengrundlage

Als Datengrundlage dienen zugangsbedingt Public Use Files von EU-SILC aus dem Jahr 2013; für eine ausreichend große Datenbasis werden die Datensätze für Deutschland, Frankreich und die Niederlande gemeinsam betrachtet, was einer Simulationsgrundlage von $N=25\,857$ Beobachtungen entspricht. Hieraus werden für jeden in Grafik 1 dargestellten Variablenblock (X, Y, Z) Variablen ausgewählt, die dem realen Fusionsszenario von EU-SILC und HBS nahe kommen. Zu den Konzepten der in diesem Abschnitt spezifizierten Variablen bietet die Dokumentation in Eurostat (2013) umfangreiche Informationen.

Als gemeinsame X -Variablen werden die von Eurostat hinsichtlich der Datenfusion von EU-SILC und HBS bereits für Deutschland, beispielsweise aufgrund von Verteilungsuntersuchungen anhand der Hellinger-Distanz, vorgeschlagenen Merkmale ausgewählt (Leulescu/Agafitei, 2013; Lamarche, 2017). [Übersicht 1](#) zeigt die entsprechenden X -Variablen einschließlich Wertebereich und Skalenniveau. Hierbei ist ersichtlich, dass die metrischen Variablen Alter und Einkommen für Random Hot-Deck kategorisiert werden müssen, während

Übersicht 1

Übersicht der gemeinsamen Variablen

Gemeinsame X -Variablen	Wertebereich/Skalenniveau	
	Eurostat	PMM
X_1 : Activity Status of RP ^{1 2}	1 bis 7 / kategorial	1 bis 7 / kategorial
X_2 : Age of RP ²	1 bis 8 / kategorial	gem. X_2 / metrisch
X_3 : Population Density Level	1 bis 3 / kategorial	1 bis 3 / kategorial
X_4 : Type of Household ^{1 3 4}	1 bis 4 / kategorial	1 bis 4 / kategorial
X_5 : Tenure Status	1 bis 5 / kategorial	1 bis 5 / kategorial
X_6 : Main Source of Income ¹	1 bis 2 / kategorial	1 bis 2 / kategorial
X_7 : Income	1 bis 5 / kategorial	gem. X_7 / metrisch

1 Zusätzlich bilden hier die fehlenden Werte eine Kategorie (kodiert als 9).

2 RP: „Reference Person“ (befragte Person des gezogenen Haushalts).

3 Approximiert durch Variable „Dwelling Type“.

4 Eigentlicher Wertebereich 1 bis 5, Kategorie 5 ist jedoch leer.

Quelle: Public Use Files der Erhebung EU-SILC 2013 für Deutschland, Frankreich und die Niederlande

bei PMM beide Variablen auf dem metrischen Skalenniveau verbleiben können.

Als spezifische Y - und Z -Variablen werden nun Stellvertretermerkmale ausgewählt, welche Einkommen und Konsum approximativ widerspiegeln sollen. Für Einkommen ist dies trivial, da die Datengrundlage auf EU-SILC basiert, worin detaillierte Einkommensmerkmale enthalten sind. Als Substitute für die Konsumvariablen muss hingegen auf ähnliche Variablengegebenheiten geachtet werden; deshalb ist es entscheidend, hierfür metrische Variablen auszuwählen. Die ausgewählten Y - und Z -Variablen zeigt [Übersicht 2](#).

Übersicht 2

Übersicht der spezifischen Variablen für EU-SILC (Y) und Household Budget Survey (Z)

Spezifische EU-SILC-Variablen (Y)	Skalenniveau
Y_1 : Total disposable household income before social transfers including old-age and survivor's benefits	metrisch
Y_2 : Interest, dividends, profit from capital investments in unincorporated business	metrisch
Spezifische HBS-Variablen (Z)	Skalenniveau
Z_1 : Total household gross income	metrisch
Z_2 : Total disposable household income before social transfers other than old-age and survivor's benefits	metrisch

Quelle: Public Use Files der Erhebung EU-SILC 2013 für Deutschland, Frankreich und die Niederlande

Eine inhaltlich exakte Überschneidung mit Konsum kann aufgrund der Datenbeschaffenheit nicht gewährleistet werden. Um die zugrundeliegenden Verfahren valide zu vergleichen, müssen jedoch Informationen über die gemeinsame Verteilung von ursprünglich nicht gemeinsam beobachteten Variablen vorliegen. Daher kann ein reales Fusionszenario nicht als Performanceevaluation für Random Hot-Deck und PMM herangezogen werden. Dementsprechend ist hier simulationsorientiert vorzugehen und es scheint, dass die Z -Variablensubstitute für Konsum tatsächlich Konsuminformationen umfassten, obwohl dies rein logisch nicht der Fall ist. Dies dient jedoch dem Zweck der Evaluation der (sonst unbekannten) gemeinsamen Verteilung von Y und Z .

3.2 Monte-Carlo-Simulation

Die eben beschriebene Datengrundlage stellt die Hilfsdatenbasis für die Monte-Carlo-Simulationsstudie (Morris und andere, 2019) dar. Hieraus werden $k=1\,000$ Zufallsstichproben (ohne Zurücklegen) gezogen. Anschließend wird für jeden Simulationszug das Datenfusionsmuster gemäß Grafik 1 erzeugt und die in Datenfile A (entspricht hier EU-SILC) fehlenden Z_1 - und Z_2 -Werte werden mittels beider Verfahren, Random Hot-Deck und PMM, entsprechend ergänzt.

Für die darauffolgende Evaluation jedes Simulationszugs sind besonders die Korrelationen zwischen den nicht gemeinsam beobachteten Merkmalen Y und Z im fusionierten Datenfile relevant. Denn Datenfusionen werden in der Regel zu genau diesem Zwecke durchgeführt, zur Erfassung und Analyse unbeobachteter Korrelationen jeweils nicht gemeinsam beobachteter Merkmale (Kiesl/Rässler, 2005). Die geschätzten Korrelationen der $k=1\,000$ Zufallszüge werden anschließend jeweils mit den wahren Korrelationen⁴ der Hilfsdatenbasis verglichen, um so die Performance beider Verfahren abschätzen zu können. Der Erhalt der bereits im Spenderdatensatz beobachteten Korrelationen zwischen den gemeinsamen Variablen und den spezifischen Konsumvariablen des HBS gilt als eine Art Mindestanforderung an eine Datenfusion und wird daher ebenso kurz beleuchtet. Um die Sensitivität beider Verfahren auf eine gleich- und übermäßige Anzahl an Beobachtungen des Spenderdatensatzes im Vergleich zu Beobachtungen des Empfängerdatensatzes abzuschätzen, wird die Monte-Carlo-Simulation sowohl mit einem gleichmäßigen Stichprobenverhältnis aus Empfänger- und Spenderbeobachtungen durchgeführt ($n_{silc}=n_{hbs}=300$), als auch unter einem Stichprobenumfang, der neunmal so viele Spender- wie Empfängerbeobachtungen enthält ($n_{silc}=300$ und $n_{hbs}=2\,700$).

Die Monte-Carlo-Simulation wird mit dem Statistikprogramm *R* (R Core Team, 2019) durchgeführt. Die wichtigsten Packages sind StatMatch (D'Orazio, 2020) für Random Hot-Deck sowie BaBooN (Meinfielder/Schnapp, 2015) für Predictive Mean Matching. Das folgende Kapitel diskutiert die Ergebnisse der Monte-Carlo-Simulation.

4 Hier ist zu beachten, dass die „wahre Korrelation“ die Benchmark-Korrelation der in EU-SILC beobachteten Variablensubstitute für Y und Z widerspiegelt, nicht jedoch die tatsächliche, aber unbeobachtete Korrelation zwischen Einkommen und Konsum.

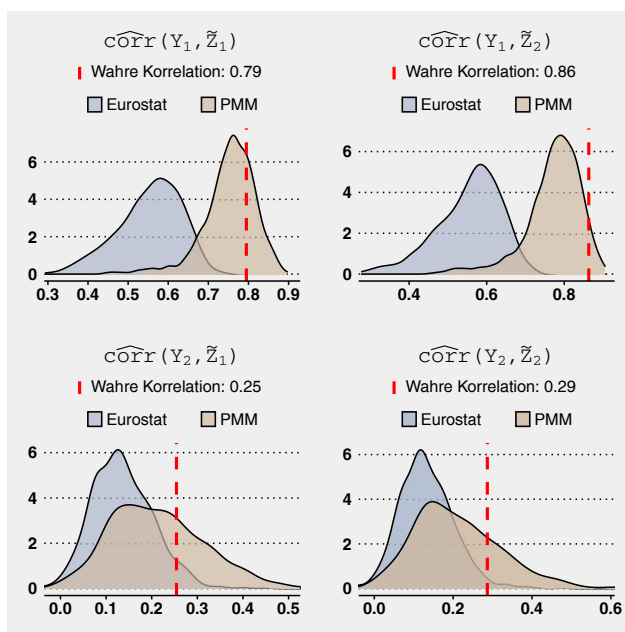
4

Ergebnisse der Monte-Carlo-Simulation

Die Verteilungen der jeweils $k=1000$ geschätzten Korrelationen zeigen unter einem gleichmäßigen Verhältnis an Empfänger- und Spenderbeobachtungen, dass PMM die tatsächlichen Zusammenhänge zwischen den simulierten Einkommens- und Konsumvariablen deutlich besser reproduziert als das Random Hot-Deck-Verfahren. [► Grafik 2](#) Für hohe Originalkorrelationen (hier in Höhe von 0.79 und 0.86) optimiert PMM den Korrelationserhalt erheblich: Während beim Random Hot-Deck die (über alle $k=1000$ Monte-Carlo-Simulationszüge) gemittelten Korrelationsschätzer 0.55 und 0.56 betragen, produziert PMM im Mittel Korrelationsschätzer von 0.75 und 0.77. Damit kommen diese relativ nah an die Originalzusammenhänge von 0.79 und 0.86 heran. Für mittlere Zusammenhänge in der Grundgesamtheit (hier in Höhe von 0.25 und 0.29) produziert PMM durchschnittliche Monte-Carlo-Korrelationen von 0.21 und ebenfalls 0.21, während die Korrelationsschätzer unter Random Hot-

Grafik 2

Dichte der Korrelationen von Y und Z bei gleichmäßigem Empfänger- und Spenderverhältnis

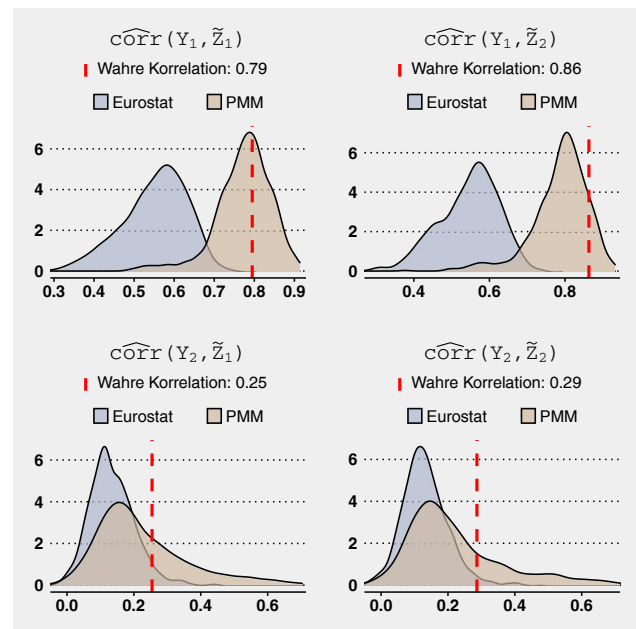


Deck im Mittel 0.14 und 0.13 betragen. Jedoch sind hier die Monte-Carlo-Varianzen für PMM höher und induzieren damit eine stärkere Unsicherheit, wenngleich PMM auch mittlere Originalkorrelationen besser abdeckt. Bei einer übermäßigen Anzahl an Spenderbeobachtungen zeigt sich ein kaum abweichendes Ergebnis, was auf eine geringe Sensitivität beider Verfahren mit Blick auf das Empfänger- und Spenderverhältnis hindeutet.

[► Grafik 3](#)

Grafik 3

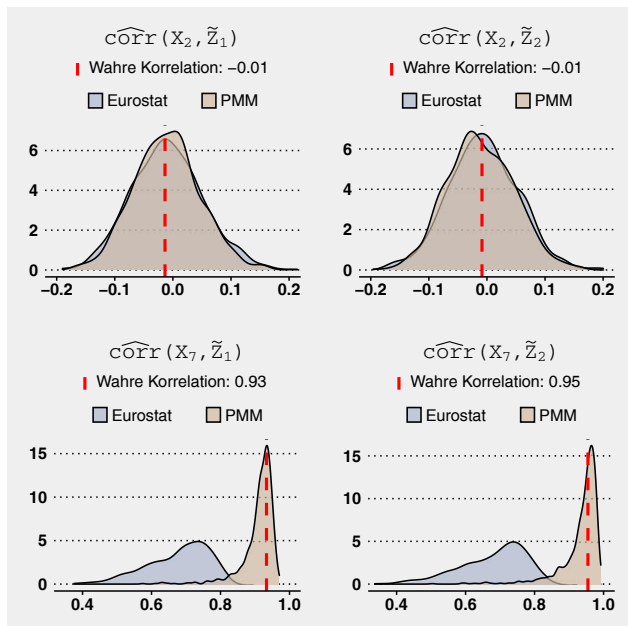
Dichte der Korrelationen von Y und Z bei übermäßiger Donorenanzahl



Die Reproduktion der bereits im Spenderdatensatz beobachteten Korrelationen zwischen X und Z stellt eine Art Mindestanforderung an eine Datenfusion dar (Kies/Rässler, 2005). Hier wird ersichtlich, dass PMM sowohl bei gleichmäßigem Empfänger- und Spenderverhältnis als auch bei einer übermäßigen Donorenanzahl dieser Mindestanforderung deutlich besser nachkommt als das Random Hot-Deck-Verfahren. Allerdings kann dies aufgrund der Datenbeschaffenheit lediglich für hohe Originalkorrelationen (hier in Höhe von 0.93 und 0.95) untersucht werden. Originalzusammenhänge nahe 0 (wie sie in diesem Fall in Höhe von -0.01 vorliegen) eignen sich hingegen für keinen Performancevergleich, da die Kor-

Grafik 4

Dichte der Korrelationen von X und Z bei gleichmäßigem Empfänger- und Spenderverhältnis



Quelle: Public Use Files der Erhebung EU-SILC 2013 für Deutschland, Frankreich und die Niederlande

PMM: Predictive Mean Matching

2021 - 0286

relationsschätzung eher einem Zufallsereignis gleicht. Deshalb erzielen hier Random Hot-Deck und PMM aggregiert relativ ähnliche Ergebnisse. ➡ Grafik 4, Grafik 5

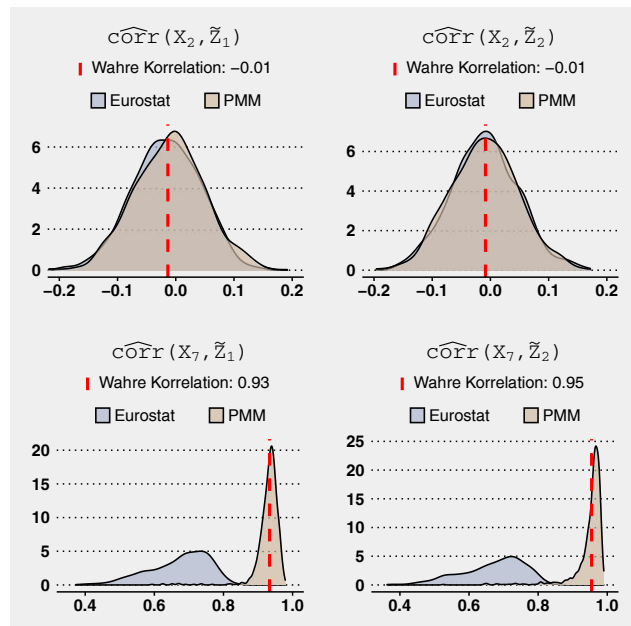
5

Fazit

Für die validere und umfassendere Beurteilung des sozialen und wirtschaftlichen Lebensstandards sowie zur präziseren Abbildung von Armutsindikatoren strebt die amtliche Statistik die gemeinsame Betrachtung von Einkommen und Konsumausgaben privater Haushalte an. Beide Informationen sind über zwei Datensätze, EU-SILC und Household Budget Survey, verteilt, weshalb beide Datenquellen mittels einer geeigneten Datenfusionsmethode verbunden werden sollen. Die Ergebnisse der Simulationsstudie zeigen, dass die nicht gemeinsam beobachtete Verteilung von Einkommen und Konsum sowie im Besonderen deren Zusammenhangsstruktur mittels Predictive Mean Matching präziser

Grafik 5

Dichte der Korrelationen von X und Z bei übermäßiger Donorenanzahl



Quelle: Public Use Files der Erhebung EU-SILC 2013 für Deutschland, Frankreich und die Niederlande


PMM: Predictive Mean Matching

2021 - 0287

reproduziert werden kann als mit der bisher verwendeten Random Hot-Deck-Methode.

Simulationsstudien können allerdings keine Allgemeingültigkeit gewährleisten. Dennoch ergibt sich aus den Ergebnissen für die amtliche Statistik die Erkenntnis, dass PMM zumindest für vergleichbare Datensituationen (also hinsichtlich einer Fusionierung von stetigen Variablen) eine vielversprechende Fusionsalternative zu herkömmlichen, nicht-parametrischen Verfahren darstellt. Eine Schwierigkeit in sämtlichen Datenfusions-szenarien ist hingegen die Annahme der bedingten Unabhängigkeit. Sie muss unterstellt werden, weil Datenfusionsverfahren (auch Random Hot-Deck und PMM) bedingte Unabhängigkeit im fusionierten Datenfile herstellen. Künftige Forschungsarbeiten könnten zur Bewältigung dessen die adäquate Einbeziehung von Hilfsinformationen aufgreifen.

Die Erkenntnisse des vorliegenden Beitrags kann die amtliche Statistik für künftige Datenfusionsvorhaben nutzen. Denn die amtliche Statistik ist heute mehr denn je bestrebt, lange Fragebogen zu vermeiden – zugun-

ten der Entlastung der Auskunftgebenden und einer höheren Datenqualität. Dennoch soll es möglich sein, aus den Datenbeständen ein möglichst hohes Analysepotenzial zu generieren, um den steigenden amtlichen Daten- und Analysebedarf zu decken. Die Implementierung eines zielführenden Datenfusionsverfahrens spielt daher auch in Zukunft eine wichtige Rolle. 

LITERATURVERZEICHNIS

- D’Orazio, Marcello/Di Zio, Marco/Scanu, Mauro. *Statistical Matching. Theory and Practice*. 2006. [Zugriff am 6. Juli 2021]. Verfügbar unter: epdf.pub
- D’Orazio, Marcello. *StatMatch: Statistical Matching or Data Fusion. R-Package. Version 1.4.0*. 2020. [Zugriff am 6. Juli 2021]. Verfügbar unter: cran.r-project.org
- Eurostat. *Description of Target Variables: Cross-sectional and Longitudinal*. EU-SILC 065, 2013 operation (Version May 2013). 2013. [Zugriff am 6. Juli 2021]. Verfügbar unter: circabc.europa.eu
- Fosdick, Bailey K./DeYoreo, Maria/Reiter, Jerome P. *Categorical Data Fusion Using Auxiliary Information*. In: The Annals of Applied Statistics. Jahrgang 10. Ausgabe 4/2016, Seite 1907 ff. [Zugriff am 6. Juli 2021]. Verfügbar unter: www.jstor.org
- Gilula, Zvi/McCulloch, Robert E./Rossi, Peter E. *A Direct Approach to Data Fusion*. In: Journal of Marketing Research. Jahrgang 43. Ausgabe 1/2006, Seite 73 ff. [Zugriff am 6. Juli 2021]. Verfügbar unter: asu.pure.elsevier.com
- Kiesl, Hans/Rässler, Susanne. *Techniken und Einsatzgebiete von Datenintegration und Datenfusion*. In: König, Christian/Stahl, Matthias/Wiegand, Erich (Herausgeber). Datenfusion und Datenintegration, 6. wissenschaftliche Tagung. Bonn 2005, Seite 17 ff. [Zugriff am 6. Juli 2021]. Verfügbar unter: www.iab.de
- Kiesl, Hans/Rässler, Susanne. *How Valid Can Data Fusion Be?* In: IAB Discussion Paper. Ausgabe 15/2006. [Zugriff am 6. Juli 2021]. Verfügbar unter: doku.iab.de
- Koller-Meinfelder, Florian. *Analysis of Incomplete Survey Data – Multiple Imputation via Bayesian Bootstrap Predictive Mean Matching*. Dissertation. Otto-Friedrich-Universität Bamberg. 2009. [Zugriff am 6. Juli 2021]. Verfügbar unter: fis.uni-bamberg.de
- Koschnick, Wolfgang J. *Standard-Lexikon für Markt- und Konsumforschung*. Band 1, A – K. München 1995.
- Lamarche, Pierre. *Measuring Income, Consumption and Wealth jointly at the micro-level*. 2017. [Zugriff am 6. Juli 2021]. Verfügbar unter: ec.europa.eu
- Leulescu, Aura/Agafitei, Mihaela. *Statistical matching: a model based approach for data integration*. In: Eurostat. Methodologies & Working papers. 2013. [Zugriff am 6. Juli 2021]. Verfügbar unter: ec.europa.eu
- Little, Roderik J. A. *Missing-Data Adjustments in Large Surveys*. In: Journal of Business & Economic Statistics. Jahrgang 6. Ausgabe 3/1988, Seite 287 ff. [Zugriff am 6. Juli 2021]. Verfügbar unter: www.tandfonline.com
- Meinfelder, Florian. *Datenfusion: Theoretische Implikationen und praktische Umsetzung*. In: Riede, Thomas/Bechtold, Sabine/Ott, Notburga (Herausgeber). Weiterentwicklung der amtlichen Haushaltsstatistiken. Auflage 1. 2013. Seite 83 ff. [Zugriff am 6. Juli 2021]. Verfügbar unter: www.konsortswd.de

LITERATURVERZEICHNIS

Meinfelder, Florian/Schnapp, Thorsten. *Package BaBooN: Bayesian Bootstrap Predictive Mean Matching – Multiple and Single Imputation for Discrete Data. R-Package. Version 0.2-0*. 2015. [Zugriff am 6. Juli 2021]. Verfügbar unter: cran.r-project.org

Morris, Tim P./White, Ian R./Crowther, Michael J. *Using simulation studies to evaluate statistical methods*. In: Statistics in Medicine. Jahrgang 38. Ausgabe 11/2019, Seite 1 ff. [Zugriff am 6. Juli 2021]. Verfügbar unter: onlinelibrary.wiley.com

Okner, Benjamin A. *Constructing a New Data Base from Existing Microdata Sets: The 1966 Merge File*. In: Annals of Economic and Social Measurement. Jahrgang 1. Ausgabe 3/1972, Seite 325 ff. [Zugriff am 6. Juli 2021]. Verfügbar unter: www.nber.org

R Core Team. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Wien 2019. [Zugriff am 6. Juli 2021]. Verfügbar unter: <https://www.R-project.org/>

Rässler, Susanne. *Statistical Matching. A Frequentist Theory, Practical Applications, and Alternative Bayesian Approaches*. New York 2002.

Rodgers, Willard. L. *An Evaluation of Statistical Matching*. In: Journal of Business & Economic Statistics. Jahrgang 2. Ausgabe 1/1984, Seite 91 ff.

Rubin, Donald B. *Statistical Matching Using File Concatenation With Adjusted Weights and Multiple Imputations*. In: Journal of Business & Economic Statistics. Jahrgang 4. Ausgabe 1/1986, Seite 87 ff. [Zugriff am 6. Juli 2021]. Verfügbar unter: www.jstor.org

Serafino, Paola/Tonkin, Richard. *Statistical matching of European Union statistics on income and living conditions (EU-SILC) and the household budget survey*. In: Statistisches Amt der Europäischen Union. Statistical working papers. Doi: 10.2785/933460. 2017. [Zugriff am 6. Juli 2021]. Verfügbar unter: ec.europa.eu

Sims, Christopher A. *Comments (zu Okner 1972)*. In: Annals of Economic and Social Measurement. Jahrgang 1. Ausgabe 3/1972, Seite 343 ff.

Singh, A. C./Mantel, H. J./Kinack M. D./Rowe, G. *Statistical Matching: Use of Auxiliary Information as an Alternative to the Conditional Independence Assumption*. In: Survey Methodology. Jahrgang 19. Ausgabe 1/1993, Seite 59 ff. [Zugriff am 6. Juli 2021]. Verfügbar unter: www150.statcan.gc.ca

Stiglitz, Joseph E./Sen, Amartya/Fitoussi, Jean-Paul. *Report by the Commission on the Measurement of Economic Performance and Social Progress*. 2009. [Zugriff am 6. Juli 2021]. Verfügbar unter: www.economie.gouv.fr

Van der Putten, Peter/Kok, Joost N./Gupta, Amar. *Data Fusion Through Statistical Matching*. In: MIT Sloan School of Management. Working Paper 4342-02. 2002. [Zugriff am 6. Juli 2021]. Verfügbar unter: liacs.leidenuniv.nl

NEUKONZEPTION DER BODENMARKTSTATISTIKEN

Eva-Maria Diehl-Wolf

➤ **Schlüsselwörter:** Preisstatistik – Bodenmarkt – Bauland – Landwirtschaftliche Grundstücke – Kaufwerte

ZUSAMMENFASSUNG

Die Bodenmarktstatistiken — das sind die Statistiken der Kaufwerte für Bauland und der Kaufwerte für landwirtschaftliche Grundstücke — bestehen seit mehreren Jahrzehnten. Neue europäische Anforderungen im Agrarstatistiksektor haben zu einer Überarbeitung sowohl der Abgrenzungen als auch der Merkmalskataloge beider Statistiken geführt. Ab dem Berichtsjahr 2021 erfolgt die Meldung der Transaktionsfälle von Grundstücken zudem erstmals über ein Onlineverfahren. Die gemeldeten Fälle werden standardisiert in einem gemeinsamen Fachverfahren aufbereitet, anschließend die Geheimhaltung maschinell durchgeführt und schließlich automatisiert in Veröffentlichungstabellen überführt. Der vorliegende Aufsatz beschreibt die Überarbeitung der beiden Statistiken und beleuchtet die sich daraus ergebenden Informationspotenziale.

➤ **Keywords:** price statistics – land market – building land – agricultural land – purchase values

ABSTRACT

Land market statistics, i. e. the statistics of purchase values of building land and purchase values of agricultural land, have existed for several decades. New European requirements for the sector of agricultural statistics have now led to a revision of both the delimitations and the lists of variables of the two sets of statistics. From the reference year 2021 onwards, transactions of land will be reported for the first time via an online procedure. The transactions reported will be processed in a standardised way, using a specialised IT application for all the statistical offices. Automatic processes will then be implemented to ensure data confidentiality and to transfer the data into publication tables. This article describes the revision of the two sets of statistics and highlights the resulting information potential.



Eva-Maria Diehl-Wolf

ist Diplom-Sozialwissenschaftlerin und seit 2015 Referentin im Referat „Agrar-, Bau- und Immobilienpreise“ des Statistischen Bundesamtes. Sie befasst sich vor allem mit der Weiterentwicklung verschiedener zentraler und dezentraler Preisstatistiken, aktuell in den Bereichen Bodenmarkt und Landwirtschaft.

1

Neue Anforderungen an die Bodenmarktstatistiken

Fast täglich berichten die Medien von steigenden Immobilienpreisen und beziehen sich dabei häufig auf Daten des Statistischen Bundesamtes. Das Interesse an solchen Daten besteht schon seit Jahrzehnten, wenngleich nicht immer so stark ausgeprägt wie heute. Zwei Statistiken mit einer weit zurückreichenden Geschichte sind die Statistik der Kaufwerte für Bauland (KWB-Statistik) und die Statistik der Kaufwerte für landwirtschaftliche Grundstücke (KWL-Statistik). Sie geben als eine Art „Grundstückswechselstatistiken“ einen Überblick über den Markt für Bauland beziehungsweise landwirtschaftliche Grundstücke in Deutschland. Die Statistischen Ämter der Länder weisen die Merkmale und Strukturen der im Berichtszeitraum vollzogenen Transaktionen zu den entsprechenden Grundstücken für die jeweiligen Bundesländer und für tiefer regionalisierte Ebenen aus. Das Statistische Bundesamt erstellt das Bundesergebnis.

Bis zum Jahr 1960 gab es im früheren Bundesgebiet eine Preisbindung für Baulandgrundstücke. Nach Aufhebung dieser Regelung entstand schnell Bedarf an einer Statistik über Baulandpreise (Statistisches Bundesamt, 1963). Um auch die Preisentwicklungen bei Käufen landwirtschaftlicher Flächen zu erfassen, wurde im Jahr 1973 die Statistik der Kaufwerte für landwirtschaftliche Grundstücke eingeführt.

In beiden Statistiken erfolgten in unregelmäßigen Abständen kleinere Anpassungen, beispielsweise um neue Veröffentlichungspositionen, jedoch keine grundlegenden Überarbeitungen. Eine neue europäische Rechtsverordnung zu Statistiken über landwirtschaftliche Betriebsmittel und die landwirtschaftliche Erzeugung (SAIO: Statistics on Agricultural Input and Output) soll auch Landwirtschaftspreise – und damit Kaufwerte für landwirtschaftliche Grundstücke – regeln. Sie legt sowohl die definitorische Grundlage der Flächen als auch die Kauffallabgrenzung neu fest und soll noch 2021 in Kraft treten. Daher ist die Statistik der Kaufwerte für landwirtschaftliche Grundstücke an die neuen europäischen Vorgaben anzupassen. In diesem Zusammenhang haben die Statistischen Ämter des Bundes

und der Länder ebenfalls eine Überarbeitung der KWB-Statistik angestoßen.

Die Neukonzeption der Bodenmarktstatistiken spiegelt sich auch im novellierten Preisstatistikgesetz wider, das in § 7 Nr. 1 wichtige Merkmale dieser Statistiken rechtlich regelt:

«Die Statistik nach § 2 Nr. 5 [Preise für Grundstücke, Gebäude und Wohnungen] erfasst

1. die Preise für nach Arten und Merkmalen bezeichnete Grundstücke, Gebäude und Wohnungen,
2. Angaben darüber, ob es sich bei den Käuferinnen und Käufern sowie den Verkäuferinnen und Verkäufern jeweils um natürliche Personen, juristische Personen des öffentlichen und des Privatrechts handelt,
3. die Angabe über das Vorhandensein einer familiären Beziehung zwischen den Käuferinnen und Käufern und den Verkäuferinnen und Verkäufern sowie
4. für die Kaufwerte landwirtschaftlicher Grundstücke zusätzlich die Angabe darüber, ob es sich bei den Käuferinnen und Käufern sowie bei den Verkäuferinnen und Verkäufern um eine Landwirtin oder einen Landwirt oder eine Nichtlandwirtin oder einen Nichtlandwirt handelt.»

Bis das novellierte Preisstatistikgesetz Anfang 2020 in dieser Form in Kraft trat, war in § 7 Nr. 1 Preisstatistikgesetz lediglich geregelt, dass die Statistiken über Grundstücke die Preise für nach Arten und Merkmalen bezeichnete Grundstücke erfassen.

Der folgende Beitrag stellt die Neuerungen der Kaufwertestatistiken vor: Kapitel 2 beschreibt die Neukonzeption im Einzelnen, Kapitel 3 stellt die künftige Datenbereitstellung und potenzielle Auswertungsmöglichkeiten dar. Das letzte Kapitel gibt einen Ausblick auf weitere Entwicklungsmöglichkeiten der Bodenmarktstatistiken.

2

Neukonzeption der Bodenmarktstatistiken

Bereits im Jahr 2003 hat der Statistische Verbund¹ die Entscheidung getroffen, den Produktionsprozess der amtlichen Statistik in Deutschland zu standardisieren (Blumöhr und andere, 2017). Demnach sind die Statistischen Ämter des Bundes und der Länder angehalten, bestimmte IT-Werkzeuge und Schnittstellen zu nutzen, wo immer dies möglich ist (Standardisierung der Prozesse, kurz: SteP). Diese Werkzeuge stellen die einheitliche Bearbeitung in allen statistischen Ämtern sicher. So erfolgte auch bei der Neukonzeption der Kaufwertestatistiken eine Prüfung, ob und wie die zur Verfügung stehenden SteP-Werkzeuge implementiert werden konnten – ein wichtiger Schritt in Richtung Vereinheitlichung der beiden Statistiken. Neben den notwendigen inhaltlichen Anpassungen sollte eine Online-Erhebung etabliert, ein gemeinsames Fachverfahren entwickelt und eine koordinierte maschinelle Geheimhaltung wie auch Tabellierung mit abgestimmtem Veröffentlichungsprogramm umgesetzt werden. Einen Überblick über den neuen Ablauf der Bodenmarktstatistiken gibt [Grafik 1](#).

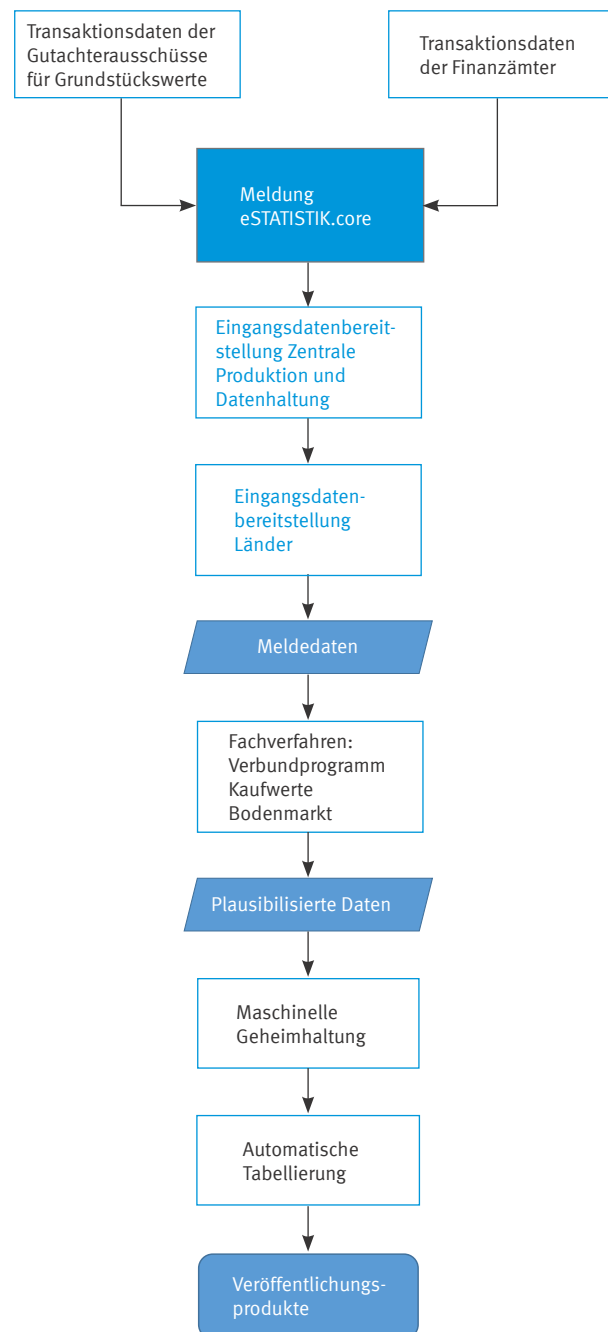
2.1 Einführung einer Online-Erhebung

Jeder Kaufakt einer Immobilie muss laut § 311b des Bürgerlichen Gesetzbuchs mittels eines Kaufvertrags notariell beurkundet werden. Diese Kaufverträge werden durch die Notare in Kopie an das zuständige Finanzamt zur Feststellung der Grunderwerbsteuer übermittelt. Darüber hinaus erhält auch der am Ort des veräußerten Grundstücks ansässige Gutachterausschuss für Grundstückswerte eine Kopie des Vertrags. Die Gutachterausschüsse führen Kaufpreissammlungen und werten diese für die Festlegung von Bodenrichtwerten und als Grundlage für gutachterliche Wertermittlungen sowie allgemein für die Dokumentation wesentlicher Entwicklungen am Grundstücksmarkt aus. Beide Institutionen stellen die Berichtsstellen der Bodenmarktstatistiken dar.

¹ Der Begriff Statistischer Verbund bezeichnet die Statistischen Ämter des Bundes und der Länder.

Grafik 1

Ablauf der Statistikproduktion der Bodenmarktstatistiken ab dem Berichtsjahr 2021



2021 - 0288

Die Hauptanforderung an eine standardisierte Online-Erhebung bei den Kaufwertestatistiken war, den Umstellungsprozess, der teilweise noch die Ablösung von Papiermeldungen enthielt, überschaubar zu halten,

um die Berichtspflichtigen möglichst wenig zusätzlich zu belasten beziehungsweise bestenfalls zu entlasten. Auch die künftige Datenabgabe sollte einfach möglich sein. Ab dem Berichtsjahr 2021 werden hauptsächlich Gutachterausschüsse als Meldestellen herangezogen. Diese sind seit einigen Jahren bereits für die zentrale Statistik des Häuserpreisindex berichtspflichtig (siehe unter anderem Schöneich/Teske, 2020). Bei dieser Statistik kommt für die Datenabgabe das Meldeverfahren eSTATISTIK.core zum Einsatz, das sollte nach Möglichkeit auch für die Kaufwertestatistiken genutzt werden (Statistische Ämter des Bundes und der Länder, 2014). Hier erstellen die Berichtsstellen entweder eine CSV-Datei nach einer festen Datensatzbeschreibung und laden diese über eine bereitstehende Webanwendung hoch oder sie nutzen eine Schnittstelle, die aus den digital vorliegenden Daten automatisch eine XML-Datei erstellt und automatisiert übermittelt. Aufgrund der Struktur der Transaktionen, die die Basis der Bodenmarktstatistiken bilden, sollten die Gutachterausschüsse und Finanzämter nicht jeden einzelnen Kauffall in einen Online-Fragebogen eintragen müssen, was den Ausschlag für .CORE gab. Seit dem Beginn des Berichtsjahres 2021 übermitteln die Berichtsstellen nun auf diesem Weg.

2.2 Neue Abgrenzungen und Merkmale

Grund für die Überarbeitung der KWL-Statistik war die Vorbereitung auf die mit der europäischen SAIO-Rahmenverordnung entstehenden Lieferverpflichtungen. Diese Rahmenverordnung wird nicht nur den landwirtschaftlichen Bodenmarkt, sondern alle Agrarstatistiken umfassen. Daher findet künftig die in diesen Statistiken übliche Abgrenzung der landwirtschaftlich genutzten Fläche (LF) auch bei der KWL-Statistik Anwendung. Im Unterschied zur bisher genutzten Fläche der landwirtschaftlichen Nutzung umfasst die LF auch Gartenbauflächen. Entsprechende Grundstücke werden also bei der KWL-Statistik künftig ebenfalls einbezogen. Bei Gartenbauflächen ist aufgrund ihrer besonderen Beschaffenheit eine andere Preissetzung als beim klassischen Acker- oder Grünland zu erwarten.

Um die Preisverhältnisse der verschiedenen Flächenarten später darstellen zu können, wurde der Merkmalskatalog neben der Gesamtfläche um die Darunterpositionen der Flächen des Ackerlands, Grünlands und der sonstigen LF erweitert. Darüber hinaus wurde die

Erhebung von Informationen über Verkäuferin/Verkäufer und Erwerberin/Erwerber um den Status Landwirtin/Landwirt beziehungsweise Nichtlandwirtin/-landwirt ergänzt. Hierdurch soll es möglich sein, außersektorale Veräußerungsfälle zu identifizieren. Der rechtliche Status wird weiterhin unterteilt nach Privatpersonen, juristischen Personen des privaten sowie des öffentlichen Rechts erfasst. Neu erfragte Merkmale sind das Vorhandensein einer familiären Beziehung zwischen den Vertragsparteien sowie die geplante künftige landwirtschaftliche Nutzung. Zudem sollen die Meldenden ab dem Berichtsjahr 2021 angeben, ob es sich bei der Transaktion um einen ungewöhnlichen Geschäftsverkehr handelt. Dieser kann beispielsweise bei besonderen Zuschnitten des Grundstücks, aber auch aufgrund der bislang nicht zu erfassenden Eigentumsübergänge aufgrund eines Tauschs, einer Schenkung, Vererbung, Flurbereinigung, Enteignung oder Zwangsversteigerung vorliegen. Bisher wurden diese Fälle von den Gutachterausschüssen beziehungsweise den Finanzämtern als solche eingestuft und nicht an die zuständigen Statistischen Landesämter übermittelt. Künftig sollen auch diese entsprechend gekennzeichneten Transaktionen gemeldet werden, um ihre Bedeutung für den Gesamtmarkt beurteilen zu können. Einen Überblick über die bisherigen und neuen Merkmale gibt [Übersicht 1](#).¹² Die bisher auszuschließenden Grundstücke¹³ sollen auch nach der neuen Abgrenzung nicht in die Statistik einfließen. Nach wie vor soll die Größe des veräußerten Grundstücks mindestens 0,1 Hektar umfassen.

Bei der KWB-Statistik wurden vorrangig begriffliche Anpassungen vorgenommen beziehungsweise Merkmalsausprägungen vereinfacht. So wurden die Rechtsformen von Veräußerin/Veräußerer beziehungsweise Erwerberin/Erwerber auf natürliche und juristische Personen des öffentlichen und privaten Rechts zusammengefasst. Das Vorhandensein eines Verwandtschaftsverhältnisses wurde um die Ehepartnerin/den Ehepartner erweitert und entspricht damit der familiären

2 Die genauen Merkmalsbezeichnungen können sich unterscheiden und werden in den Übersichten 1 und 2 allgemeiner beziehungsweise vereinfacht dargestellt.

3 Bei der bisherigen KWL-Statistik ausgeschlossen waren Transaktionen von Baulandgrundstücken, Verkäufe von landwirtschaftlichen Hofstellen, Verkäufe von lebendem oder totem Inventar, Eigentumsübergänge aufgrund eines Tauschs, einer Schenkung, Vererbung, Flurbereinigung, Enteignung oder Zwangsversteigerung, Grundstückserwerbe unter besonderen Bedingungen entsprechend dem Grunderwerbsteuergesetz sowie preisbegünstigte Grundstücksveräußerungen im Sinne der Flächenerwerbsverordnung.

Übersicht 1

Merkmalkatalog der Statistik der Kaufwerte für landwirtschaftliche Grundstücke (KWL-Statistik)

Bis Berichtsjahr 2020	Ab Berichtsjahr 2021
Regionalangaben	
Kaufdatum	
Rechtsform Veräußerer/Veräußerin	
Rechtsform Erwerber/Erwerberin	
	Status Landwirt Veräußerer/Veräußerin
	Status Landwirt Erwerber/Erwerberin
	Familiäre Beziehung
Art des veräußerten Grundstücks	
Gesamtfläche	
	Fläche Ackerland
	Fläche Grünland
Fläche der landwirtschaftlichen Nutzung	Sonstige landwirtschaftlich genutzte Fläche
Geldleistung	Kaufpreis
Übernommene Hypotheken usw.	
Wert Tauschgrundstück	
Sonstige Leistungen	
Summe Gegenleistungen	
	Gesamtpreis für mehrere Flurstücke ja/nein
	Umfang Transaktion
	Künftige landwirtschaftliche Nutzung
Ertragsmesszahl	
	Ungewöhnlicher Geschäftsverkehr

Beziehung. Bei der Art des Grundstücks fallen baureifes Land und Rohbauland nun unter den Oberbegriff Wohnbauland. Das Industrieland wurde umbenannt in „Wirtschaftlich genutztes Bauland“, Land für Verkehrszwecke und Freiflächen werden gemeinsam mit allen Baugrundstücksarten, die nicht unter die vorgegebenen Kategorien fallen, subsumiert. Auf den ersten Blick scheint das Merkmal Art der Baufläche neu zu sein. Hierunter verbirgt sich die ehemalige Art des Baugebiets. Das Merkmal entspricht somit nun den Begrifflichkeiten der Baunutzungsverordnung (Verordnung über die bauliche Nutzung der Grundstücke). Die Grundstücksfläche wird auch weiterhin erhoben. Der Kaufpreis kann künftig auf zwei Arten übermittelt werden: als Vertragspreis, der dem Kaufvertrag entstammt, und/oder als bereinigter Preis. Während der Vertragspreis einen Gesamtpreis für das Grundstück einschließlich weiterer Wertgegenstände, Rechte und/oder Pflichten darstellt, ist im berei-

nigten Preis ausschließlich der reine Grundstückspreis enthalten. Weisen Preise eine ungewöhnliche Struktur auf, können ab dem Berichtsjahr 2021 der beitrags- und abgabenrechtliche Zustand sowie der Bodenrichtwert als Plausibilisierungsmerkmale herangezogen werden. Wie auch bei der KWL-Statistik sollen bei der KWB-Statistik ungewöhnliche Geschäftsverkehre künftig durch die Berichtsstellen gemeldet werden, jedoch als solche gekennzeichnet. An den bisherigen Erhebungsabgrenzungen ändert sich nichts. Die Gemeinsamkeiten und Unterschiede in den Erhebungsmerkmalen fasst [Übersicht 2](#) zusammen.

Übersicht 2

Merkmalkatalog der Statistik der Kaufwerte für Bauland (KWB-Statistik)

Bis Berichtsjahr 2020	Ab Berichtsjahr 2021
Regionalangaben	
Kaufdatum	
Rechtsform Veräußerer/Veräußerin	
Rechtsform Erwerber/Erwerberin	
Verwandtschaftsverhältnis	Familiäre Beziehung
Art des Grundstücks	
Art des Baugebiets	Art der Baufläche
Kaufpreis	Vertragspreis
	Bereinigter Preis
Grundstücksfläche	
	Beitrags- und abgabenrechtlicher Zustand
	Bodenrichtwert
	Ungewöhnlicher Geschäftsverkehr

Mit der Überarbeitung der Kaufwertestatistiken wurden zum einen die Begrifflichkeiten dieser beiden Erhebungen angeglichen. Zum anderen wurde – wo sinnvoll – auch die Begrifflichkeit beim Häuserpreisindex berücksichtigt, der oft gemeinsam mit der KWB-Statistik für die Beurteilung des Immobilienmarkts durch Nutzerinnen und Nutzer verwendet wird.

2.3 Neuentwicklung eines Fachverfahrens

Bereits zu Beginn der Arbeiten zur Neukonzeption der Kaufwertestatistiken stand fest, dass ein gemeinsames Fachverfahren entwickelt werden sollte, um die Melde-daten so effizient wie möglich zu verarbeiten und im Verbund nach einheitlichen Kriterien zu gestalten. Zudem wurde eine Standardisierung der Plausibilitätsprüfungen angestrebt.

Das Verbundprogramm „Kaufwerte Bodenmarkt“, kurz KWBo, wurde als Anwendung mit je einem Programmteil für die beiden Kaufwertestatistiken entwickelt. Die Bearbeiterinnen und Bearbeiter in den Statistischen Landesämtern arbeiten dabei auf der gleichen Oberfläche, die mit statistikspezifischen Aufbereitungsschritten ausgestattet ist.

Die Daten der Gutachterausschüsse und Finanzämter gehen über das Online-Meldeverfahren eSTATISIK.core ein und werden von dort an die zentrale Eingangsdatenbank des Landes weitergeleitet, das für die Bodenmarktstatistiken die Produktion und Datenhaltung zentral hostet (ZPD). Von dort gelangen sie in die Eingangsdatenbanken der einzelnen Statistischen Landesämter (siehe Grafik 1). Über eine Importschnittstelle werden die Kauffälle sodann in das KWBo-Programm geladen. Hier startet die Aufbereitung der Einzeldaten. Zur Prüfung auf Plausibilität wurden weitgehend verbundweit gültige Parameter definiert. Darunter fallen Prüfungen auf formale Korrektheit, beispielsweise Datumsangaben oder die Existenz des Gemeindeschlüssels. Darüber hinaus werden die Merkmalsausprägungen auf Sinnhaftigkeit geprüft. Dies umfasst unter anderem die Kombination bestimmter Merkmale, beispielsweise ob die Erwerberin/der Erwerber Landwirtin/Landwirt ist, aber künftig keine landwirtschaftliche Nutzung der Fläche vorgesehen ist. Um auf regionale Unterschiede in den erzielten Kaufpreisen einzugehen, werden zudem jährlich Ober- und Untergrenzen der Quadratmeter- (KWB) beziehungsweise Hektarpreise (KWL) auf Kreisebene ermittelt und im Programm hinterlegt. Das Statistische Bundesamt verwaltet die benötigten Metadaten (Plausibilitätsprüfungen, Gemeindeverzeichnis, Ober- und Untergrenzen sowie Gemeindeschlüssel-Berichtsstellenlisten) als Leitdateien. So wird sichergestellt, dass alle Statistischen Landesämter mit einem identischen Stand arbeiten. Sofern vordefinierte Fehler der Plausibilitätsprüfungen erkannt werden, obliegt es den Bearbei-

terinnen und Bearbeitern, den Fall durch Rückfragen bei den Berichtsstellen zu klären oder den Fehler auf Basis der vorliegenden Informationen zu bereinigen.

Ist die Datenaufbereitung abgeschlossen, exportiert das Statistische Bundesamt die plausibilisierten Einzeldaten, um sie für den nächsten Verarbeitungsschritt, die statistische Geheimhaltung, bereitzustellen.

3

Künftige Datenbereitstellung und potenzielle Auswertungsmöglichkeiten

3.1 Einführung einer koordinierten Geheimhaltung

Die Geheimhaltungspflicht in der Bundesstatistik ist in § 16 Bundesstatistikgesetz geregelt. Demnach sind Einzelangaben, die zu statistischen Ergebnissen zusammengefasst werden, geheim zu halten. Es darf nicht möglich sein, auf Werte einzelner Unternehmen oder Personen zu schließen. Bei den Kaufwertestatistiken sollen auch die Daten einzelner Transaktionen geheim gehalten werden.

Bei der Überarbeitung der Bodenmarktstatistiken hat der Statistische Verbund beschlossen, künftig ein einheitliches Geheimhaltungsverfahren anzuwenden und dieses zu automatisieren. Nach dem Abschluss der Aufbereitung in den Ländern werden die Einzeldaten ab dem Berichtsjahr 2021 mittels der Zellspernung mit p%-Regel primär und sekundär geheim gehalten (Giessing/Dittrich, 2006). Eine Geheimhaltung wird durchgeführt, wenn die Differenz zwischen dem Tabellenwert und dem nächstkleineren Einzelwert den größten Einzelwert – in diesem Fall – um weniger als 5 % übersteigt (p%-Regel). Außerdem greift die Mindestfallzahlregel, wenn weniger als drei Transaktionen zum Ergebnis beitragen. Auf diese Art primär gesperrte Werte benötigen eine Gegen-spernung, um Rückrechnungen auf Einzelfälle auszu-schließen. Das gewählte Verfahren wird automatisiert mithilfe des Softwarepakets τ -ARGUS zentral für alle Länderdaten durchgeführt. Dazu ist es notwendig, alle Einzeldaten zeitgleich verarbeiten zu können. Eine nach-trägliche Integration weiterer Daten ist nicht möglich.

Grundvoraussetzung für die Entwicklung der Geheimhaltung war ein abgestimmtes Veröffentlichungsprogramm. Nach dessen Struktur werden die Einzeldaten zu den zu veröffentlichenden Ergebnissen aggregiert und können so in Veröffentlichungsprodukte überführt werden. Auch dieser Schritt erfolgt nach der Überarbeitung der Kaufwertestatistiken automatisiert.

3.2 Neues Veröffentlichungsprogramm

Bislang galt für die KWL-Statistik keine verpflichtende Datenlieferung an das Statistische Amt der Europäischen Union (Eurostat). Mit dem Inkrafttreten der SAIO-Rahmenverordnung wird sich dies ändern. Das Statistische Bundesamt wird ab dem Berichtsjahr 2021 jährlich Kaufwerte für landwirtschaftliche Flächen, unterteilt nach Ackerland und (Dauer)Grünland, ab NUTS2-Ebene (Regierungsbezirke und ehemalige Regierungsbezirke) bereitstellen (Statistisches Bundesamt, 2021). Diese Kaufwerte werden gemäß den europäischen Vorgaben flächengewichtet und national als europäisch harmonisierte Kaufwerte veröffentlicht. Von der Ergebniserstellung ausgeschlossen sind Kaufakte zwischen Personen mit einer familiären Beziehung, solche, bei denen die öffentliche Hand beteiligt war, und solche, bei denen für die Flächen keine weitere landwirtschaftliche Nutzung vorgesehen ist. Die Berechnung der nationalen Kaufwerte durch die Statistischen Ämter der Länder und das Statistische Bundesamt folgt weiterhin der Methodik der einfachen Durchschnitte auf ihrer jeweiligen regionalen Ebene. Wo bislang jedoch nur Ergebnisse für die Gesamtfläche und die Fläche der landwirtschaftlichen Nutzung vorlagen, sollen die Veröffentlichungstabellen ab dem Berichtsjahr 2021 jeweils nach Gesamtfläche, Ackerland, (Dauer)Grünland, landwirtschaftlich genutzter Fläche und sonstigen landwirtschaftlich genutzten Flächen unterschieden werden. Da es zu dieser detaillierten Abgrenzung bislang nur sehr wenige Erfahrungswerte gibt, muss sich im Berichtsjahr 2021 zeigen, ob eine Veröffentlichung tatsächlich für alle Flächenarten infrage kommt, oder ob aufgrund von zu geringen Fallzahlen darauf verzichtet werden muss. Als maximales Veröffentlichungsprogramm sollen Informationen zu den Veräußerungsfällen, veräußerten Flächen, Kaufsummen, Kaufwerten, Flächen je Veräußerungsfall sowie Ertragsmesszahlen je Hektar zur Verfügung gestellt werden. Diese werden gegliedert in Verkäufe zum Verkehrs-

wert ohne weitere Gegenstände, Rechte und Pflichten (teilweise nach Ertragsmesszahlen und Größenklassen aufgeschlüsselt) sowie mit weiteren Gegenständen, Rechten und Pflichten. Auch Zeitreihen sind vorgesehen, jedoch erst sinnvoll, wenn die überarbeitete Statistik bereits längere Zeit produziert worden ist. Schließlich soll es einen Überblick über die Struktur der ungewöhnlichen Geschäftsverkehre (siehe Abschnitt 2.2) geben. Erstmals sollen die neuen Veröffentlichungstabellen zum 30. Juni 2022 für das Berichtsjahr 2021 verfügbar sein, mehr als zwei Monate schneller im Vergleich zum bisherigen Vorgehen.

Bis zum Berichtsjahr 2020 wurden auch für Quartalsergebnisse der KWB-Statistik ausführliche Veröffentlichungstabellen erzeugt. Ab dem Berichtsjahr 2021 werden diese in einer Übersichtstabelle zusammengefasst, die die wichtigsten Merkmale Kauffälle, Kaufflächen, Kaufsummen und Kaufwerte untergliedert nach Art des Grundstücks und Art der Baufläche enthält. Die Quartalsergebnisse sollen 125 Tage nach dem Quartalsabschluss und somit etwa zwei Wochen früher als bisher zur Verfügung stehen. Sie werden nur die Transaktionen enthalten, die tatsächlich für das entsprechende Quartal durch die Gutachterausschüsse beziehungsweise Finanzämter gemeldet wurden. Diese vorläufigen Ergebnisse werden im Folgejahr zum 30. Juni um nachträglich gelieferte Kauffälle aktualisiert und zeitgleich mit dem Jahresergebnis sowie dem Jahresergebnis der KWL-Statistik bereitgestellt. Die Jahrestabellen⁴ informieren umfassender über die stattgefundenen Transaktionen als die Übersichtstabelle mit den Quartalsergebnissen. Die Kauffälle, Kaufflächen, Kaufsummen und Kaufwerte werden jeweils gegliedert nach Art des Grundstücks beziehungsweise Art der Baufläche sowie unterteilt nach Gemeindegrößenklassen, Grundstücksgrößenklassen sowie Preisklassen. Darüber hinaus sind gesonderte Tabellen für die Kombinationen Grundstücksgrößenklassen und Gemeindegrößenklassen sowie Preisklassen und Gemeindegrößenklassen vorgesehen. Zusätzlich findet weiterhin eine Befüllung der Regionaldatenbank statt. Auch die bislang bei der KWB-Statistik vorliegenden langen Reihen sollen fortgesetzt werden. Ob und in welchem Umfang dies möglich ist, ist noch nicht abschließend geprüft.

4 Prinzipiell stellen die dargestellten Tabellen den maximalen Veröffentlichungsrahmen dar, auf dem die Geheimhaltungsroutinen beruhen.


Über welche Medien es die Ergebnisse verbreitet, entscheidet jedes statistische Amt selbst. Das Statistische Bundesamt wird weiterhin vor allem die Datenbank GENESIS-Online nutzen.

4

Fazit und Ausblick

Die Neukonzeption der Bodenmarktstatistiken erfüllt nicht nur rechtliche Vorgaben und europäische Datenanforderungen. Sowohl für die Meldenden als auch die Nutzerinnen und Nutzer der Statistiken ergeben sich weitere Vorteile. Vor allem bei der KWL-Statistik bietet das neue Merkmalsprogramm zusätzliches Informationspotenzial, beispielsweise über die verschiedenen Flächenarten Ackerland, Grünland und landwirtschaftlich genutzte Fläche. Das eingesetzte Online-Meldev erfahren ermöglicht es, Transaktionsdaten schneller, standardisiert und sicher zu übermitteln. Die zentrale Geheimhaltung und automatisierte Tabellierung führen zu einem einheitlichen Veröffentlichungsprogramm und damit zur Vergleichbarkeit der bundesweit veröffentlichten regionalen Daten. Durch den zusätzlichen Einsatz des neuen Verbundprogramms liegen die Quartalsergebnisse der KWB-Statistik künftig zwei Wochen und die Jahresergebnisse beider Statistiken sogar über zwei Monate schneller vor.

Auch nach der generellen Überarbeitung in allen Bereichen der Statistikproduktion besteht bei beiden Bodenmarktstatistiken nach wie vor Entwicklungspotenzial. Bei der KWL-Statistik ist es durch die Erhebung des Status der Landwirtin/des Landwirts beispielsweise künftig theoretisch möglich, eine Aussage über außersektorale Verkäufe zu tätigen. Konkrete Pläne zur tiefergehenden Auswertung der Merkmale bestehen derzeit jedoch noch nicht. Sicher wird es auch einige Zeit benötigen, bis die Berichtsstellen die neuen Merkmalskataloge weitgehend vollständig umsetzen können und somit das gesamte Veröffentlichungspotenzial ausgeschöpft werden kann. Die mit der KWB-Statistik erhaltenen Informationen sollen später für Testrechnungen zu einem hedonischen Baulandindex genutzt werden. Damit könnten die Immobilienpreisstatistiken um eine weitere Statistik ergänzt werden.

Anfang Mai 2021 wurde das Fachverfahren KWBo erfolgreich eingeführt. Wie bei allen neuen Systemen ist auch hier eine Evaluationsphase vorgesehen mit dem Ziel zu klären, wie die Bearbeitung in den Statistischen Landesämtern noch komfortabler und einheitlicher gestaltet werden kann. 

LITERATURVERZEICHNIS

Blumöhr, Torsten/Teichmann, Corina/Noack, Anke. [Standardisierung der Prozesse: 14 Jahre AG SteP](#). In: WISTA Wirtschaft und Statistik. Ausgabe 5/2017, Seite 58 ff.

Giessing, Sarah/Dittrich, Stefan. [Tabellengeheimhaltung im statistischen Verbund – ein Verfahrensvergleich am Beispiel der Umsatzsteuerstatistik](#). In: Wirtschaft und Statistik. Ausgabe 8/2006, Seite 805 ff.

Schöneich, Cordula/Teske, Markus. [Regionalisierung des Häuserpreisindex](#). In: WISTA Wirtschaft und Statistik. Ausgabe 1/2020, Seite 32 ff.

Statistische Ämter des Bundes und der Länder. *Melden über .CORE*. 2014. [Zugriff am 21. Juni 2021]. Verfügbar unter: erhebungsportal.estatistik.de

Statistisches Bundesamt. *Fachserie M Preise, Löhne, Wirtschaftsrechnungen – Reihe 5 Preise und Preisindices für Bauwerke und Bauland – II. Baulandpreise 3. und 4. Vierteljahr 1961 und Jahr 1962*. Wiesbaden 1963.

Statistisches Bundesamt. *Fachserie B Land- und Forstwirtschaft, Fischerei – Reihe 5 Betriebe, Arbeitskräfte und technische Betriebsmittel – IV. Kaufpreissammlung für landwirtschaftliche Betriebe und Stückländereien 1961-1971*. Wiesbaden 1973.

Statistisches Bundesamt. *Fachserie B Land- und Forstwirtschaft, Fischerei – Reihe 5 Betriebe, Arbeitskräfte und technische Betriebsmittel – IV. Kaufwerte für landwirtschaftlichen Grundbesitz 1974*. Wiesbaden 1976.

Statistisches Bundesamt. *NUTS-Klassifikation. Die Einteilung der Europäischen Union in EU-Regionen*. 2021. [Zugriff am 28. Juni 2021]. Verfügbar unter: www.destatis.de

RECHTSGRUNDLAGEN

Bürgerliches Gesetzbuch (BGB) in der Fassung der Bekanntmachung vom 2. Januar 2002 (BGBl. I Seite 42, 2909; 2003 I Seite 738), das zuletzt durch Artikel 11 des Gesetzes vom 9. Juni 2021 (BGBl. I Seite 1666) geändert worden ist.

Gesetz über die Preisstatistik (PreisStatG) in der im Bundesgesetzblatt Teil III, Gliederungsnummer 720-9, veröffentlichten bereinigten Fassung, das zuletzt durch Artikel 9 des Gesetzes vom 22. Februar 2021 (BGBl. I Seite 266) 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 2 des Gesetzes vom 14. Juni 2021 (BGBl. I Seite 1751) geändert worden ist.

Verordnung über die bauliche Nutzung der Grundstücke (Baunutzungsverordnung – BauNVO) in der Fassung der Bekanntmachung vom 21. November 2017 (BGBl. I Seite 3786), die durch Artikel 2 des Gesetzes vom 14. Juni 2021 (BGBl. I Seite 1802) geändert worden ist.



Florian Peter

ist Volkswirt und Referent im Referat „Erzeugerpreise, Außenhandelspreise, Großhandelsverkaufspreise“ des Statistischen Bundesamtes. Seine Arbeitsschwerpunkte sind Energiepreise und die Weiterentwicklung der Erzeugerpreise gewerblicher Produkte, Außenhandelspreise und Großhandelsverkaufspreise.

NEUE STATISTIK DER STROM- UND ERDGASDURCHSCHNITTSPREISE

Florian Peter

➤ **Schlüsselwörter:** Preisstatistik — Strompreis — Erdgaspreis — Durchschnittspreise — Energie

ZUSAMMENFASSUNG

Für die neue Statistik der Strom- und Erdgasdurchschnittspreise hat das Statistische Bundesamt erstmals im Jahr 2020 Durchschnittspreise für Strom und Erdgas beim Verkauf an Endkunden im Haushaltssektor und Nicht-Haushaltssektor erfasst. Die ermittelten Preise werden gegliedert nach verschiedenen Jahresverbräuchen, Preisbestandteile wie Steuern, Netzentgelte sowie Vertrieb und Energie getrennt ausgewiesen. Mit der ersten Erhebung für den Berichtszeitraum zweites Halbjahr 2019 wurden erstmals die Vorgaben der Verordnung über europäische Erdgas- und Strompreisstatistik aus dem Jahr 2016 umgesetzt. Der Aufsatz erläutert die Anforderungen der neuen EU-Verordnung sowie deren Umsetzung in Deutschland.

➤ **Keywords:** price statistics – electricity price – natural gas price – average prices – energy

ABSTRACT

In 2020, the Federal Statistical Office for the first time calculated the average prices of electricity and natural gas sold to final household and non-household customers in order to compile the new statistics on average electricity and natural gas prices. The prices are structured according to different annual consumption bands. Price components such as taxes and network costs as well as distribution and energy are shown separately. The first survey was conducted for the reporting period covering the second half of 2019 and for the first time implemented the requirements of the Regulation on European statistics on natural gas and electricity prices of 2016. The article explains the requirements of the new EU Regulation and its implementation in Germany.

1

Einleitung

Im Jahr 2020 hat das Statistische Bundesamt erstmals Durchschnittspreise für Strom und Erdgas erhoben. Die Statistik wird auf Grundlage der Verordnung (EU) 2016/1952 durchgeführt.

Bis 2019 ermittelte der Bundesverband für Energie- und Wasserwirtschaft e.V. (BDEW) im Auftrag des Bundesministeriums für Wirtschaft und Energie die Durchschnittspreise nach den Vorgaben der Richtlinie 2008/92/EG. Die Neuregelungen durch die neue Verordnung 2016/1952 sollen verschiedene Entwicklungen des Energiemarktes berücksichtigen.

Eine wesentliche Entwicklung der letzten dreißig Jahre war der Prozess der Liberalisierung der Energiemärkte und der Entflechtung der Energieversorgung und des Netzbetriebs (Burger, 2014). Mit der Liberalisierung der Energiemärkte ist auch ein europäischer Energiebinnenmarkt entstanden, der europaweit vergleichbare und transparente Statistiken benötigt, die wiederum eine Analyse der energiepolitischen Instrumente ermöglichen. Die zweite, eher national bedeutsame Entwicklung war die der Belastung der Energie durch Steuern und Abgaben, in Deutschland vor allem auf Strom. Auch für diesen Bereich werden belastbare Statistiken benötigt, zum einen um nationale Daten zur Verfügung zu stellen. Zum anderen ist die Besteuerung der verschiedenen Energieträger im europäischen Vergleich von besonderem Interesse. Benötigt werden europaweite Preisdaten, zum Beispiel zur Analyse von Klimaschutzinstrumenten, die über einen Preismechanismus funktionieren, wie der EU-Emissionshandel¹ und der ab dem 1. Januar 2021 eingeführte nationale Emissionshandel².

Für Konzeption und Aufbau der für das Statistische Bundesamt neuen Statistik waren umfangreiche Vorarbeiten notwendig. Es wurden verschiedene Verwaltungsdaten für die Statistik nutzbar gemacht sowie eine eigene Erhebung aufgebaut. Voraussetzung für diese Arbeiten

waren Anpassungen der nationalen Gesetzgebung und umfassende Vorbefragungen von und Beratungen mit Energieversorgern und Netzbetreibern sowie dem BDEW.

Die folgenden Kapitel stellen die Anforderungen und Konzepte der Verordnung (EU) 2016/1952 dar (Kapitel 2) und beschreiben die Umsetzung in der deutschen amtlichen Statistik (Kapitel 3). Kapitel 4 erläutert die Erhebung und die Ergebnisse bis zum zweiten Halbjahr 2020.

2

Anforderungen und Konzepte der europäischen Verordnung über Erdgas- und Strompreisstatistik

Ziel der Verordnung (EU) 2016/1952 ist, eine höhere Transparenz über Energiekosten und -preise und eine europaweit vergleichbare Datengrundlage zu schaffen. Insbesondere die Beurteilung der klimapolitischen Ziele und des europäischen Green Deal (Europäische Kommission, 2019) hängt von einer verlässlichen und vergleichbaren Preismessung der verschiedenen Energiearten ab. So setzen die marktorientierten Mechanismen, wie die europäische und die nationale CO₂-Bepreisung, insbesondere auf preispolitische Anreize. Auch eventuelle Effekte einer Energiepreisreform und die Wirkung von Abgaben und Steuern können nur mit ausreichenden statistischen Informationen beurteilt werden (Sachverständigenrat zur Begutachtung der gesamtwirtschaftlichen Entwicklung, 2020, hier: Kapitel 4).

Hier setzt auch die neue Verordnung mit ihren Datenanforderungen an: halbjährliche Durchschnittspreise für Erdgas und Strom, unterteilt in Preise für Haushaltskunden und Endkunden des Nicht-Haushaltssektors.³ Die beiden Kundengruppen wiederum werden für Erdgas und Strom in Verbrauchsbänder unterteilt. Diese Verbrauchsbänder geben eine Ober- und Untergrenze des Jahresverbrauchs der Energiekunden an. Um europäische Aggregate der Durchschnittspreise bilden zu können, wird zusätzlich zu den Preisen eine prozentuale Verteilung der Verbrauchsmengen auf die Verbrauchsbänder benötigt.

1 Grundlage für den EU-Emissionshandel ist die Richtlinie über ein System für den Handel mit Treibhausgasemissionszertifikaten in der Gemeinschaft.

2 Mit dem Gesetz über einen nationalen Zertifikatehandel für Brennstoffemissionen (BEHG) wurde ein nationaler Emissionshandel eingeführt, der zur Bepreisung von Emissionen führen soll, die nicht durch den EU-Emissionshandel erfasst werden.

3 Die Verordnung (EU) 2016/1952 definiert Haushaltskunden, anders als § 3 Energiewirtschaftsgesetz, ausschließlich als Kunden, die nur für den privaten Verbrauch Energie nutzen.

2.1 Unterteilung in Preiskategorien

Die Durchschnittspreise für die Kunden- und Verbrauchsgruppen werden in drei Abgrenzungen ausgewiesen: (1) Bruttopreise mit allen Steuern und Abgaben, auch der Umsatzsteuer, (2) Preise ohne abzugsfähige Steuern und (3) Nettopreise ohne Steuern und Abgaben. Enthalten in den Preisen sind in allen drei Abgrenzungen die Netzentgelte sowie Energie- und Versorgungskosten. Mit den Preismeldungen für das zweite Halbjahr sind die Preise in Teilkomponenten zu gliedern. Diese Teilkomponenten sind zum einen Netzentgelte sowie Versorgungs- und Energiekosten, zum anderen sind es Steuern und andere staatliche Abgaben, die wiederum für Erdgas in fünf und für Strom in sechs Unterkategorien unterteilt sind.

Die Gliederung der Steuern, Abgaben und Umlagen sowie die Netzentgelte wurden im Vergleich zur Richt-

linie 2008/92/EG erweitert und neu strukturiert. Die [Übersichten 1 und 2](#) vergleichen die Gliederung für Steuern, Abgaben und Umlagen, getrennt nach Erdgas und Strom. Die nationalen Steuern und Abgaben werden in diese vorgegebenen Unterkategorien eingeordnet. Wichtigste Änderung ist die neue Gliederung der staatlichen Abgaben, die dem Umweltschutz dienen sollen. Sie wurden unterteilt in Abgaben, die erneuerbare Energiequellen fördern sollen, und Abgaben, die in Zusammenhang mit der Luftqualität und anderen Umweltschutzgründen stehen.

Entscheidend ist die Zuordnung der nationalen Steuern, Abgaben und Umlagen in diese Gliederung. Für Erdgas sind dies nur die Energiesteuer, die Konzessionsabgabe und die Umsatzsteuer. Seit dem 1. Januar 2021 werden Erdgas und andere fossile Brennstoffe zusätzlich durch das Brennstoffemissionshandelsgesetz mit einem CO₂-Preis von zunächst 25 Euro je Tonne CO₂

Übersicht 1

Gegenüberstellung der Unterkategorien Steuern, Abgaben und Umlagen für Erdgas

Verordnung (EU) 2016/1952	Richtlinie 2008/92/EG	Zuordnung deutscher Steuern, Abgaben und Umlagen
Mehrwertsteuer im Sinne der Richtlinie 2006/112/EG des Rates	Mehrwertsteuer	Mehrwertsteuer
Steuern, Abgaben und sonstige staatlich verursachte Belastungen und Gebühren im Zusammenhang mit der Förderung von erneuerbaren Energiequellen, Energieeffizienz und Energiegewinnung durch Kraft-Wärme-Kopplung	Umweltsteuern, -abgaben oder -belastungen; dies betrifft in der Regel entweder die Förderung erneuerbarer Energiequellen oder der Kraft-Wärme-Kopplung oder aber Abgaben auf CO ₂ -, SO ₂ - oder andere Emissionen, die mit dem Klimawandel in Zusammenhang stehen	
Steuern, Abgaben und sonstige staatlich verursachte Belastungen und Gebühren im Zusammenhang mit strategischen Vorräten, Kapazitätsszahlungen und Energieversorgungssicherheit; Steuern auf die Erdgasverteilung; verlorene Kosten und sonstige staatlich verursachte Belastungen bezüglich der Finanzierung von Energieregulierungsbehörden oder Marktteilnehmern und Netzbetreibern	Andere Steuern, Abgaben oder sonstige staatlich verursachte Belastungen im Energiesektor: zum Beispiel Abgaben zur Erfüllung gemeinschaftlicher Verpflichtungen, Gebühren zur Finanzierung von Energieregulierungsbehörden und so weiter	
Steuern, Abgaben und sonstige staatlich verursachte Belastungen und Gebühren im Zusammenhang mit der Luftqualität und zu anderen Umweltschutzgründen; Steuern auf Emissionen von CO ₂ oder anderen Treibhausgasen		Energiesteuer
Alle sonstigen nicht unter die vier genannten Kategorien fallenden Steuern, Abgaben und sonstigen staatlich verursachten Belastungen und Gebühren: Fernwärmeförderung; Abgaben an Gebietskörperschaften; Ausgleich für Insellage; Konzessionsabgaben im Zusammenhang mit Lizenzen und Gebühren für die Nutzung von Land und staatlichem oder privatem Eigentum für das Netz oder andere Versorgungseinrichtungen	Konzessionsabgaben; dies betrifft in der Regel Lizenzen und Gebühren für die Nutzung von Land-, Staats- oder Privatbesitz für das Netz oder andere Gasversorgungseinrichtungen	Konzessionsabgabe
	Andere Steuern oder Abgaben, die nicht mit dem Energiesektor verknüpft sind, zum Beispiel nationale, lokale oder regionale Steuern auf den Energieverbrauch, Steuern auf die Gasverteilung und so weiter	

belastet. Die Erhebung der Gaspreise für das erste Halbjahr 2021 wird diese Abgabe erstmals erfassen. Die an die Marktgebietsverantwortlichen⁴ entrichtete Bilanzierungsumlage wird nicht als staatliche Abgabe oder Steuer erfasst, sondern nur als Teil des Gesamtpreises berücksichtigt. Umsatzsteuer und Konzessionsabgaben werden den entsprechenden Unterkategorien zugeordnet, die Energiesteuer der Kategorie der Steuern und Abgaben im Zusammenhang mit der Luftqualität und dem Umweltschutz. Zwar waren auch fiskalische Gründe mit der Einführung der Energiesteuer im Jahr 2006 ver-

knüpft, jedoch war der Grundgedanke der Energiesteuer der Umweltschutz.

Strom wird mit deutlich mehr Steuern und Abgaben belastet, sowohl in der Anzahl der unterschiedlichen Abgaben als auch als Anteil am Gesamtpreis. Analog zur Energiesteuer auf Erdgas wird die Stromsteuer der Kategorie der Steuern und Abgaben im Zusammenhang mit der Luftqualität und des Umweltschutzes zugeordnet. Die Umlagen nach dem Erneuerbare-Energien-Gesetz (EEG-Umlage) und zur Kraft-Wärme-Kopplung (KWK-Umlage) gehören zur Kategorie der Steuern im Zusammenhang mit der Förderung von erneuerbaren Energiequellen, Energieeffizienz und Energiegewinnung

⁴ Ab Oktober 2021 sollen die bisherigen Marktgebiete GASPOOL und NetConnect Germany zu einem Marktgebiet zusammengelegt werden.

Übersicht 2

Gegenüberstellung der Unterkategorien Steuern, Abgaben und Umlagen für Strom

Verordnung (EU) 2016/1952	Richtlinie 2008/92/EG	Zuordnung deutscher Steuern, Abgaben und Umlagen
Mehrwertsteuer gemäß der Definition der Richtlinie 2006/112/EG	Mehrwertsteuer	Mehrwertsteuer
Steuern, Abgaben und sonstige staatlich verursachte Belastungen und Gebühren im Zusammenhang mit der Förderung von erneuerbaren Energiequellen, Energieeffizienz und Energiegewinnung durch Kraft-Wärme-Kopplung (KWK)	Umweltsteuern, -abgaben oder -belastungen — dies betrifft in der Regel entweder die Förderung erneuerbarer Energiequellen oder der Kraft-Wärme-Kopplung oder aber Abgaben auf CO ₂ -, SO ₂ - oder andere Emissionen, die mit dem Klimawandel in Zusammenhang stehen	EEG-Umlage KWK-Umlage
Steuern, Abgaben und sonstige staatlich verursachte Belastungen und Gebühren im Zusammenhang mit Kapazitätzahlungen, Energieversorgungssicherheit und angemessener Stromerzeugung; Steuer auf die Umstrukturierung des Kohlesektors; Steuern auf die Stromverteilung; verlorene Kosten und sonstige staatlich verursachte Belastungen bezüglich der Finanzierung von Energieregulierungsbehörden oder Marktteilnehmern und Netzbetreibern	Andere Steuern, Abgaben oder sonstige staatlich verursachte Belastungen im Energiesektor; zum Beispiel Abgaben zur Erfüllung gemeinwirtschaftlicher Verpflichtungen, Gebühren zur Finanzierung von Energieregulierungsbehörden und so weiter	§ 19-Umlage Offshore-Netzumlage Umlage für abschaltbare Lasten
Steuern, Abgaben und sonstige staatlich verursachte Belastungen und Gebühren im Zusammenhang mit der Luftqualität und zu anderen Umweltschutzgründen; Steuern auf Emissionen von CO ₂ oder anderen Treibhausgasen		Stromsteuer
Steuern, Abgaben und sonstige staatlich verursachte Belastungen und Gebühren im Zusammenhang mit der Kernkraft einschließlich Stilllegung und Beaufsichtigung von Kernkraftanlagen und diesbezügliche Gebühren	Kernkraftsteuer und sonstige Aufsichtsabgaben — dies betrifft unter anderem Stilllegungsgebühren für Kernkraftwerke, Aufsichtsabgaben und Gebühren für Kernkraftanlagen und so weiter	
Alle sonstigen nicht unter die fünf genannten Kategorien fallenden Steuern, Abgaben und sonstigen staatlich verursachten Belastungen und Gebühren: Fernwärmeförderung; Abgaben an Gebietskörperschaften; Ausgleich für Inselflage; Konzessionsabgaben im Zusammenhang mit Lizenzen und Gebühren für die Nutzung von Land und staatlichem oder privatem Eigentum für das Netz oder andere Versorgungseinrichtungen	Konzessionsabgaben — dies betrifft in der Regel Lizenzen und Gebühren für die Nutzung von Land-, Staats- oder Privatbesitz für das Stromnetz oder andere Stromversorgungseinrichtungen	Konzessionsabgabe
	Andere Steuern oder Abgaben, die nicht mit dem Energiesektor verknüpft sind, zum Beispiel nationale, lokale oder regionale Steuern auf den Energieverbrauch, Steuern auf die Stromverteilung und so weiter	

durch Kraft-Wärme-Kopplung. Die § 19-Umlage, die Offshore-Netzumlage und die Umlage für abschaltbare Lasten zählen zur Kategorie Steuern im Zusammenhang mit Kapazitätszahlungen, Energieversorgungssicherheit und angemessener Stromerzeugung.

2.2 Kundentypen und Verbrauchsbänder

Grundsätzlich werden Preisdaten für Erdgas und Strom mit der neuen europäischen Verordnung für Haushaltskunden, also private Haushalte, und Endkunden des Nicht-Haushaltssektors erfasst. Diese beiden Gruppen werden wiederum nach ihrem Jahresverbrauch in verschiedene Verbrauchsbänder unterteilt. Aufgrund der bisher geltenden Richtlinien haben die Mitgliedstaaten der Europäischen Union (EU) Daten für private Haushalte nur freiwillig erhoben, die neue Verordnung verpflichtet sie nun dazu. Diese rechtliche Verpflichtung soll die Vergleichbarkeit zwischen den EU-Mitgliedstaaten erhöhen und auf einem „zunehmend komplexen Energiebinnenmarkt“⁵ zuverlässige Erdgas- und Strompreisdaten gewinnen. Neu ist auch die Definition der Endkunden des Nicht-Haushaltssektors; nach den bisherigen Richtlinien wurden Preise für Industriekunden erfasst, dadurch waren Bereiche wie der Staat oder auch der Handel ausgeschlossen.⁶

Auch bei den Verbrauchsgruppen gab es Neuerungen. Neu eingeführt wurden die obersten Verbrauchsbänder IG (mehr als 150 000 Megawattstunden Jahresverbrauch an Strom) und I6 (mehr als 4 000 000 Gigajoule Erdgas Jahresverbrauch). Bislang wurden diese Großverbraucher nicht separat erfasst. Die neuen Bänder führen zu einem transparenteren und stärker differenzierten Bild der mengenabhängigen Preisstruktur. Insbesondere die Großverbraucher zahlen aufgrund ihrer großen Abnahmemengen in der Regel einen niedrigeren Durchschnittspreis. Das führte dazu, dass das Preisniveau in den bisherigen obersten Verbrauchsbändern I5 für Erdgas und IF für Strom nach der alten Gliederung tendenziell niedriger war als nach der neuen Gliederung.

➤ Übersicht 3, Übersicht 4

⁵ Begründung zur Verordnung (EU) 2016/1952, Punkt 3.

⁶ Ausnahmen, die eine praktikable Erhebung für die EU-Mitgliedstaaten ermöglichen sollten, waren in den bisherigen Richtlinien zugelassen, falls eine Trennung zwischen industriellen Endverbrauchern und anderen Endverbrauchern nicht möglich war.

Übersicht 3

Verbrauchsbänder Erdgas

	Jährlicher Energieverbrauch in Gigajoule	
	niedrigster Wert	höchster Wert
Haushaltskunden		
Gruppe D1		< 20
Gruppe D2	≥ 20	< 200
Gruppe D3	≥ 200	
Endkunden des Nicht-Haushaltssektors		
Gruppe I1		< 1 000
Gruppe I2	≥ 1 000	< 10 000
Gruppe I3	≥ 10 000	< 100 000
Gruppe I4	≥ 100 000	< 1 000 000
Gruppe I5	≥ 1 000 000	< 4 000 000
Gruppe I6	≥ 4 000 000	

Übersicht 4

Verbrauchsbänder Strom

	Jährlicher Energieverbrauch in Megawattstunden	
	niedrigster Wert	höchster Wert
Haushaltskunden		
Gruppe DA		< 1
Gruppe DB	≥ 1	< 2,5
Gruppe DC	≥ 2,5	< 5
Gruppe DD	≥ 5	< 15
Gruppe DE	≥ 15	
Endkunden des Nicht-Haushaltssektors		
Gruppe IA		< 20
Gruppe IB	≥ 20	< 500
Gruppe IC	≥ 500	< 2 000
Gruppe ID	≥ 2 000	< 20 000
Gruppe IE	≥ 20 000	< 70 000
Gruppe IF	≥ 70 000	< 150 000
Gruppe IG	≥ 150 000	

2.3 Kostenanteile der Energienetze

Von großer Bedeutung für die Analyse der Preise und die Beurteilung der Netzkosten insgesamt ist die neue Datenanforderung, Anteile der Netzkosten für Fernleitungs- beziehungsweise Übertragungsnetze einerseits und Verteilnetze andererseits zu ermitteln. Hierbei soll der relative Anteil der Fernleitungs- und Übertragungsnetzkosten jeweils für Haushaltskunden und Nicht-Haushaltskunden und der Anteil der Verteilungskosten für die beiden Kundengruppen festgestellt werden.

Gerade diese Daten sind für die Beurteilung der Versorgungssicherheit der Energienetze wichtig. Auch um den für die Energiewende und den vermehrten Einsatz erneuerbarer Energien notwendigen Netzausbau einzuschätzen sind Informationen über die Zusammensetzung der Netzkosten von entscheidender Bedeutung (Umweltbundesamt, 2020).

3

Datenquellen

Die neue europäische Verordnung über Erdgas- und Strompreisstatistik ermöglicht einen großen Spielraum bei der Nutzung verschiedener Datenquellen, um die Durchschnittspreise für Strom und Erdgas zu ermitteln. Explizit führt sie statistische Erhebungen, Verwaltungsquellen und sonstige Quellen auf, bei denen statistische Schätzmethoden zur Anwendung kommen.

Verschiedene Stellen besitzen Daten zum Energiesektor, zum einen fallen sie im Zuge der deutschen und europäischen Energiestatistik an, zum anderen bei verschiedenen Behörden als Verwaltungsdaten zu Steuern und Abgaben. Daneben verfügen die Regulierungsbehörden, das sind die Bundesnetzagentur und das Bundeskartellamt, über Daten zu den Energienetzen und Netzbetreibern. Bereits bei der Konzeption der Statistik und der Datenerfassung konnte daher auf die verschiedenen Datenquellen zurückgegriffen werden, um Doppelerhebungen zu vermeiden.

3.1 Preiserhebung

Die Bundesnetzagentur erhebt im Zusammenhang mit dem Energiemonitoring auch Preisinformationen der Energieversorger. Diese erfüllen jedoch nicht die Anforderungen der Verordnung (EU) 2016/1952 hinsichtlich der Kundentypen und Verbrauchsgruppen, zusätzlich liegen diese Daten nur jährlich mit einem zu großen Abstand zum Berichtszeitraum vor. Auch sonstige Datenquellen aus der Finanzverwaltung oder der Energieregulierung liefern keine ausreichenden Informationen über die Gesamtpreise für Strom oder Erdgas. Um die Halbjahresdurchschnittspreise für Strom und Erdgas zu ermitteln, war daher eine eigene Erhebung bei Energieversorgern und Netzbetreibern aufzubauen. Ergänzend

genutzt werden Verwaltungsdaten zur Stromsteuer und Energiesteuer sowie zu EEG-Umlage, KWK-Umlage und Offshore-Netzumlage.

Die Erhebung der Preisdaten erfolgt über eine geschichtete Zufallsstichprobe. Die Grundgesamtheit umfasst dabei alle Unternehmen, die Endkunden mit Erdgas oder Strom beliefern, sowie alle Verteilnetzbetreiber. Auswahlgrundlage bilden die Energiestatistik und das Marktstammregister der Bundesnetzagentur. Die Schichten der Stichprobe werden sowohl regional als auch nach Umsatzgröße der Unternehmen gebildet. Die gebildeten Regionen berücksichtigen neben der geographischen Lage auch wirtschaftliche Aspekte und die unterschiedliche Nutzung von Energieerzeugungsquellen:

- › Region 1: Schleswig-Holstein, Hamburg, Niedersachsen, Bremen, Mecklenburg-Vorpommern
- › Region 2: Nordrhein-Westfalen
- › Region 3: Hessen, Rheinland-Pfalz, Saarland
- › Region 4: Baden-Württemberg, Bayern
- › Region 5: Berlin, Brandenburg, Sachsen, Sachsen-Anhalt, Thüringen

Weitere Schichtungsmerkmale sind der Anteil des Umsatzes eines Unternehmens in Relation zum Gesamtumsatz des deutschen Erdgas- beziehungsweise Strommarkts sowie der meistbeliebteste Kundentyp (Haushalts-/Nicht-Haushaltskunden). Für die zufällige Stichprobenziehung wurde für alle Schichten der erforderliche Stichprobenumfang bestimmt. Insgesamt wurden 370 Stromversorger und 210 Erdgasversorger ausgewählt sowie jeweils 50 Stromnetzbetreiber und Gasnetzbetreiber.

Die ausgewählten Energieversorger sollen Auskunft geben zu ihren mengengewichteten Durchschnittspreisen sowie den Preisbestandteilen je Kundentyp und Verbrauchsband. Mit den ebenfalls abgefragten Verbrauchsmengen je Verbrauchsband zur Gewichtung wird ein bundesweiter Durchschnittspreis für Strom und Erdgas berechnet. Dieser enthält noch nicht die zwischen Verbrauchern und Netzbetreibern unabhängigen individuellen Netzentgelte, zu denen die Netzbetreiber befragt werden. Die individuell abgerechneten Netzentgelte werden zu den bei den Energieversorgern ermittelten Netzentgelten hinzugerechnet.

3.2 Verwaltungsdaten

Den größten Anteil an den Strom- und Erdgaspreisen haben in der Regel die staatlichen Abgaben. Entscheidend für die Endpreise ist hier die tatsächliche Höhe der gezahlten Steuern und Abgaben. Für Stromsteuer, EEG-Umlage, KWK-Umlage und Offshore-Umlage, aber auch für die Energiesteuer bestehen für Unternehmen jeweils verschiedene Möglichkeiten zur Reduzierung, dagegen bestehen für Haushaltskunden keine Möglichkeiten der Reduzierung. Mithilfe von Daten des Bundesamtes für Wirtschaft und Ausfuhrkontrolle und der Generalzolldirektion werden diese Beträge ermittelt.

Die Besondere Ausgleichsregelung erlaubt stromkostenintensiven Unternehmen und Schienenbahnen auf Antrag eine Begrenzung der EEG-Umlage. Sie hängt von der Stromkostenintensität des Unternehmens ab und ist somit individuell für jedes Unternehmen zu berechnen (Bundesamt für Wirtschaft und Ausfuhrkontrolle, 2020). Der Antrag ist jährlich neu zu stellen, die Nachweisdaten zum Stromverbrauch stammen dabei jeweils aus dem Vorjahr. Aus diesen Daten kann das Bundesamt für Wirtschaft und Ausfuhrkontrolle die Höhe der tatsächlich von den berechtigten Unternehmen gezahlten EEG-Umlage für den jeweiligen Berichtszeitraum der Statistik der Durchschnittspreise ermitteln. Auch ist eine Zuordnung zu den verschiedenen Verbrauchsbändern möglich. Allerdings sind diese Verbrauchswerte aus dem Antragsverfahren zur Besonderen Ausgleichsregelung Vergangenheitswerte, sie dienen als Schätzung für den tatsächlichen Verbrauch des Berichtsjahrs.

Daten zur Begrenzung der KWK-Umlage stehen ebenfalls dem Bundesamt für Wirtschaft und Ausfuhrkontrolle zur Verfügung. Eine Begrenzung der EEG-Umlage durch die Besondere Ausgleichsregelung ist in der Regel Voraussetzung für eine Begrenzung der KWK-Umlage. Hier kann erneut auf die Daten aus dem Antrag zur Besonderen Ausgleichsregelung zurückgegriffen werden. Die Begrenzung der Offshore-Netzumlage wiederum ergibt sich aus der Begrenzung der KWK-Umlage, auch hier ist es möglich, die Daten des Bundesamtes für Wirtschaft und Ausfuhrkontrolle aus dem Antragsverfahren zur Besonderen Ausgleichsregelung für die Berechnung der tatsächlichen Belastung des Strompreises heranzuziehen.

Das Stromsteuergesetz und das Energiesteuergesetz bieten verschiedene Möglichkeiten, um die Strom-

oder Energiesteuer zu reduzieren beziehungsweise eine Rückerstattung zu erlangen. Die Stromsteuer wird für bestimmte stromintensive Prozesse oder Verfahren erstattet. Zudem gilt für Unternehmen des Produzierenden Gewerbes sowie der Land- und Forstwirtschaft ein reduzierter Stromsteuersatz und durch den sogenannten Spitzenausgleich können Unternehmen des Produzierenden Gewerbes die verbleibende Stromsteuer erstattet bekommen. Für die auf Erdgas anfallende Energiesteuer gibt es Entlastungen für bestimmte energieintensive Produktionsprozesse, ebenso wird die Energiesteuer auf Antrag für Unternehmen des Produzierenden Gewerbes und der Land- und Forstwirtschaft reduziert. Auch hier ist eine teilweise Erstattung der Energiesteuer über den Spitzenausgleich für energieintensive Unternehmen möglich. Daten über die so entlasteten Strom- und Erdgasmengen sowie über den Entlastungsbetrag stellt die Generalzolldirektion zur Verfügung, sodass auch hier die tatsächlich gezahlten Steuerbeträge beziehungsweise die Steuererstattungen berechnet werden können.

Die so berechneten tatsächlichen Steuern und Abgaben und die Ergebnisse der beiden Erhebungen bei den Energieversorgern und Netzbetreibern zu den dort gezahlten Durchschnittspreisen und Verbrauchsmengen werden dann zur Berechnung entsprechender mengengewichteter Durchschnittspreise herangezogen. Das erfolgt jeweils für Strom und Erdgas für die beiden Kundentypen und die verschiedenen Verbrauchsbänder.

Daten über die Verteilung der Netzkosten auf Verteilernetze und Fernleitungs- beziehungsweise Übertragungsnetze liegen der Bundesnetzagentur nicht direkt vor. Angaben hierzu können aus den vorliegenden Daten der Netzregulierung und dem Energiemonitoring (Bundesnetzagentur, 2020) geschätzt werden. Für Strom verfügt die Bundesnetzagentur über Angaben zu den Erlösobergrenzen der Übertragungsnetzbetreiber sowie den Erlösobergrenzen der Verteilernetzbetreiber in Zuständigkeit der Bundesnetzagentur. Diese Erlösobergrenzen der einzelnen Netzbetreiber dienen als Schätzer für die tatsächlichen Kosten der Netzbetreiber (Bundesnetzagentur, 2021). Auf die Verteilernetzbetreiber, die der Regulierung durch die Bundesnetzagentur unterliegen, entfallen etwa 85 % der bundesweiten Erlösobergrenze, sodass dieser Wert auf alle Verteilernetzbetreiber hochgerechnet werden muss. In einem weiteren Schritt werden die Erlösobergrenzen der Verteilernetzbetrei-

ber um die darin enthaltenen vorgelagerten Netzkosten bereinigt. Aus diesen bereinigten Netzkosten der Verteilernetzbetreiber und der Übertragungsnetzbetreiber wird dann der Kostenanteil der jeweiligen Netzebene berechnet.

Für die Gasnetze liegen ebenfalls die Erlösobergrenzen für Verteilernetze und Fernleitungsnetze vor. Auch bei den Gasnetzen fallen nicht alle Verteilernetzbetreiber in die Zuständigkeit der Bundesnetzagentur, weswegen auch hier wieder eine Hochrechnung auf alle Netzbetreiber erfolgt. Doppelzählungen beim Ausspeisen von Erdgas von einem Verteilernetz in ein anderes können hier nicht bereinigt werden. Der Kostenanteil der Fernleitungsnetze ergibt sich abermals aus dem Verhältnis der Erlösobergrenzen von Verteilernetzbetreibern und Fernleitungsnetzbetreibern.

4

Datenerhebung und Ergebnisse

4.1 Datenerhebung

Das Statistische Bundesamt hat die Durchschnittspreise für Strom und Erdgas erstmals für das zweite Halbjahr 2019 ermittelt. Die bis zum ersten Halbjahr 2019 durch das Statistische Amt der Europäischen Union (Eurostat) veröffentlichten [Preise](#) hat für Deutschland der BDEW geliefert.⁷ Nach Inkrafttreten der Verordnung (EU) 2016/1952 nutzte Deutschland einen Aufschub von zwei Jahren, um die notwendigen Änderungen in das Preisstatistikgesetz zu integrieren, die Erhebung aufzubauen und die Verwaltungsdatennutzung vorzubereiten. Neben der Möglichkeit zur Verwaltungsdatennutzung waren auch die Ermittlung der Verbrauchsmengen sowie die halbjährliche Periodizität der Statistik in das Preisstatistikgesetz aufzunehmen.

Um die Erhebung bei den Energieversorgern und Netzbetreibern vorzubereiten, hat das Statistische Bundesamt umfangreiche Vorbefragungen durchgeführt und einzelne Versorger auch zu Gesprächen eingeladen.

Zusätzlich war bei der Konzeption der Statistik auch der BDEW zu einzelnen Punkten einbezogen, um auch dessen Erfahrungen bei der bisherigen Ermittlung der Preise zu nutzen. Bei der Konzeption der Statistik hat sich früh gezeigt, dass die Ermittlung der notwendigen Durchschnittspreise und auch der Preisbestandteile für die Energieversorger durchaus schwierig sein kann. Da gegebenenfalls Anpassungen in der jeweiligen Auswertungssoftware erforderlich sein können, wurden die in die Stichprobe gezogenen Unternehmen frühzeitig informiert.

Die Erhebung für das zweite Halbjahr 2019 erfolgte im ersten Quartal 2020. Die Daten sind spätestens drei Monate nach Ende des Berichtszeitraums an Eurostat zu melden. Somit lag der Zeitraum der Erhebung und der Preismeldungen der befragten Versorger und Netzbetreiber genau in der ersten Welle der Corona-Pandemie im Februar und März 2020. Dies führte zu Schwierigkeiten im Erhebungsablauf und aufgrund schlechterer Erreichbarkeit der Unternehmen zu einer unerwartet hohen Ausfallquote. Grund dafür war, dass Rückfragen, die sich bei der Plausibilisierung und Prüfung der Meldungen ergaben, nicht rechtzeitig geklärt werden konnten. Dadurch lag die Ausfallquote beziehungsweise die Quote der Meldungen, die nicht für die Berechnung genutzt werden konnten, für Erdgas bei etwa 30 %, für Strom bei etwa 35 %. Dennoch waren Qualität und Menge der Preisdaten gerade noch ausreichend, um Durchschnittspreise für alle Kundentypen und Verbrauchsbänder zu berechnen.

4.2 Ergebnisse

Unabhängig von diesen Problemen kam es durch den Übergang der Statistik zum Statistischen Bundesamt und die Umsetzung der neuen Verordnung zu einem erwarteten Bruch in den Zeitreihen einzelner Verbrauchsbänder. Gründe waren die veränderten Erhebungsmethoden, zum einen die Stichprobengröße betreffend, zum anderen die Auswahl der befragten Unternehmen. Ebenso ließ die Verwendung der Verwaltungsdaten zu den Steuern und Abgaben insbesondere bei den Nicht-Haushalten in den höheren Verbrauchsbändern Abweichungen zu den bisherigen Ergebnissen erwarten, sowohl bei Erdgas als auch bei Strom.

⁷ Ab dem Berichtszeitraum zweites Halbjahr 2019 veröffentlicht das Statistische Bundesamt die Ergebnisse in seiner Datenbank [GENESIS-Online](#).

Bei der Interpretation der Ergebnisse ist zu beachten, dass die Entwicklung absoluter Durchschnittspreise beobachtet wird. Anders als bei sonstigen Preisstatistiken, die die reine Preisentwicklung unabhängig von Mengen- oder Qualitätsveränderungen verfolgen, spielt die Mengenentwicklung bei der Beobachtung der Durchschnittspreise eine entscheidende Rolle. Insbesondere die Gliederung in Verbrauchsbänder kann dafür sorgen, dass Verschiebungen im Verbrauchsverhalten sowohl bei Haushalten als auch bei Nicht-Haushalten zu Änderungen der Durchschnittspreise führen, auch wenn keine wirklichen Preisänderungen stattfanden. Insbesondere bei den großen Verbrauchsbändern können einzelne Kundenwechsel eines befragten Versorgers oder geänderte Verbräuche Änderungen der gemeldeten Durchschnittspreise verursachen. Auch sorgen mengenunabhängige Preisbestandteile in den Abrechnungen der Energiekunden (wie Grundpreis oder Messstellen- und Abrechnungsgebühren) dafür, dass sich Durchschnittspreise nur aufgrund von Verbrauchsänderungen unterscheiden.

↘ **Tabelle 1** stellt die Ergebnisse für das zweite Halbjahr 2019 bis zum zweiten Halbjahr 2020 für die Haushalte dar, ↘ **Tabelle 2** enthält die Ergebnisse für die Nicht-Haushalte. Auffallend ist sowohl für Strom als auch für Erdgas, dass die Durchschnittspreise stark von der Ver-

brauchsmenge abhängen, was wie oben beschrieben an den konstanten Preiskomponenten liegt. Je höher der Jahresverbrauch ist, desto niedriger ist der gezahlte Durchschnittspreis. Deutlich unterscheiden sich die Anteile der Preiskomponenten für Strom und Erdgas. Während für Erdgas die Energie- und Vertriebskosten sowohl bei Haushalten als auch bei Nicht-Haushalten den größten Anteil am Gesamtpreis ausmachen, liegt beim Strompreis der Anteil der Steuern und Abgaben für fast alle Verbrauchsbänder über 50 %.

Im Vergleich mit den anderen EU-Mitgliedstaaten zeigen sich auch deutliche Unterschiede zwischen Erdgas und Strom. Die ↘ **Tabellen 3 und 4** auf Seite 106 zeigen beispielhaft jeweils ein mittleres Verbrauchsband für Strom und Erdgas für Nicht-Haushalte. Die Preise sind ohne Mehrwertsteuer und andere abzugsfähige Steuern als Durchschnitt aus den beiden Halbjahren 2020 ausgewiesen. Die nicht abzugsfähigen Steuern wurden als Differenz zwischen den Preiskomponenten Netzkosten, Vertrieb und Energie und dem Preis ohne abzugsfähige Steuern berechnet. Bei den Erdgaspreisen liegt Deutschland etwa im europäischen Mittelfeld leicht über dem EU-Durchschnitt, auch der Anteil der nicht abzugsfähigen Steuern und Abgaben befindet sich im europäischen Mittelfeld. Anders verhält es sich bei Strom; sowohl beim Preis als auch beim Anteil an nicht abzugsfähigen

Tabelle 1
Strom- und Erdgaspreise für private Haushalte

	2. Halbjahr 2019				1. Halbjahr 2020				2. Halbjahr 2020			
	Gesamtpreis	Energie und Vertrieb	Netz-entgelte	Steuern, Abgaben und Umlagen	Gesamtpreis	Energie und Vertrieb	Netz-entgelte	Steuern, Abgaben und Umlagen	Gesamtpreis	Energie und Vertrieb	Netz-entgelte	Steuern, Abgaben und Umlagen
Cent je Kilowattstunde												
Jahresverbrauch von ... bis unter ... Kilowattstunden	Strom											
unter 1 000	45,70	9,24	18,08	18,37	47,52	12,38	16,13	19,00	44,20	11,79	14,59	17,82
1 000 – 2 500	32,44	6,83	9,38	16,22	34,30	7,47	9,95	16,88	33,43	7,36	9,87	16,20
2 500 – 5 000	28,78	5,81	7,40	15,56	30,43	5,68	8,62	16,13	30,06	5,83	8,68	15,55
5 000 – 15 000	26,44	5,47	5,92	15,04	28,63	5,75	7,20	15,69	28,00	5,81	7,08	15,11
15 000 und mehr	23,70	4,92	4,44	14,32	25,79	4,92	5,93	14,93	25,23	5,14	5,62	14,47
Haushalte insgesamt	29,92	6,17	8,04	15,71	31,94	6,58	9,00	16,35	31,16	6,60	8,85	15,71
Jahresverbrauch von ... bis unter ... Gigajoule	Erdgas											
unter 20	7,90	3,57	2,34	1,98	8,29	3,68	2,57	2,04	8,31	3,84	2,57	1,89
20 – 200	5,88	2,80	1,51	1,57	5,97	2,84	1,56	1,58	6,20	3,09	1,62	1,49
200 und mehr	5,28	2,56	1,26	1,46	5,37	2,57	1,32	1,48	5,33	2,70	1,27	1,36
Haushalte insgesamt	5,83	2,78	1,49	1,56	5,99	2,84	1,57	1,58	6,12	3,05	1,60	1,48

Steuern und Abgaben liegt Deutschland an erster Stelle in der EU. Diese Feststellung lässt sich auf fast alle Verbrauchsbänder sowohl für Nicht-Haushaltskunden als auch für Haushaltskunden von Strom und Erdgas übertragen.

Im zweiten Halbjahr 2019 zahlten die Haushalte im Durchschnitt 29,92 Cent je Kilowattstunde Strom und 5,83 Cent je Kilowattstunde Erdgas (siehe Tabelle 1). Im ersten Halbjahr 2020 stiegen die Durchschnittspreise über alle Verbrauchsgruppen auf 31,94 Cent je Kilowattstunde für Strom und auf 5,99 Cent je Kilowattstunde für Erdgas. Die zum 1. Juli 2020 als Maßnahme des Konjunkturpakets von der Bundesregierung beschlossene Senkung des Mehrwertsteuersatzes konnte aufgrund einer Sonderregelung (Bundesministerium der Finanzen, 2020) von den Energieversorgern für Strom und Erdgas auch schon im ersten Halbjahr 2020 genutzt werden. Der überwiegende Teil der Energieversorger wendete die Senkung des Mehrwertsteuersatzes aber erst ab dem zweiten Halbjahr 2020 an. Bei den Haushalten profitierten etwa 13 % der Stromkunden und 8 % der

Erdgaskunden von der Sonderregelung und zahlten auch im ersten Halbjahr 2020 die reduzierte Mehrwertsteuer. Im zweiten Halbjahr 2020 profitierte hingegen der überwiegende Teil der Haushaltskunden⁸ vom gesenkten Mehrwertsteuersatz. Für Stromkunden sorgte dieser über alle Verbrauchsbänder für einen gesunkenen Preis gegenüber dem ersten Halbjahr 2020. Im Durchschnitt zahlten alle Haushalte 31,16 Cent je Kilowattstunde im zweiten Halbjahr 2020. Erdgaskunden zahlten im Durchschnitt 6,12 Cent je Kilowattstunde und damit mehr als im ersten Halbjahr.

Nicht-Haushalte zahlten im zweiten Halbjahr 2019 im Durchschnitt 17,56 Cent je Kilowattstunde für Strom und 3,15 Cent je Kilowattstunde für Erdgas, jeweils ohne Mehrwertsteuer und alle abzugsfähigen Steuern (siehe Tabelle 2). Im ersten Halbjahr 2020 mussten für Strom im Durchschnitt nur noch 14,89 Cent je Kilowattstunde

8 Für die Energieversorger besteht bei rollierender Abrechnung ihrer Kunden auch die Möglichkeit, Strom und Erdgas, das im zweiten Halbjahr 2020 verbraucht wurde, mit dem vollen Mehrwertsteuersatz abzurechnen, wenn die Abrechnung erst im Jahr 2021 stattfindet.

Tabelle 2

Strom und Erdgaspreise für Nicht-Haushaltskunden ohne Mehrwertsteuer und andere abzugsfähige Steuern

	2. Halbjahr 2019				1. Halbjahr 2020				2. Halbjahr 2020			
	Gesamtpreis	Energie und Vertrieb	Netzentgelte	Steuern, Abgaben und Umlagen	Gesamtpreis	Energie und Vertrieb	Netzentgelte	Steuern, Abgaben und Umlagen	Gesamtpreis	Energie und Vertrieb	Netzentgelte	Steuern, Abgaben und Umlagen
	Cent je Kilowattstunde											
Jahresverbrauch von ... bis unter ... Megawattstunden	Strom											
unter 20	22,89	5,25	6,82	10,82	24,55	6,27	7,19	11,08	24,85	6,78	6,91	11,16
20 – 500	18,63	3,47	5,25	9,91	20,12	4,36	5,57	10,19	20,36	4,77	5,38	10,20
500 – 2 000	16,08	3,04	4,14	8,90	17,81	4,20	4,29	9,32	18,18	4,27	4,58	9,33
2 000 – 20 000	13,61	3,10	3,13	7,38	15,09	3,60	3,46	8,03	15,34	4,00	3,30	8,03
20 000 – 70 000	10,93	3,10	2,04	5,79	12,06	2,86	2,45	6,75	12,71	3,58	2,44	6,68
70 000 – 150 000	9,25	3,10	1,42	4,73	10,07	2,94	1,40	5,72	11,22	3,26	1,64	6,32
150 000 und mehr	7,68	2,36	1,12	4,20	6,43	2,77	0,65	3,02	7,33	3,10	0,80	3,43
Nicht-Haushalte insgesamt	17,56	3,17	3,27	11,12	14,89	3,74	3,51	7,64	15,08	4,09	3,38	7,62
Jahresverbrauch von ... bis unter ... Gigajoule	Erdgas											
unter 1 000	4,23	2,34	1,27	0,62	4,45	2,57	1,28	0,60	4,42	2,44	1,38	0,60
1 000 – 10 000	3,72	2,12	1,02	0,58	3,70	2,10	1,04	0,56	3,70	2,03	1,12	0,55
10 000 – 100 000	3,00	1,90	0,63	0,47	2,89	1,85	0,60	0,45	2,89	1,81	0,63	0,44
100 000 – 1 000 000	2,50	1,77	0,35	0,38	2,30	1,60	0,30	0,40	2,34	1,60	0,33	0,40
1 000 000 – 4 000 000	1,90	1,53	0,11	0,26	1,82	1,28	0,14	0,40	1,88	1,33	0,14	0,41
4 000 000 und mehr	1,83	1,47	0,10	0,26	1,71	1,15	0,12	0,43	1,78	1,17	0,16	0,46
Nicht-Haushalte insgesamt	3,15	1,96	0,72	0,47	2,75	1,74	0,55	0,46	2,67	1,67	0,54	0,46

Tabelle 3

Nicht abzugsfähige Steuern bei einem Stromverbrauch von 2 000 bis 19 999 Megawattstunden (Gruppe ID Nicht-Haushaltskunde) im europäischen Vergleich¹ 2020

	Preis einschließlich nicht abzugsfähiger Steuern (berechnet)	Nicht abzugsfähige Steuern in Cent je Kilowattstunde (berechnet)	Anteil nicht abzugsfähiger Steuern
	EUR		%
Deutschland	15,22	8,04	52,81
Italien	12,95	4,69	36,19
Euroraum	11,34	4,10	36,13
Niederlande	9,80	3,44	35,10
Europäische Union	10,70	3,56	33,24
Portugal	10,17	2,96	29,07
Polen	9,72	2,74	28,15
Belgien	9,92	2,76	27,79
Slowakei	12,4	3,42	27,58
Österreich	9,98	2,72	27,25
Norwegen	3,93	0,89	22,65
Griechenland	9,20	2,02	21,96
Schweden	5,53	0,41	7,33
Kosovo ²	6,16	0,37	5,93
Irland	10,82	0,61	5,68
Ukraine	5,56	0,19	3,42
Island	4,96	0,15	3,02
Bosnien und Herzegowina	7,01	0,19	2,64
Moldau	7,34	0,11	1,57
Malta	11,84	0,16	1,31
Bulgarien	7,85	0,10	1,27
Nordmazedonien	7,21	0,02	0,21

1 Top 10 und Bottom 10.

2 Gemäß der Resolution 1244/99 des Sicherheitsrates der Vereinten Nationen.

Quelle: Eurostat/eigene Berechnung

gezahlt werden. Dieser starke Rückgang resultiert zum Teil tatsächlich aus den gesunkenen Preisen für Strom, vor allem Großverbraucher profitierten von den niedrigen Börsenpreisen im ersten Halbjahr 2020. Zum Großteil aber ergab sich der Preisrückgang aus einer Verschiebung der Verbrauchsstruktur der Nicht-Haushalte insgesamt. Dies dürfte eine Folge der Corona-Pandemie sein und des damit verbundenen ersten Lockdowns im Jahr 2020. Durch diese Verbrauchsänderungen erhielten die niedrigeren Preise der hohen Verbrauchsbänder ein höheres Gewicht bei der Berechnung des Durchschnittspreises, da diese im Vergleich einen weniger starken Verbrauchsrückgang verzeichneten. Betrachtet man die Verbrauchsbänder einzeln, stiegen in fast allen die Preise gegenüber dem Vorhalbjahr. Bei den Erdgaspreisen

Tabelle 4

Nicht abzugsfähige Steuern bei einem Erdgasverbrauch zwischen 100 000 und 1 000 000 Gigajoule (Gruppe I4 Nicht-Haushaltskunde) im europäischen Vergleich¹ 2020

	Preis einschließlich nicht abzugsfähiger Steuern (berechnet)	Nicht abzugsfähige Steuern in Cent je Kilowattstunde (berechnet)	Anteil nicht abzugsfähiger Steuern
	EUR		%
Schweden	3,34	1,09	32,63
Dänemark	2,10	0,56	26,67
Niederlande	2,28	0,60	26,15
Österreich	2,37	0,60	25,32
Deutschland	2,32	0,41	17,67
Euroraum	2,21	0,31	14,03
Estland	2,31	0,30	12,99
Europäische Union	2,21	0,29	12,93
Türkei	1,84	0,23	12,50
Litauen	1,89	0,24	12,47
Frankreich	2,12	0,26	12,06
Slowenien	2,48	0,29	11,69
Slowakei	2,54	0,13	5,12
Rumänien	2,18	0,10	4,37
Bulgarien	1,63	0,06	3,68
Polen	2,31	0,08	3,25
Bosnien und Herzegowina	3,86	0,11	2,85
Serbien	3,01	0,08	2,66
Portugal	2,18	0,05	2,07
Luxemburg	1,85	0,01	0,54
Moldau	2,16	0,00	0,00
Ukraine	1,56	0,00	0,00

1 Top 10 und Bottom 10.

Quelle: Eurostat/eigene Berechnung

zeigt sich ein ähnlicher Mengeneffekt wie bei Strom, allerdings profitierten die Nicht-Haushaltskunden hier deutlicher von den gesunkenen Beschaffungspreisen für Erdgas.


In der zweiten Jahreshälfte 2020 begannen die Strom- und Erdgaspreise an der Börse wieder zu steigen. Insbesondere die Großverbraucher bekamen dies in ihren Einkaufspreisen zu spüren, da deren Preisanpassungen in der Regel in kürzeren Perioden erfolgen. Für Strom mussten Nicht-Haushalte mit 15,08 Cent je Kilowattstunde etwas mehr zahlen als im ersten Halbjahr 2020, für Erdgas sank der Durchschnittspreis über alle Verbrauchsgruppen gegenüber der ersten Jahreshälfte 2020 auf 2,67 Cent je Kilowattstunde. Dies lag aber wieder an einer Verschiebung der Verbrauchsstruktur.

5

Fazit

Der europäische Energiebinnenmarkt wurde in den letzten Jahrzehnten zunehmend komplexer und hat sehr große Bedeutung für die künftige europäische Klimapolitik. Entsprechend hoch sind die Anforderungen an die Energiestatistiken, die diesen Markt beobachten und abbilden sollen. Die Verordnung (EU) 2016/1952 gibt einen gemeinsamen europäischen Rahmen vor, nach dem die Durchschnittspreise für Erdgas und Strom zu ermitteln sind. Es gibt nicht mehr nur eine rechtliche Verpflichtung, Preise für Industriekunden zu erfassen, sondern auch die Preise aller anderen Endkunden werden erfasst. Durch die neu gefasste Gliederung der staatlichen Abgaben und der Netzentgelte wird mehr Transparenz der Preise auf den verschiedenen Stufen und eine bessere Vergleichbarkeit zwischen den EU-Mitgliedstaaten erreicht. Indem alle Teilkomponenten der Energiepreise einzeln erfasst werden, können eventuelle Auswirkungen von einzelnen steuerrechtlichen oder anderen politischen Maßnahmen, wie etwa die vorübergehende Senkung des Mehrwertsteuersatzes im Jahr 2020, beurteilt werden. Die Verordnung lässt ein breites Spektrum an Datenquellen zu, so können alle Verwaltungsdaten, die über den Energiemarkt existieren, genutzt werden und zusätzliche bürokratische Belastungen werden vermieden durch die Anlehnung an das Once-Only-Prinzip. Dennoch ist zusätzlich zu den genutzten Verwaltungsdaten eine Preiserhebung bei Versorgern und Netzbetreibern durchzuführen. Der deutsche Energiemarkt unterscheidet sich in weiten Teilen deutlich von anderen europäischen Märkten. Auch die Vielzahl an Marktteilnehmern macht eine größere Stichprobenerhebung notwendig, um verlässliche Preisdaten zu ermitteln.

Im europäischen Vergleich sind in Deutschland die höchsten Strompreise über fast alle Verbrauchsbänder zu zahlen. Dies liegt überwiegend am überdurchschnittlich hohen Anteil an Steuern und Abgaben am Gesamtpreis, sowohl bei Haushaltskunden als auch bei Nicht-Haushaltskunden. Die Erdgaspreise liegen in Deutschland in der Regel leicht über dem EU-Durchschnitt.

Der Energiemarkt wird zweifelsohne auch in Zukunft erheblichen Veränderungen ausgesetzt sein, die durch die neue Verordnung erfasst werden sollen. Diese Veränderungen des Marktes bedeuten aber auch eine fortwährende Anpassung der Statistik, der Weiterentwicklung der Erhebungsmethodik und der Nutzung anderer Datenquellen. Die nächste größere Änderung ist die Einführung der CO₂-Bepreisung, die mit den Ergebnissen für das erste Halbjahr 2021 erfasst wird. Neue Datenquellen werden sich in Zukunft durch die Digitalisierung des Strom- und Gasmarktes ergeben und müssen für die Statistik nutzbar gemacht werden. 

LITERATURVERZEICHNIS

Bundesamt für Wirtschaft und Ausfuhrkontrolle. *Hintergrundinformationen zur Besonderen Ausgleichsregelung*. 2020. [Zugriff am 2. Juli 2021]. Verfügbar unter: www.bafa.de

Bundesministerium der Finanzen. *Rundschreiben: Befristete Absenkung des allgemeinen und ermäßigten Umsatzsteuersatzes zum 1. Juli 2020*. 2020. [Zugriff am 2. Juli 2021]. Verfügbar unter: www.bundesfinanzministerium.de

Bundesnetzagentur. *Monitoringbericht 2020*. 2020. [Zugriff am 2. Juli 2021]. Verfügbar unter: www.bundesnetzagentur.de

Bundesnetzagentur. *Wesentliche Elemente der Anreizregulierung*. 2021. [Zugriff am 2. Juli 2021]. Verfügbar unter: www.bundesnetzagentur.de

Burger, Alexander. *Liberalisierung des deutschen Strommarktes – ein Erfolg?* In: WiSt Wirtschaftswissenschaftliches Studium. Ausgabe 1/2014, Seite 45 ff.

Europäische Kommission. *Der europäische Grüne Deal*. 2019. [Zugriff am 5. Juli 2021]. Verfügbar unter: <https://ec.europa.eu>

Sachverständigenrat zur Begutachtung der gesamtwirtschaftlichen Entwicklung. *Corona-Krise gemeinsam bewältigen, Resilienz und Wachstum stärken. Jahresgutachten 2020/21*. Wiesbaden 2020. [Zugriff am 13. Juli 2021]. Verfügbar unter: www.sachverstaendigenrat-wirtschaft.de

Umweltbundesamt. *Verwirklichung des Potenzials der erneuerbaren Energien durch Höherauslastung des Bestandsnetzes und zügigen Stromnetzausbau auf Verteilnetzebene*. 2020. [Zugriff am 5. Juli 2021]. Verfügbar unter: www.umweltbundesamt.de

RECHTSGRUNDLAGEN

Energiesteuergesetz (EnergieStG) vom 15. Juli 2006 (BGBl. I Seite 1534; 2008 I Seite 660, 1007), das zuletzt durch Artikel 4 des Gesetzes vom 30. März 2021 (BGBl. I Seite 607) geändert worden ist.

Gesetz für den Ausbau erneuerbarer Energien (Erneuerbare-Energien-Gesetz – EEG 2021) vom 21. Juli 2014 (BGBl. I Seite 1066), das zuletzt durch Artikel 1 des Gesetzes vom 21. Dezember 2020 (BGBl. I Seite 3138) geändert worden ist.

Gesetz für die Erhaltung, die Modernisierung und den Ausbau der Kraft-Wärme-Kopplung (Kraft-Wärme-Kopplungsgesetz – KWKG 2020) vom 21. Dezember 2015 (BGBl. I Seite 2498), das zuletzt durch Artikel 17 des Gesetzes vom 21. Dezember 2020 (BGBl. I Seite 3138) geändert worden ist.

Gesetz über die Elektrizitäts- und Gasversorgung (Energiewirtschaftsgesetz – EnWG) vom 7. Juli 2005 (BGBl. I Seite 1970, 3621), das zuletzt durch Artikel 3 des Gesetzes vom 18. Mai 2021 (BGBl. I Seite 1122) geändert worden ist.

Gesetz über die Preisstatistik in der im Bundesgesetzblatt Teil III, Gliederungsnummer 720-9, veröffentlichten bereinigten Fassung, das zuletzt durch Artikel 9 des Gesetzes vom 22. Februar 2021 (BGBl. I Seite 266) geändert worden ist.

Gesetz über einen nationalen Zertifikatehandel für Brennstoffemissionen (Brennstoffemissionshandelsgesetz – BEHG) vom 12. Dezember 2019 (BGBl. I Seite 2728), das durch Artikel 1 des Gesetzes vom 3. November 2020 (BGBl. I Seite 2291) geändert worden ist.

Richtlinie 2003/87/EG des Europäischen Parlaments und des Rates vom 13. Oktober 2003 über ein System für den Handel mit Treibhausgasemissionszertifikaten in der Gemeinschaft und zur Änderung der Richtlinie 96/61/EG des Rates (Amtsblatt der EU Nr. L 275, Seite 32).

Richtlinie 2008/92/EG des Europäischen Parlaments und des Rates vom 22. Oktober 2008 zur Einführung eines gemeinschaftlichen Verfahrens zur Gewährleistung der Transparenz der vom industriellen Endverbraucher zu zahlenden Gas- und Strompreise (Neufassung) (Amtsblatt der EU Nr. L 298, Seite 9).

Stromsteuergesetz (StromStG) vom 24. März 1999 (BGBl. I Seite 378; 2000 I Seite 147), das zuletzt durch Artikel 6 des Gesetzes vom 30. März 2021 (BGBl. I Seite 607) geändert worden ist.

Verordnung (EU) 2016/1952 des Europäischen Parlaments und des Rates vom 26. Oktober 2016 über europäische Erdgas- und Strompreisstatistik und zur Aufhebung der Richtlinie 2008/92/EG (Amtsblatt der EU Nr. L 311, Seite 1).

Verordnung über die Entgelte für den Zugang zu Elektrizitätsversorgungsnetzen (Stromnetzentgeltverordnung – StromNEV) vom 25. Juli 2005 (BGBl. I Seite 2225), die zuletzt durch Artikel 34 des Gesetzes vom 23. Juni 2021 (BGBl. I Seite 1858) geändert worden ist.



Frauke Mischler

ist Diplom-Geographin und seit 2004 im Statistischen Bundesamt tätig. Ihre Schwerpunkte als Referentin im Referat „Verdienststrukturerhebung, Arbeitskostenerhebung, Tarifstatistiken“ liegen in der konzeptionellen und methodischen Weiterentwicklung der Verdienststrukturerhebung sowie in der Analyse des Gender Pay Gap.

VERDIENSTUNTERSCHIEDE ZWISCHEN MÄNNERN UND FRAUEN

Eine Ursachenanalyse auf Grundlage der Verdienststrukturerhebung 2018

Frauke Mischler

➤ **Schlüsselwörter:** geschlechtsspezifische Verdienstunterschiede – Verdienststrukturerhebung – Oaxaca-Blinder-Dekomposition – Lohnerhöhung – Gender Pay Gap – GPG-Sensitivität

ZUSAMMENFASSUNG

Anknüpfend an frühere Analysen stellt der Beitrag Ergebnisse zum unbereinigten und bereinigten Gender Pay Gap auf Basis der Verdienststrukturerhebung für das Jahr 2018 vor und erläutert die zugrunde gelegte Methodik. Eine modellbasierte Analyse zeigt die (eingeschränkte) Sensitivität des unbereinigten Gender Pay Gap bei einem hypothetisch höheren Mindestlohn für Frauen und Männer oder einem Verdienstzuschlag auf den Bruttostundenverdienst ausschließlich für Frauen. Deutlich wird hierbei, dass der unbereinigte Gender Pay Gap sich nur geringfügig bei einem hypothetischen Mindestlohn von 12 Euro ändert. Höher wäre die Wirkung bei einem einseitigen Verdienstzuschlag für Frauen um 50 Cent auf den Bruttostundenverdienst.

➤ **Keywords:** gender pay gap – structure of earnings survey – Oaxaca-Blinder decomposition – wage increase – GPG sensitivity

ABSTRACT

Following on from previous analyses, this article presents results on the unadjusted and adjusted gender pay gap (GPG) based on the structure of earnings survey for 2018 and explains the underlying methodology. A model-based analysis shows the (limited) sensitivity of the unadjusted GPG to a hypothetically higher minimum wage for women and men or an earnings supplement to gross hourly earnings for women only. It becomes clear here that the unadjusted gender pay gap will change only slightly in the case of a hypothetical minimum wage of 12 euros. The effect would be higher in the case of a unilateral earnings supplement of 50 cents to the gross hourly earnings of women.

1

Einleitung

Trotz der politisch und gesellschaftlich geforderten Chancengleichheit von Männern und Frauen liegt der durchschnittliche Bruttostundenverdienst von Arbeitnehmerinnen weiterhin deutlich unter dem der Arbeitnehmer. Der Verdienstabstand zwischen Frauen und Männern (Gender Pay Gap) ist daher sowohl in der Nachhaltigkeitsstrategie für Deutschland (Ziel 5: Geschlechtergleichstellung – Gleichstellung und partnerschaftliche Aufgabenteilung fördern; [Indikator 5.1.a Verdienstabstand zwischen Frauen und Männern](#)) nachzuweisen als auch als Teil des [Monitorings der Agenda 2030](#) der Vereinten Nationen.

Im europäischen Vergleich rangierte Deutschland mit dem unbereinigten Gender Pay Gap für das Berichtsjahr 2019 auf dem 24. Platz. Von den 27 EU-Staaten wies Luxemburg den geringsten Gender Pay Gap auf (Eurostat, 2021).

Aus der Verdienststrukturhebung 2018 liegen nun Daten vor, die tiefere Analysen für den unbereinigten und den bereinigten Gender Pay Gap ermöglichen. Im Zentrum der vorliegenden Auswertungen stehen die Ermittlung des bereinigten Gender Pay Gap sowie eine detaillierte Ursachenanalyse auf Basis der Verdienststrukturhebung 2018. Die Analysen knüpfen an Auswertungen zum bereinigten Gender Pay Gap auf Grundlage der Verdienststrukturhebungen 2006, 2010 und 2014 an.

Die der Berechnung des Gender Pay Gap zugrunde gelegte Methodik, die Verdienststrukturhebung als Datenbasis und deren Ergebnisse aus der Erhebung 2018 erläutern die Kapitel 2 bis 4.

Der zweite Teil des Beitrags beschäftigt sich in Kapitel 5 mit der Frage, wie sensitiv der Gesamtwert des unbereinigten Gender Pay Gap auf

- a) einen hypothetisch höheren Mindestlohn 2018 beziehungsweise
- b) einen (alleinigen) Verdienstaufschlag zum Bruttostundenverdienst für Frauen

reagiert. Höhere Mindestlöhne sind immer wieder in der politischen Diskussion und es stellt sich in diesem Zusammenhang die Frage, wie sich ein hypothetisch höherer Mindestlohn auf den Gender Pay Gap auswirken würde. Im Vergleich dazu soll überprüft werden, wie stark sich Verdienstaufschläge für Frauen je nach Höhe auf den Gender Pay Gap niederschlagen. Dabei steht weniger im Fokus, dass ein Verdienstaufschlag für Frauen den Gender Pay Gap verkleinert, sondern wie stark beziehungsweise weniger stark die Wirksamkeit wäre.

Der Beitrag schließt mit einem Fazit und einem Ausblick auf die Ergebnisse für 2019 und 2020.

2

Zur Berechnung des Gender Pay Gap

Der Begriff „Gender Pay Gap“ beschreibt den geschlechtsspezifischen Verdienstunterschied zwischen Männern und Frauen. Generell ist zwischen dem bereinigten und dem unbereinigten Gender Pay Gap zu unterscheiden. Der unbereinigte Gender Pay Gap vergleicht die absoluten Bruttostundenverdienste der Arbeitnehmerinnen und Arbeitnehmer miteinander und erfasst somit auch den Teil des Verdienstunterschieds, der sich auf unterschiedliche Qualifikationen, Tätigkeiten und Erwerbsbiografien zurückführen lässt. Der bereinigte Gender Pay Gap misst dagegen den Verdienstabstand von Frauen und Männern mit vergleichbaren Eigenschaften (soweit möglich). Die Modellrechnung ermöglicht es, den Teil des Verdienstunterschieds herauszurechnen, der auf strukturelle Unterschiede (beispielsweise unterschiedliche Bildungsstruktur) zwischen den Geschlechtergruppen zurückzuführen ist.

2.1 Unbereinigter Gender Pay Gap

Der unbereinigte Gender Pay Gap ist für internationale Vergleiche bedeutsam. Daher verwendet das Statistische Bundesamt bei seiner Berechnung die europaweit einheitlichen Vorgaben von Eurostat, dem Statistischen Amt der Europäischen Union (Finke, 2011, hier: Seite 37).

Der Gender Pay Gap wird als Differenz des durchschnittlichen Bruttostundenverdienstes der Männer und Frauen in Relation zu den durchschnittlichen Bruttostundenverdiensten der Männer definiert (Eurostat, 2008).

$$(1) \text{GPG}_{\text{unadj}} = \frac{\bar{y}_m - \bar{y}_f}{\bar{y}_m} \cdot 100$$

Dabei sind:

\bar{y}_m = durchschnittlicher Bruttostundenverdienst ohne Sonderzahlungen von Männern

\bar{y}_f = durchschnittlicher Bruttostundenverdienst ohne Sonderzahlungen von Frauen

Einbezogen werden die absoluten durchschnittlichen Bruttostundenverdienste aller Arbeitnehmerinnen und Arbeitnehmer. Andere lohnbestimmende Faktoren, beispielsweise die Berufserfahrung oder der Bildungsabschluss, werden nicht gesondert in der Berechnung berücksichtigt. Unregelmäßig anfallende Sonderzahlungen (zum Beispiel Urlaubs- oder Weihnachtsgeld) werden ausgeschlossen. Der unbereinigte Gender Pay Gap kann thematisch nach einzelnen Merkmalen (Altersgruppe, Branche, Bundesland) berechnet werden.

Nach europäischer Definition werden die Wirtschaftsbereiche B bis S (Produzierendes Gewerbe und Dienstleistungsbereich) ohne O (Öffentliche Verwaltung, Verteidigung; Sozialversicherung) und Betriebe ab zehn Beschäftigten in die Berechnung einbezogen. National liegen für die Berichtsjahre 2014 und 2018 Daten für die Wirtschaftsabschnitte A bis S (Land- und Forstwirtschaft, Fischerei, Produzierendes Gewerbe und Dienstleistungsbereich) und Betriebe ab einem Beschäftigten vor. Somit sind auch Auswertungen für die Gesamtwirtschaft möglich.

Berücksichtigt werden nicht nur die Verdienste von Arbeitnehmerinnen und Arbeitnehmern in Vollzeit, sondern auch in Teilzeit, Altersteilzeit sowie von geringfügig Beschäftigten. Ebenso werden Auszubildende in die Analyse mit einbezogen. Die Berechnung erfolgt bewusst auf Basis der Bruttostundenverdienste und nicht der Bruttomonats- oder Bruttojahresverdienste. Grund dafür ist, dass der Beschäftigungsumfang von Männern und Frauen das Ergebnis verfälschen würde, da Frauen häufiger in Teilzeit tätig sind als Männer (Eurostat, 2007, hier: Seiten 6 und 9). Die Definition gibt vor,

den Gender Pay Gap mit dem Bezug zu den Männerverdiensten zu berechnen.

2.2 Bereinigter Gender Pay Gap

Für die Berechnung des bereinigten Gender Pay Gap nutzt das Statistische Bundesamt die Oaxaca-Blinder-Dekomposition (Oaxaca, 1973; Blinder, 1973). Es handelt sich hierbei um die am häufigsten verwendete Methode zur Ermittlung von Verdienstunterschieden (Hübler, 2003, hier: Seite 557; Neumark, 2004, hier: Seite 8). Sie ermöglicht Aussagen zur Höhe des Unterschieds in den Bruttostundenverdiensten von Frauen und Männern mit vergleichbaren Eigenschaften.

Regressionsverfahren

Die Dekomposition basiert auf zwei nach dem Geschlecht differenzierten, semilogarithmischen Regressionsfunktionen. Mithilfe des Verfahrens der multiplen Regression wird der Einfluss unterschiedlicher Strukturmerkmale auf den logarithmierten Bruttostundenverdienst bestimmt (Finke, 2011, hier: Seite 38).

$$(2) \ln Y^F = \beta_0^F + \sum_{j=1}^n \beta_j^F x_j^F + e^F$$

$$(3) \ln Y^M = \beta_0^M + \sum_{j=1}^n \beta_j^M x_j^M + e^M$$

Dabei sind:

$\ln Y$ = logarithmierter Bruttostundenverdienst

β_j = Regressionskoeffizient eines Merkmals j

β_0 = Regressionskonstante

x_j = beobachtetes Merkmal j

e = Störterm

M = Männer

F = Frauen

➤ **Übersicht 1** zeigt die in die Berechnung als unabhängige Variable in das Regressionsmodell aufgenommenen Merkmale.

Übersicht 1

In die Regressionsanalysen einbezogene Variable

	Definition/Ausprägung
Unabhängige Variablen	
Höchster beruflicher Ausbildungsabschluss	<ul style="list-style-type: none"> › Dummy-Kodierung, Referenzkategorie: Abschluss einer anerkannten Berufsausbildung › Weitere Kategorien: ohne beruflichen Ausbildungsabschluss, Meister-/Techniker- oder gleichwertiger Fachschulabschluss, Bachelor, Diplom/Magister/Master/Staatsexamen und Promotion
Berufserfahrung ¹ Berufserfahrung quadriert ²	› Berufserfahrung = Alter – Ausbildungsjahre – 6
Dienstalter	› Dienstalter = Berichtsjahr – Eintrittsjahr
Leistungsgruppe	<ul style="list-style-type: none"> › Dummy-Kodierung, Referenzkategorie: angelernte Arbeitnehmer › Weitere Kategorien: Arbeitnehmer in leitender Stellung, herausgehobene Fachkräfte, Fachangestellte, ungelernte Arbeitnehmer, geringfügig Beschäftigte, Auszubildende
Berufshauptgruppe (nach ISCO-08 ³)	<ul style="list-style-type: none"> › Dummy-Kodierung, Referenzkategorie: ISCO 4 (Bürokräfte und verwandte Berufe) › Weitere Kategorien: ISCO 1, ISCO 2, ISCO 3, ISCO 5, ISCO 6, ISCO 7, ISCO 8, ISCO 9
Art des Arbeitsvertrags	<ul style="list-style-type: none"> › Dummy-Kodierung, Referenzkategorie: unbefristet › Weitere Kategorie: befristet
Beschäftigungsumfang	<ul style="list-style-type: none"> › Dummy-Kodierung, Referenzkategorie: vollzeitbeschäftigt › Weitere Kategorie: teilzeitbeschäftigt
Altersteilzeit	<ul style="list-style-type: none"> › Dummy-Kodierung, Referenzkategorie: keine Altersteilzeit › Weitere Kategorie: Altersteilzeit
Tariffbindung des Betriebs ⁴	<ul style="list-style-type: none"> › Dummy-Kodierung, Referenzkategorie: keine Tariffbindung › Weitere Kategorie: mit Tariffbindung
Zulagen für Schicht-, Wochenend-, Feiertags- und Nachtarbeit	<ul style="list-style-type: none"> › Dummy-Kodierung, Referenzkategorie: keine Zulagen › Weitere Kategorie: Zulagen
Gebietsstand	<ul style="list-style-type: none"> › Dummy-Kodierung, Referenzkategorie: Früheres Bundesgebiet (einschließlich Land Berlin) › Weitere Kategorie: Neue Länder
Differenzierter Regionstyp ⁵	<ul style="list-style-type: none"> › Dummy-Kodierung, Referenzkategorie: ländlicher Raum › Weitere Kategorie: städtischer Raum
Unternehmensgröße	<ul style="list-style-type: none"> › Dummy-Kodierung, Referenzkategorie: 1 000 Arbeitnehmer und mehr › Weitere Kategorien: 10 bis 49 Arbeitnehmer, 50 bis 249 Arbeitnehmer, 250 bis 499 Arbeitnehmer, 500 bis 999 Arbeitnehmer
Einfluss der öffentlichen Hand auf die Unternehmensführung	<ul style="list-style-type: none"> › Dummy-Kodierung, Referenzkategorie: kein oder eingeschränkter Einfluss der öffentlichen Hand auf die Unternehmensführung durch Kapitalbeteiligung (50 % und weniger), Satzung oder sonstige Bestimmungen › Weitere Kategorie: beherrschender Einfluss der öffentlichen Hand auf die Unternehmensführung durch Kapitalbeteiligung (mehr als 50 %), Satzung oder sonstige Bestimmungen
Wirtschaftsgruppe	› Dummy-Kodierung, Referenzkategorie: Wirtschaftsgruppe Q 86.1 (Krankenhäuser)
Abhängige Variable	
ln(Bruttostundenverdienst)	<ul style="list-style-type: none"> › Logarithmierter Bruttostundenverdienst › Bruttostundenverdienst = (Bruttomonatsverdienst – Sonderzahlungen)/bezahlte Stunden

1 Um die Berufserfahrung als erklärende Variable auch in Untersuchungen berücksichtigen zu können, in denen das Merkmal nicht direkt abgefragt wurde, erfolgt in der Regel eine näherungsweise Ermittlung dieser Größe über das Alter und die Ausbildungsdauer (Achatz und andere, 2005). Den Arbeitnehmerinnen und Arbeitnehmern wird im Rahmen dieses Vorgehens ein unterbrochener Erwerbsverlauf unterstellt. Da jedoch vor allem bei Frauen Erwerbsunterbrechungen zu beobachten sind, führt dies insbesondere bei Arbeitnehmerinnen zu einer Überschätzung der Berufserfahrung (Hinz/Gartner, 2005, hier: Seite 26). Zur Umrechnung des höchsten (Aus-)Bildungsabschlusses in Jahre siehe Finke und andere, 2017, hier: Tabelle 2. Aufgrund von Änderungen in Bezug auf die Erfassung der Bildungsinformationen in der Verdienststrukturerhebung ergeben sich Unterschiede zwischen den für 2006 und 2010 durchgeführten Berechnungen und den vorliegenden Analysen.

2 An dieser Stelle wird auch der quadrierte Term in die Regressionsanalyse einbezogen, da zwischen der Berufserfahrung und dem Verdienst ein kurvilinearere Zusammenhang besteht.

3 ISCO (International Standard Classification of Occupations) ist eine von der Internationalen Arbeitsorganisation (ILO) vorgenommene Klassifikation der Berufe.

4 Für die Berechnung des bereinigten Gender Pay Gap 2006 und 2010 wurden Angaben zur Tariffindung des Arbeitnehmers herangezogen. Für die Analysen der Verdienststrukturerhebung 2018, wie auch der Verdienststrukturerhebung 2014, erfolgte die Berücksichtigung der Tariffindung des Betriebs. Die Betriebsangabe ist die Angabe, die üblicherweise für Analysen der Tariffindung genutzt wird. Vergleichsrechnungen, in denen – ceteris paribus – die Tariffindung des Arbeitnehmers beziehungsweise alternativ die des Betriebs herangezogen wurde, zeigen keine relevanten Auswirkungen auf die Höhe des bereinigten Gender Pay Gap. Auch der Beitrag beider Merkmale auf den erklärten Anteil des Gender Pay Gap unterscheidet sich nur unwesentlich.

5 Während unter „Ballungsraum“ für die 2006 und 2010 durchgeführten Analysen Kreistypen subsumiert wurden, die entsprechend der Typologie des Bundesamtes für Bauwesen und Raumordnung (BBR) als Kernstädte in Agglomerationen und (hoch-)verdichtete Kreise in Agglomerationen definiert wurden, fasste die Ausprägung „kein Ballungsraum“ alle übrigen Kreistypen zusammen. Da die Kreistypen in der beschriebenen Form nicht mehr existieren, wurde im Zusammenhang mit den vorliegenden Untersuchungen auf den in der Verdienststrukturerhebung 2018 wie auch der Verdienststrukturerhebung 2014 vorliegenden differenzierten Regionstyp zurückgegriffen. Dieser kategorisiert die Regionen in städtische und ländliche Räume. Die Nutzung dieser Information soll die Vergleichbarkeit mit den Ergebnissen der bisherigen Untersuchungen gewährleisten. Für Informationen zur aktuellen Kategorisierung der siedlungsstrukturellen Kreistypen siehe www.bbsr.bund.de (zuletzt abgerufen am 6. Juli 2021).

Die im Zusammenhang mit den Ergebnissen der Regressionsanalyse dargestellten β -Koeffizienten sind als sogenannte partielle Regressionskoeffizienten zu interpretieren. Diese Koeffizienten geben die Effektstärke einer unabhängigen Variablen auf die abhängige Variable unter Konstanthaltung der Effekte aller übrigen in das Modell einbezogenen Merkmale an (Finke, 2011, hier: Seite 40).

Dekompositionsverfahren

Durch Subtraktion und Umformung der geschlechtsspezifischen Lohnfunktion ergibt sich folgende Gleichung (Finke und andere, 2017, hier: Seite 45):

$$(4) \overline{\ln Y_M} - \overline{\ln Y_F} = \underbrace{(\beta_0^M - \beta_0^F) + \sum_j \bar{x}_j^F (\beta_j^M - \beta_j^F)}_{\text{unexplained gap}} + \underbrace{\sum_j \beta_j^M (\bar{x}_j^M - \bar{x}_j^F)}_{\text{explained gap}}$$

Dabei sind:

$\ln Y$ = logarithmierter Bruttostundenverdienst

β_j = Regressionskoeffizient eines Merkmals j

β_0 = Regressionskonstante

\bar{x}_j = arithmetisches Mittel eines Merkmals j

M = Männer

F = Frauen

Die Lohndifferenz wird in zwei Bestandteile zerlegt: den erklärten und den unerklärten Gender Pay Gap. Der erklärte Gender Pay Gap beschreibt den Teil des Verdienstunterschieds, der auf geschlechtsspezifische Unterschiede in den erklärenden Variablen zurückzuführen ist. Man bezeichnet ihn auch als Merkmals- oder Ausstattungseffekt.

Mit dem unerklärten Gender Pay Gap (bereinigter Gender Pay Gap) wird der Teil des Verdienstunterschieds deutlich, der sich nicht durch bekannte geschlechtsspezifische Ausstattungsunterschiede begründen lässt. Er teilt sich in zwei Bereiche auf. Ein Teil basiert auf Unterschieden hinsichtlich der Koeffizienten. Er hängt also mit der unterschiedlichen Bewertung der erklärenden Variablen zusammen. Ein Masterabschluss wird beispielsweise für

Frauen und Männer finanziell unterschiedlich belohnt (zweiter Term). Der andere Teil kann nicht auf die erklärenden Variablen zurückgeführt werden (erster Term). Selbst wenn Frauen und Männer identische relevante Ausstattungsmerkmale haben und dafür gleich entlohnt würden, bleibt ein Verdienstunterschied bestehen, der nicht auf die berücksichtigten Modellvariablen zurückgeführt werden kann (Finke und andere, 2017, hier: Seite 45; Beck, 2018, hier: Seite 29).

Bei der Interpretation ist einerseits zu beachten, dass die ungleiche Verteilung der Geschlechtergruppen nach bestimmten Ausstattungsmerkmalen, zum Beispiel die ungleiche Verteilung auf Wirtschaftsbranchen, möglicherweise selbst schon das Resultat gesellschaftlich benachteiligender Strukturen sein könnte, und somit das Ausmaß der Benachteiligung gegebenenfalls unterschätzt würde (Finke, 2011, hier: Seite 41).

Andererseits ist zu berücksichtigen, dass der ermittelte Wert des unerklärten Gender Pay Gap eine Obergrenze ist. Dieser Wert würde geringer ausfallen, stünden weitere lohnrelevante Einflussfaktoren zur Erklärung von geschlechtsspezifischen Verdienstunterschieden zur Verfügung. Beispiele dafür sind Angaben zu Erwerbsunterbrechungen, zum individuellen Verhalten in Lohnverhandlungen oder zur tatsächlichen Berufserfahrung, die in der Verdienststrukturerhebung nicht erhoben werden (Finke und andere, 2017, hier: Seite 45).

3

Datenbasis

Seit 2006 erfolgen die Berechnungen zum unbereinigten und bereinigten Gender Pay Gap auf Basis der Verdienststrukturerhebung. Ziel ist, europaweit harmonisiert zuverlässige Daten über die Verteilung der Verdienste zu gewinnen und zu den wichtigsten Faktoren, von denen diese abhängen. Die Verdienststrukturerhebung wird in Deutschland als geschichtete Stichprobe durchgeführt; die Statistischen Ämter der Länder erheben alle vier Jahre bei etwa 60 000 Betrieben die Angaben zu Verdiensten und ihre bestimmenden Faktoren (Statistisches Bundesamt, 2020a). Im Zuge der Erhebung werden nicht nur Angaben zu Bruttoverdiensten der Beschäftigten, sondern auch zu einer Vielzahl von lohnbestimmenden Merkmalen ermittelt. Für die Analyse der

Verdienstunterschiede zwischen Männern und Frauen ist diese Datenquelle daher besonders gut geeignet. Zu den erhobenen arbeitnehmerbezogenen Angaben zählen das Geschlecht, das Geburtsjahr, der Schulabschluss, der Ausbildungsabschluss, die ausgeübte Tätigkeit, die Art der Beschäftigung (befristet/unbefristet) sowie der Beschäftigungsumfang (Vollzeit/Teilzeit). Weiterhin werden auch Angaben zum Betrieb erhoben, beispielsweise der Standort, der Wirtschaftszweig, die Betriebsgröße und, ob ein Tarifvertrag angewendet wird. Selbstständige sind in die Erhebung nicht eingeschlossen.

4

Ergebnisse für das Berichtsjahr 2018

Nach der europaweit einheitlichen Definition wurden in die Auswertungen die Wirtschaftsbereiche B bis S (Produzierendes Gewerbe und Dienstleistungsbereich) ohne O (Öffentliche Verwaltung, Verteidigung; Sozialversicherung) und Betriebe ab zehn Beschäftigten einbezogen.

4.1 Unbereinigter Gender Pay Gap

Nach den Auswertungen aus der Verdienststrukturerhebung 2018 lag der durchschnittliche Bruttostundenverdienst von Frauen mit 17,33 Euro um 20,1 % unter dem der Männer (21,70 Euro). Im Vergleich zu 2014 ist der Gender Pay Gap damit von 22,3 % um 2,2 Prozentpunkte gesunken. Nach wie vor fällt der unbereinigte Gender Pay Gap in den neuen Bundesländern deutlich geringer aus als im früheren Bundesgebiet (einschließlich Berlins). Im Jahr 2018 betrug der unbereinigte Gender Pay Gap im Westen 21,4 %, während er im Osten bei 7,3 % lag. [↗ Tabelle 1](#)

Tabelle 1

Entwicklung des unbereinigten Gender Pay Gap

	Verdienststrukturerhebung 2014			Verdienststrukturerhebung 2018		
	Bruttostundenverdienst Frauen	Bruttostundenverdienst Männer	Gender Pay Gap	Bruttostundenverdienst Frauen	Bruttostundenverdienst Männer	Gender Pay Gap
	EUR		%	EUR		%
Deutschland	15,44	19,87	22,3	17,33	21,70	20,1
Früheres Bundesgebiet einschließlich Berlins	15,73	20,58	23,6	17,58	22,36	21,4
Neue Bundesländer	13,52	14,78	8,5	15,64	16,87	7,3

Ergebnisse der Verdienststrukturerhebung. Wirtschaftsbereiche B bis S (Produzierendes Gewerbe und Dienstleistungsbereich) ohne O (Öffentliche Verwaltung, Verteidigung; Sozialversicherung). Ohne Kleinbetriebe.

4.2 Bereinigter Gender Pay Gap

Ausgehend vom unbereinigten Gender Pay Gap (20,1 %) lassen sich 14,2 Prozentpunkte und damit 70,8 % durch die in die Analyse einbezogenen Faktoren erklären. Die restlichen 5,9 Prozentpunkte beziehungsweise 29,2 % des Lohnabstands sind nicht durch die berücksichtigten Merkmale zu erklären, sie entsprechen dem bereinigten Gender Pay Gap. Dies bedeutet, dass der Gender Pay Gap um rund 14 Prozentpunkte geringer wäre, wenn Männer und Frauen über dieselben, in der Analyse berücksichtigten Ausstattungsmerkmale verfügten. [↗ Tabelle 2](#)

Tabelle 2

Zerlegung des Gender Pay Gap 2018

	Verdienstunterschied	Anteil
	%	
Unbereinigter Gender Pay Gap	20,1	100
Erklärter Teil	14,2	70,8
Unerklärter Teil (= bereinigter Gender Pay Gap)	5,9	29,2

Ergebnisse der Verdienststrukturerhebung. Wirtschaftsbereiche B bis S (Produzierendes Gewerbe und Dienstleistungsbereich) ohne O (Öffentliche Verwaltung, Verteidigung; Sozialversicherung). Ohne Kleinbetriebe.

Erkenntnisse aus den Regressionsanalysen – separat für Frauen und Männer

Ein Blick auf die Einzelregressionsmodelle für Männer und Frauen zeigt, dass sich 80 % der Varianz des Bruttostundenverdienstes der Männer erklären lassen; bei den Frauen sind dies 77 %. Es geht sowohl bei Männern als auch bei Frauen ein signifikanter Einfluss von einem Großteil der Variablen aus.

Die Regressionskoeffizienten in [↗ Tabelle 3](#) geben näherungsweise die relative Änderung der abhängigen Variablen an, wenn sich die unabhängige Variable um eine Einheit ändert. Merkmale, die als sogenannte Dummy-

Tabelle 3
Ergebnisse der Regressionsanalyse auf Basis der Verdienststrukturerhebung 2018

	Männer		Frauen	
	Koeffizient	p	Koeffizient	p
Neue Länder (Referenz: Früheres Bundesgebiet einschließlich Berlins)	- 0,165	<.0001	- 0,117	<.0001
Städtischer Raum (Referenz: ländlicher Raum)	0,031	<.0001	0,041	<.0001
Berufshauptgruppen nach ISCO-08 (Referenz: ISCO 4)				
ISCO 1	0,373	<.0001	0,261	<.0001
ISCO 2	0,164	<.0001	0,116	<.0001
ISCO 3	0,111	<.0001	0,052	<.0001
ISCO 5	0,012	<.0001	- 0,089	<.0001
ISCO 6	- 0,011	0,096	- 0,040	0,000
ISCO 7	0,014	<.0001	- 0,087	<.0001
ISCO 8	- 0,021	<.0001	- 0,117	<.0001
ISCO 9	- 0,044	<.0001	- 0,134	<.0001
Dienstalter	0,005	<.0001	0,005	<.0001
Unternehmensgröße (Referenz: 1 000 Arbeitnehmer und mehr)				
10 bis 49 Arbeitnehmer	- 0,151	<.0001	- 0,104	<.0001
50 bis 249 Arbeitnehmer	- 0,101	<.0001	- 0,072	<.0001
250 bis 499 Arbeitnehmer	- 0,056	<.0001	- 0,038	<.0001
500 bis 999 Arbeitnehmer	- 0,027	<.0001	- 0,022	<.0001
Höchster beruflicher Ausbildungsabschluss (Referenz: Abschluss einer anerkannten Berufsausbildung)				
Ohne beruflichen Ausbildungsabschluss	- 0,034	<.0001	- 0,023	<.0001
Meister-/Techniker- oder gleichwertiger Fachschulabschluss	0,046	<.0001	0,046	<.0001
Bachelor	0,058	<.0001	0,049	<.0001
Diplom/Magister/Master/Staatsexamen	0,162	<.0001	0,140	<.0001
Promotion	0,293	<.0001	0,361	<.0001
Leistungsgruppe (Referenz: angelernte Arbeitnehmer)				
Arbeitnehmer in leitender Stellung	0,542	<.0001	0,501	<.0001
Herausgehobene Fachkräfte	0,347	<.0001	0,331	<.0001
Fachangestellte	0,133	<.0001	0,133	<.0001
Ungelernte Arbeitnehmer	- 0,075	<.0001	- 0,045	<.0001
Geringfügig Beschäftigte	- 0,101	<.0001	- 0,071	<.0001
Auszubildende	- 0,884	<.0001	- 0,798	<.0001
Tariffindung des Betriebs (Referenz: keine Tariffindung)	0,066	<.0001	0,079	<.0001
Potenzielle Berufserfahrung				
Berufserfahrung	0,012	<.0001	0,009	<.0001
Berufserfahrung (quadriert)	0,000	<.0001	0,000	<.0001
Befristeter Arbeitsvertrag (Referenz: unbefristeter Arbeitsvertrag)	- 0,071	<.0001	- 0,061	<.0001
Teilzeitbeschäftigt (Referenz: vollzeitbeschäftigt)	- 0,044	<.0001	- 0,008	<.0001
Beherrschender Einfluss der öffentlichen Hand (Referenz: kein oder eingeschränkter Einfluss der öffentlichen Hand)	- 0,028	<.0001	- 0,015	<.0001
Altersteilzeit (Referenz: keine Altersteilzeit)	0,311	<.0001	0,291	<.0001
Zulagen für Schicht-, Wochenend-, Feiertags- und Nachtarbeit (Referenz: keine Zulagen)	0,062	<.0001	0,076	<.0001
Wirtschaftsgruppen-Dummies	ja ¹	ja ¹	ja ¹	ja ¹
Konstante	2,614		2,559	
R ²	0,797		0,773	
Korrigiertes R ²	0,796		0,772	
n (Stichprobenumfang)	419 613		349 577	
Fallzahl hochgerechnet	16 435 250		14 465 894	
Abhängige Variable: logarithmierter Bruttostundenverdienst				

Wirtschaftsbereiche B bis S (Produzierendes Gewerbe und Dienstleistungsbereich) ohne O (Öffentliche Verwaltung, Verteidigung; Sozialversicherung) und ohne Kleinstbetriebe.

1 Die Regressionskoeffizienten und die dazugehörigen beobachteten Signifikanzniveaus der Wirtschaftsgruppendifferenzen können auf Anfrage bei der Autorin bezogen werden. Aus Übersichtsgründen wurde auf eine Darstellung der entsprechenden Werte verzichtet.

Variablen kodiert wurden, sind im Rahmen der Regressionsanalyse stets in Bezug auf die jeweilige Referenzgruppe zu interpretieren. Der Koeffizient von $-0,151$ bei der Regressionsanalyse der Männer der Dummy-Variable „Unternehmensgröße 10 bis 49 Arbeitnehmer“, Referenzgruppe: 1 000 Arbeitnehmer und mehr, ist beispielhaft wie folgt zu interpretieren: Männliche Arbeitnehmer, die in Unternehmen mit 10 bis 49 Arbeitnehmern tätig sind, haben gegenüber männlichen Beschäftigten in Unternehmen mit über 1 000 Arbeitnehmern einen um 15,1 % geringeren Stundenverdienst.

Werden alle anderen in die Analyse einbezogenen Merkmale konstant gehalten, ergeben sich unter anderem folgende Ergebnisse:

- › Arbeitnehmerinnen in leitender Stellung verdienen 50 % mehr als angelernte Arbeitnehmerinnen. Arbeitnehmer profitierten mit 54 % sogar noch etwas stärker von ihrer Position in leitender Stellung.
- › Der durchschnittliche Bruttostundenverdienst von Frauen in Unternehmen mit 10 bis 49 Arbeitnehmern lag um 10 % unter dem von Frauen in Unternehmen mit 1 000 und mehr Beschäftigten. Männer in kleineren Unternehmen verdienen sogar 15 % weniger als ihre Kollegen in großen Unternehmen mit über 1 000 Beschäftigten.
- › „Führungskräfte“ (ISCO 1) erzielten einen um 37 % höheren Verdienst als männliche Bürokräfte. Für die Kolleginnen betrug diese Differenz 26 %. „Hilfsarbeitskräfte“ erhielten im Vergleich zu Bürokräften (ISCO 4) dagegen einen niedrigeren Verdienst.
- › Beschäftigte im Osten verdienen weniger als Beschäftigte im Westen. Dieser Effekt ist bei den Männern etwas stärker ausgeprägt als bei den Frauen.
- › Arbeitnehmerinnen und Arbeitnehmer mit einem Hochschulabschluss wiesen wesentlich höhere Verdienste auf als Beschäftigte mit abgeschlossener Berufsausbildung.

Erkenntnisse aus der detaillierten Zerlegung des unbereinigten Gender Pay Gap

Die detaillierte Zerlegung des unbereinigten Gender Pay Gap zeigt den Beitrag einzelner Faktoren auf den erklärten und unerklärten Teil des Gender Pay Gap.

🔗 [Tabelle 4 auf Seite 118](#)

Bei Betrachtung des erklärten Teils des Gender Pay Gap (erste Spalte in Tabelle 4) zeigt sich, dass 6,2 Prozentpunkte auf den höheren Anteil von Frauen in eher niedrig bezahlten Tätigkeiten zurückzuführen sind. „Tätigkeit“ bezeichnet in diesem Zusammenhang den gemeinsamen Erklärungsbeitrag der Faktoren „Beruf“ und „Wirtschaftszweig“, da beide Beiträge inhaltlich kaum zu trennen sind. Zudem lassen sich 4,3 Prozentpunkte auf Unterschiede in den Leistungsgruppen zurückführen, während 2 Prozentpunkte damit zusammenhängen, dass Frauen eher in Teilzeit arbeiten. Weitere 0,6 Prozentpunkte lassen sich darauf zurückführen, dass Frauen auch eher geringfügig beschäftigt sind.

Neben dem erklärten Teil des Gender Pay Gap ist auch der unerklärte Teil des Gender Pay Gaps in Tabelle 4 dargestellt (Spalte 3). Hier spielt die Konstante die bedeutendste Rolle. Ihr positiver Wert weist darauf hin, dass Frauen unabhängig von den berücksichtigten Einflussfaktoren im Durchschnitt 1,6 Prozentpunkte weniger verdienen würden. Der unerklärte Teil des Gender Pay Gap enthält jedoch auch Effekte, die auf lohnrelevante Faktoren zurückgehen, die nicht bei der Modellrechnung berücksichtigt wurden. Denkbar sind hier beispielsweise Erwerbsunterbrechungen oder das individuelle Verhalten bei Lohnverhandlungen. Zu diesen Faktoren liegen in der Verdienststrukturerhebung keine Informationen vor. Sie werden von der Konstanten aufgefangen. Daher lässt sich der Wert der Konstanten nicht allein auf Diskriminierung von Arbeitnehmerinnen zurückführen (Achatz und andere, 2005, hier: Seite 478). Der Rest des unerklärten Teils des Gender Pay Gap resultiert daraus, dass bei gleicher Ausstattung die Entlohnung auf dem Arbeitsmarkt für einige Merkmale zu Ungunsten von Frauen und für andere Merkmale benachteiligend gegenüber Männern ausfällt. Am stärksten wirkt sich das Merkmal Berufserfahrung aus. Ein zusätzliches Jahr an Berufserfahrung wird bei Frauen entsprechend der Ergebnisse dieser Analyse in geringerem Maße belohnt als bei Männern. Es handelt sich hierbei allerdings lediglich um die potenzielle und nicht die tatsächliche Berufserfahrung (siehe Übersicht 1). Darüber hinaus wirken auch der Gebietsstand und die Leistungsgruppe zu Lasten von Frauen. Bezogen auf die Leistungsgruppen ist jedoch zu berücksichtigen, dass auch innerhalb der Leistungsgruppen eine vertikale Segregation besteht. Dieser Effekt kann nicht erfasst werden und schlägt sich daher im unerklärten Teil des Gender Pay Gap nieder.

Tabelle 4**Detaillierte Zerlegung des unbereinigten Gender Pay Gap nach Erklärungsfaktoren 2018**

	Erklärter Teil des Gender Pay Gap	Anteil des erklärten Teils am gesamten Gender Pay Gap	Unerklärter Teil des Gender Pay Gap: bereinigter Gender Pay Gap	Anteil des unerklärten Teils am gesamten Gender Pay Gap
	Prozentpunkte	%	Prozentpunkte	%
Konstante	X	X	1,6	7,8
Gebietsstand	0,2	0,8	2,1	10,3
Differenzierter Regionstyp	0,1	0,7	– 0,2	– 0,9
Tätigkeit	6,2	30,6	– 3,3	– 16,2
Dienstalter	0,1	0,5	– 0,2	– 1,2
Unternehmensgröße	– 0,3	– 1,6	– 0,3	– 1,5
Ausbildungsabschluss	0,5	2,7	1,0	5,0
Leistungsgruppen	4,3	21,2	1,6	7,8
Geringfügige Beschäftigung	0,6	3,2	– 0,3	– 1,4
Auszubildende	– 0,2	– 1,1	– 0,3	– 1,4
Tarifbindung	– 0,1	– 0,6	0,1	0,3
Potenzielle Berufserfahrung	0,0	0,1	4,3	21,5
Art des Arbeitsvertrags	0,3	1,3	0,3	1,6
Beschäftigungsumfang	2,0	9,8	– 0,4	– 1,8
Einfluss der öffentlichen Hand auf die Unternehmensführung	0,2	1,1	0,5	2,5
Altersteilzeit	0,1	0,5	– 1,1	– 5,7
Zulagen	0,3	1,4	0,5	2,4
Insgesamt	14,2	70,8	5,9	29,2

Ergebnisse der Verdienststrukturerhebung, Wirtschaftsbereiche B bis S (Produzierendes Gewerbe und Dienstleistungsbereich) ohne O (Öffentliche Verwaltung, Verteidigung; Sozialversicherung). Ohne Kleinbetriebe.

Zu Lasten der Männer wirken insbesondere die Faktoren Tätigkeit und Altersteilzeit.

5

Sensitivität des unbereinigten Gender Pay Gap auf Lohnerhöhung am unteren Rand

Die Entwicklung des unbereinigten Gender Pay Gap ist nicht nur einmal im Jahr anlässlich des Equal Pay Day ein politisch stark diskutiertes Thema. Die Bundesregierung hat sich zum Ziel gesetzt, den unbereinigten Gender Pay Gap bis zum Jahr 2030 auf 10 % zu reduzieren (Bundesregierung, 2018). Im Jahr 2018 lag er bei 20,1 %. Neben indirekten Maßnahmen (Frauen in Führungspositionen, Ausbau der Kinderbetreuung) könnten eine Lohnerhöhung am unteren Rand (höherer Mindestlohn) oder auch ein einseitiger Verdienstaufschlag für Frauen mögliche Mittel darstellen.

5.1 Vorgehen und Annahmen

Im Folgenden wird die Wirkungsweise eines höheren Mindestlohns beziehungsweise eines Verdienstaufschlags auf den unbereinigten Gender Pay Gap mithilfe eines einfachen Rechenmodells auf Basis der Daten der Verdienststrukturerhebung 2018 mit folgenden Annahmen simuliert:

1. Erhöhter Mindestlohn

- › Der Bruttostundenverdienst wird für alle Datensätze unterhalb des neuen simulierten Mindestlohns (zum Beispiel 12 Euro) auf diesen gesetzt, wenn die Beschäftigten aktuell unter das Mindestlohngesetz fallen würden (der Bruttostundenverdienst von Auszubildenden, Praktikanten und unter 18-Jährigen wird damit per Definition nicht angepasst).
- › Vom erhöhten Mindestlohn profitieren Frauen wie Männer gleichermaßen.

- › Die Analyse erfolgt unter sonst gleichen Bedingungen, das heißt ohne Sekundäreffekte des höheren Mindestlohns auf obere Lohngruppen beziehungsweise eine mögliche Stundenreduzierung der durch die Mindestlohnerhöhung Besserverdienenden.

2. Einseitiger Verdienstaufschlag für Frauen

- › Jede Frau erhält unabhängig von ihrer Position und der Art der Beschäftigung einen Aufschlag auf ihren Bruttostundenverdienst.

Um die Vergleichbarkeit mit den Ergebnissen zum Gender Pay Gap zu gewährleisten, werden nach europäischer Definition die Wirtschaftsbereiche B bis S (Produzierendes Gewerbe und Dienstleistungsbereich) ohne O (Öffentliche Verwaltung, Verteidigung; Sozialversicherung) und Betriebe ab zehn Beschäftigten einbezogen.

5.2 Erhöhter Mindestlohn und Verdienstaufschlag für Frauen

▢ Grafik 1 zur Verteilung der Bruttostundenverdienste zeigt, dass im Vergleich zu den Männern ein deutlich höherer Anteil an Frauen im April 2018 zwischen 9 und 11 Euro je Stunde verdient hat. Von einer Erhöhung des Mindestlohns auf 12 Euro könnten diese Frauen profitieren.

ren und es wäre denkbar, dass der unbereinigte Gender Pay Gap sich dadurch verbessert, also geringer ausfällt.

Das Ergebnis der Simulationsrechnung ergab, dass eine Erhöhung des Mindestlohns auf 12 Euro für alle einbezogenen Beschäftigten (außer Auszubildende, Praktikanten und Beschäftigte jünger als 18 Jahre) unter sonst gleichen Bedingungen den unbereinigten Gender Pay Gap lediglich um 1,2 Prozentpunkte auf 18,9% senken würde. Hätte der Mindestlohn 15 Euro betragen, läge der unbereinigte Gender Pay Gap bei 16,8%.

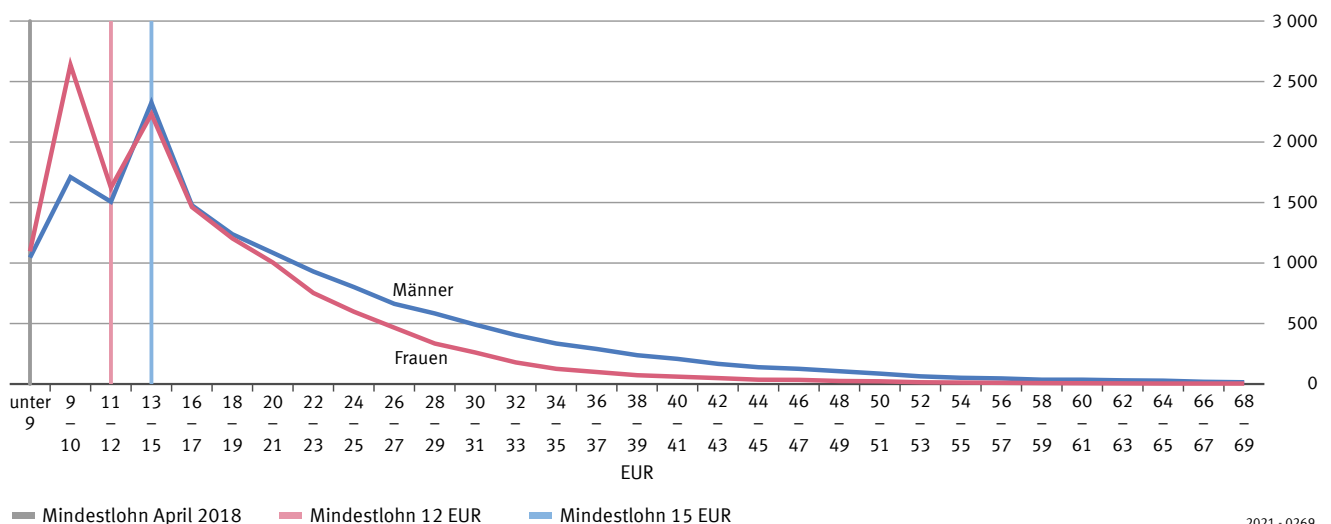
Ein weiterer Ansatz wäre, den Mindestlohn gezielt nur in Wirtschaftsbereichen zu erhöhen, die besonders stark von Frauen besetzt sind.

Auswertungen aus dem Mikrozensus zeigen, dass in folgenden Wirtschaftsbereichen Frauen besonders häufig tätig sind:

- › G Handel; Instandhaltung und Reparatur von Kraftfahrzeugen,
- › I Gastgewerbe,
- › K Erbringung von Finanz- und Versicherungsdienstleistungen,
- › M Erbringung von freiberuflichen, wissenschaftlichen und technischen Dienstleistungen,

Grafik 1

Beschäftigungsverhältnisse nach Bruttostundenverdiensten im April 2018
in 1000



- › N Erbringung von sonstigen wirtschaftlichen Dienstleistungen,
- › P Erziehung und Unterricht,
- › Q Gesundheits- und Sozialwesen,
- › R Kunst, Unterhaltung und Erholung sowie
- › S Erbringung von sonstigen Dienstleistungen (Statistisches Bundesamt, 2018).

Innerhalb dieser Wirtschaftsbereiche liegt der Anteil der Frauen bei mindestens 49 %. Für die weiteren Auswertungen wurde der Wirtschaftsbereich R ausgeschlossen, da hier die Anzahl der Frauen insgesamt zu gering ist, um einen nachhaltigen Effekt auf den gesamtwirtschaftlichen Gender Pay Gap bewirken zu können. ➔ [Tabelle 5](#)

Tabelle 5

Anteile erwerbstätiger Frauen nach Wirtschaftszweigen im Jahr 2018

	2018	Anteil der Frauen im Wirtschaftszweig		2018	Anteil der Frauen im Wirtschaftszweig
	1 000	%		1 000	%
A Land- und Forstwirtschaft, Fischerei	523	X	K Erbringung von Finanz- und Versicherungsdienstleistungen	1 254	X
Frauen	165	32	Frauen	635	51
Männer	358	X	Männer	619	X
B Bergbau und Gewinnung von Steinen und Erden	72	X	L Grundstücks- und Wohnungswesen	206	X
Frauen	9	13	Frauen	98	48
Männer	63	X	Männer	108	X
C Verarbeitendes Gewerbe	7 991	X	M Erbringung von freiberuflichen, wissenschaftlichen und technischen Dienstleistungen	2 381	X
Frauen	2 191	27	Frauen	1 169	49
Männer	5 800	X	Männer	1 212	X
D Energieversorgung	320	X	N Erbringung von sonstigen wirtschaftlichen Dienstleistungen	2 098	X
Frauen	85	27	Frauen	1 049	50
Männer	235	X	Männer	1 049	X
E Wasserversorgung; Entsorgung, Beseitigung von Umweltverschmutzung	250	X	O Öffentliche Verwaltung, Verteidigung; Sozialversicherung	2 883	X
Frauen	48	19	Frauen	1 449	50
Männer	202	X	Männer	1 434	X
F Baugewerbe	2 821	X	P Erziehung und Unterricht	2 816	X
Frauen	389	14	Frauen	2 019	72
Männer	2 432	X	Männer	797	X
G Handel; Instandhaltung und Reparatur von Kraftfahrzeugen	5 836	X	Q Gesundheits- und Sozialwesen	5 449	X
Frauen	2 898	50	Frauen	4 187	77
Männer	2 938	X	Männer	1 262	X
H Verkehr und Lagerei	2 116	X	R Kunst, Unterhaltung und Erholung	558	X
Frauen	527	25	Frauen	275	49
Männer	1 589	X	Männer	283	X
I Gastgewerbe	1 574	X	S Erbringung von sonstigen Dienstleistungen	1 180	X
Frauen	872	55	Frauen	795	67
Männer	702	X	Männer	385	X
J Information und Kommunikation	1 324	X			
Frauen	437	33			
Männer	887	X			

Abschnitte der Klassifikation der Wirtschaftszweige, Ausgabe 2008 (WZ 2008). Ergebnisse des Mikrozensus.

Anteil der Frauen beträgt mindestens 49 %.

Anteil der Frauen beträgt zwar mindestens 49 %, aber wenige Frauen sind in diesem Wirtschaftszweig tätig. Er ist daher nicht in der Auswertung enthalten.

Wirtschaftszweig O ist nicht in der Gender-Pay-Gap-Abgrenzung enthalten und daher auch nicht in der Auswertung.

Verdienstunterschiede zwischen Männern und Frauen

➤ **Tabelle 6** zeigt die Ergebnisse der Modellrechnung des höheren Mindestlohns von 12 beziehungsweise 15 Euro, wenn diese nun gezielt auf die ausgewählten Wirtschaftsbereiche angewendet wird.

Eine branchenspezifische Erhöhung des Mindestlohns auf 12 Euro reduziert den unbereinigten Gender Pay Gap in den Bereichen N (– 3,5 Prozentpunkte), I (– 3,0 Prozentpunkte) sowie G (– 2,0 Prozentpunkte) am stärksten. Auf die Wirtschaftsbereiche K (– 0,1 Prozentpunkte), Q (– 0,3 Prozentpunkte), M (– 0,5 Prozentpunkte) und S (– 0,8

Prozentpunkte) wirkt sich der erhöhte Mindestlohn kaum aus. Die Erhöhung des Mindestlohns auf 12 Euro führt im Bereich P sogar zu einer geringfügigen Verschlechterung des Gender Pay Gap (+ 0,1 Prozentpunkte).

Die Erhöhung des Mindestlohns für einzelne Branchen senkt zum Teil den unbereinigten Gender Pay Gap der jeweiligen Branche. Wendet man die Mindestlohn-erhöhung auf 12 Euro für alle ausgewählten Branchen (I, G, K, M, N, P, Q, S) insgesamt an, sinkt der gesamtwirtschaftliche Gender Pay Gap von 20,1 auf 19,0 %.

Tabelle 6

Auswirkungen auf den Gender Pay Gap bei einer Erhöhung des Mindestlohns

	Bruttostundenverdienst		Gender Pay Gap	Differenz zum Gender Pay Gap 2018
	Frauen	Männer		
	EUR		%	Prozentpunkte
Verdienststrukturerhebung 2018				
G Handel; Instandhaltung und Reparatur von Kraftfahrzeugen	14,78	19,02	22,3	X
I Gastgewerbe	10,98	11,98	8,3	X
K Erbringung von Finanz- und Versicherungsdienstleistungen	24,26	32,01	24,2	X
M Erbringung von freiberuflichen, wissenschaftlichen und technischen Dienstleistungen	19,92	28,23	29,4	X
N Erbringung von sonstigen wirtschaftlichen Dienstleistungen	12,64	14,69	14,0	X
P Erziehung und Unterricht	21,17	23,55	10,1	X
Q Gesundheits- und Sozialwesen	17,78	23,39	24,0	X
S Erbringung von sonstigen Dienstleistungen	17,06	20,41	16,4	X
B bis S ohne O	17,33	21,70	20,1	X
Modellrechnung höherer Mindestlohn 12 Euro				
G Handel; Instandhaltung und Reparatur von Kraftfahrzeugen	15,48	19,42	20,3	– 2,0
I Gastgewerbe	12,58	13,29	5,3	– 3,0
K Erbringung von Finanz- und Versicherungsdienstleistungen	24,32	32,05	24,1	– 0,1
M Erbringung von freiberuflichen, wissenschaftlichen und technischen Dienstleistungen	20,21	28,41	28,9	– 0,5
N Erbringung von sonstigen wirtschaftlichen Dienstleistungen	13,83	15,46	10,5	– 3,5
P Erziehung und Unterricht	21,28	23,70	10,2	+ 0,1
Q Gesundheits- und Sozialwesen	18,05	23,66	23,7	– 0,3
S Erbringung von sonstigen Dienstleistungen	17,63	20,90	15,6	– 0,8
B bis S ohne O	17,87	22,03	18,9	– 1,2
Modellrechnung höherer Mindestlohn 15 Euro				
G Handel; Instandhaltung und Reparatur von Kraftfahrzeugen	16,87	20,32	17,0	– 5,3
I Gastgewerbe	15,01	15,42	2,7	– 5,6
K Erbringung von Finanz- und Versicherungsdienstleistungen	24,48	32,15	23,9	– 0,3
M Erbringung von freiberuflichen, wissenschaftlichen und technischen Dienstleistungen	20,90	28,76	27,3	– 2,1
N Erbringung von sonstigen wirtschaftlichen Dienstleistungen	16,08	17,16	6,3	– 7,7
P Erziehung und Unterricht	21,60	24,07	10,3	+ 0,2
Q Gesundheits- und Sozialwesen	18,83	24,28	22,4	– 1,6
S Erbringung von sonstigen Dienstleistungen	18,63	21,78	14,5	– 1,9
B bis S ohne O	18,94	22,76	16,8	– 3,3

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

Wendet man den Mindestlohn von 15 Euro auf alle ausgewählten Branchen gleichzeitig an, senkt sich der gesamte unbereinigte Gender Pay Gap von 20,1 auf 16,8%. Um das Ziel der Bundesregierung (Gender Pay Gap=10%) zu erreichen, hätte der Mindestlohn 2018 in allen Branchen bei 23,20 Euro liegen müssen. Würde man sich auf die von Frauen stark besetzten Branchen (G, I, K, M, N, P, Q, S) konzentrieren, wäre rechnerisch ein Mindestlohn von 21,00 Euro erforderlich gewesen.

Die Simulationen zeigen, dass die Erhöhung des Mindestlohns auf 12 oder 15 Euro in von Frauen stark besetzten Branchen nur geringe Auswirkungen auf den unbereinigten Gender Pay Gap haben.

Dies lässt sich möglicherweise auch darauf zurückführen, dass sich der Gender Pay Gap weniger im unteren Lohnbereich, sondern vielmehr im oberen abspielt beziehungsweise sich durch alle Lohnstufen zieht. Wie die Auswertungen zeigen, lässt sich ein Großteil des unbereinigten Gender Pay Gap darauf zurückführen, dass die ausgeübten Tätigkeiten der Frauen nicht dem Führungs- und Qualifikationsanspruch der Männer entsprechen.

Da sich der Gender Pay Gap vermutlich durch alle Lohnstufen zieht, stellt sich die Frage, wie sich der unbereinigte Gender Pay Gap verändert, wenn man die Verdienste der Frauen um einen Fixbetrag erhöhen würde. In der Simulationsrechnung wurde mit diesem Grundgedanken der Bruttostundenverdienst der Frauen zum einen um 50 Cent und zum anderen um 1 Euro erhöht. Die Bruttostundenverdienste der Männer wurden nicht verändert.

Werden die Bruttostundenverdienste der Frauen um 50 Cent erhöht, senkt sich der Gender Pay Gap von 20,1 auf 17,8%. Erhöht man die Bruttostundenverdienste der Frauen um 1 Euro, so senkt sich der Gender Pay Gap auf 15,5%. ➔ Tabelle 7

Tabelle 7

Auswirkung auf den Gender Pay Gap bei einem Verdienstaufschlag für Frauen

	Bruttostundenverdienst in Euro		Gender Pay Gap	Differenz zum Gender Pay Gap 2018
	Frauen	Männer		
	EUR		%	Prozentpunkte
Verdienststrukturerhebung 2018	17,33	21,70	20,1	X
Prämie 0,50 Euro je Stunde	17,83	21,70	17,8	- 2,3
Prämie 1,00 Euro je Stunde	18,33	21,70	15,5	- 4,6
Prämie 2,20 Euro je Stunde	19,53	21,70	10,0	- 10,1

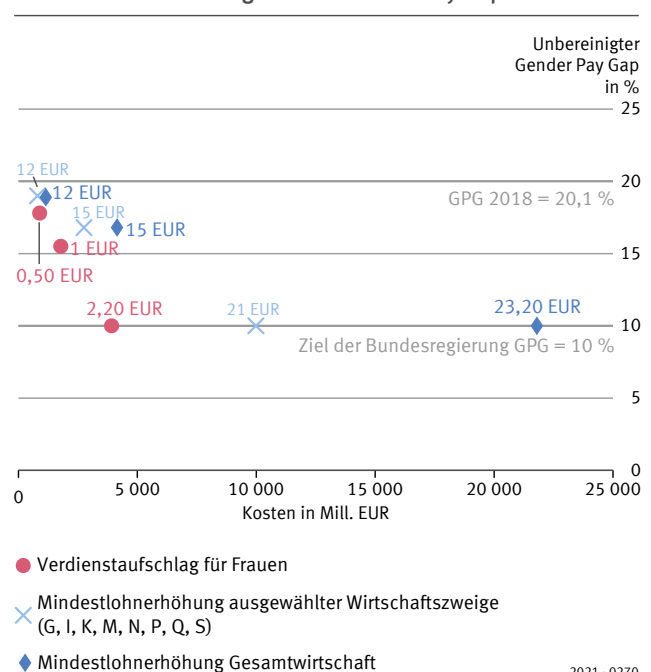
Prämienaufschlag für Frauen (geschlechterspezifisches Ungleichmoment in Erhöhung).

Eine flächendeckende Erhöhung der Bruttostundenverdienste der Frauen um 50 Cent erzielt damit eine stärkere Wirkung auf den Gender Pay Gap als die Erhöhung des gesetzlichen Mindestlohns von 8,84 Euro (2018; Gender Pay Gap = 20,1 %) auf 12 Euro für alle Branchen (Gender Pay Gap = 18,9 %). Die Erhöhung der Bruttostundenverdienste der Frauen um 1 Euro würde den unbereinigten Gender Pay Gap bereits um 4,6 Prozentpunkte senken. Um das Ziel der Bundesregierung zu erreichen, den Gender Pay Gap bis 2030 auf 10% zu reduzieren, müsste der Bruttostundenverdienst aller Frauen um 2,20 Euro erhöht werden.

➔ Grafik 2 stellt der erzielten Veränderung des Gender Pay Gap den damit verbundenen Mehraufwand der Wirt-


Grafik 2

Kosten und Auswirkung auf den Gender Pay Gap



schaft oder auch des Staates (wenn die Mehrkosten als Subventionen gewährt würden) gegenüber.

Es wird deutlich, dass das 10%-Ziel durch einen frauenspezifischen Verdienstaufschlag mit einem wesentlich geringeren Kostenumfang – unter den gegebenen vereinfachten Annahmen – erzielt werden würde. Gleichwohl ist die Maßnahme als die am wenigsten geschlechtergerechte einzustufen. Der theoretische Ansatz dieser Modellrechnung verfolgt nicht das Ziel, Frauen bei der Ausübung der gleichen Tätigkeit wie Männer höher zu entlohnen. Bei dieser Modellrechnung handelt es sich lediglich um eine Ex-post-Analyse, die die Wirkung der unterschiedlichen Maßnahmen verglichen hat.

weniger verdient als Männer. 2020 lag der unbereinigte Gender Pay Gap bei 18 % (Statistisches Bundesamt, 2021). Der Verdienstunterschied zwischen Männern und Frauen war damit um rund 2 Prozentpunkte geringer als 2018. 

6

Fazit

Auch für das Jahr 2018 konnten auf Basis der Verdienststrukturhebung neue Ergebnisse zum unbereinigten und bereinigten Gender Pay Gap ermittelt und eine detaillierte Ursachenanalyse durchgeführt werden.

Der unbereinigte Gender Pay Gap lag im Jahr 2018 nach der Abgrenzung ohne die Wirtschaftsabschnitte A und O und Kleinstbetriebe bei 20,1 %. Der bereinigte Gender Pay Gap lag für das gleiche Jahr bei 5,9 %.

Ergänzende Modellrechnungen (als Ceteris-paribus-Analyse) konnten simulieren, wie sich eine Erhöhung des Mindestlohns auf 12 und 15 Euro sowie ein Verdienstzuschlag für Frauen von 50 Cent oder 1 Euro auf den unbereinigten Gender Pay Gap auswirken würden.

Hierbei zeigt sich, dass ein Verdienstzuschlag von 50 Cent den unbereinigten Gender Pay Gap bereits stärker senkt als eine Erhöhung des Mindestlohns auf 12 Euro. Ein Verdienstzuschlag von 1 Euro führt zu einer beträchtlichen Senkung des unbereinigten Gender Pay Gap um 4,6 Prozentpunkte auf 15,5 %. Eine Erhöhung des Mindestlohns auf 15 Euro für alle Branchen reduziert den unbereinigten Gender Pay Gap jedoch nur um 3,3 Prozentpunkte.

In den Jahren ohne Verdienststrukturhebung kann das Statistische Bundesamt nur den unbereinigten Gender Pay Gap fortschätzen und veröffentlichen. Im Jahr 2019 haben Frauen in Deutschland durchschnittlich 19 %

LITERATURVERZEICHNIS

Achatz, Juliane/Gartner, Hermann/Glück, Timea. *Bonus oder Bias? Mechanismen geschlechtsspezifischer Entlohnung*. In: Kölner Zeitschrift für Soziologie und Sozialpsychologie. Jahrgang 57. Heft 3/2005, Seite 466 ff.

Beck, Martin. [Verdienstunterschiede zwischen Männern und Frauen nach Bundesländern](#). In: WISTA Wirtschaft und Statistik. Ausgabe 4/2018, Seite 26 ff.

Bundesregierung. *Ziele nachhaltiger Entwicklung. Gleichstellung von Frauen und Männern*. 2018. [Zugriff am 29. Juni 2021]. Verfügbar unter: www.bundesregierung.de

Blinder, Alan S. *Wage Discrimination: Reduced Form and Structural Estimates*. In: The Journal of Human Resources. Band 8. 1973. Ausgabe 4, Seite 436 ff.

Bundesamt für Bauwesen und Raumordnung. *Siedlungsstrukturelle Kreistypen*. [Zugriff am 25. März 2021]. Verfügbar unter: www.bbsr.bund.de

Eurostat (Herausgeber). *Task Force on Gender Pay Gap. Final Report*. Arbeitsunterlage. Luxemburg 2007.

Eurostat (Herausgeber). *Working Group Labour Market Statistics. State of the Play on the SES 2006 and Gender Pay Gap 2007*. Arbeitsunterlage. Luxemburg 2008.

Eurostat. *Auswertung: Geschlechtsspezifischer Lohnunterschied ohne Anpassungen*. [Zugriff am 8. Juli 2021]. Verfügbar unter: ec.europa.eu

Finke, Claudia. [Verdienstunterschiede zwischen Männern und Frauen](#). In: Wirtschaft und Statistik. Ausgabe 1/2011, Seite 36 ff.

Finke, Claudia/Dumpert, Florian/Beck, Martin. [Verdienstunterschiede zwischen Männern und Frauen](#). In: WISTA Wirtschaft und Statistik. Ausgabe 2/2017, Seite 43 ff.

Hinz, Thomas/Gartner, Hermann. *Geschlechtsspezifische Lohnunterschiede in Branchen, Berufen und Betrieben*. In: Zeitschrift für Soziologie. Jahrgang 34. Ausgabe 1/2005, Seite 22 ff.

Hübler, Olaf. *Geschlechtsspezifische Lohnunterschiede*. In: Mitteilungen aus der Arbeitsmarkt- und Berufsforschung. Jahrgang 36. Ausgabe 4/2003, Seite 539 ff.

International Labour Organization. *ISCO International Standard Classification of Occupations*. [Zugriff am 25. März 2021]. Verfügbar unter: www.ilo.org.

Neumark, David. *Sex Differences in Labor Markets*. London, New York 2004.

Oaxaca, Ronald. *Male-Female Wage Differentials in Urban Labour Markets*. In: International Economic Review. Band 14. Ausgabe 3/1973, Seite 693 ff.

Statistisches Bundesamt. *Verdienststrukturerhebung. Qualitätsbericht*. 2020a. [Zugriff am 25. März 2021]. Verfügbar unter: www.destatis.de

LITERATURVERZEICHNIS

Statistisches Bundesamt. *Auswertung: Erwerbstätige: Deutschland, Jahre, Wirtschaftszweige (WZ2008), Geschlecht*. 2018. [Zugriff am 25. März 2021]. Verfügbar unter: www-genesis.destatis.de

Statistisches Bundesamt. *Gender Pay Gap 2020: Frauen verdienen 18 % weniger als Männer*. Pressemitteilung Nr. 106 vom 9. März 2021. [Zugriff am 30. Juni 2021]. Verfügbar unter: 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.

Sabrina Estatico

ist staatlich geprüfte kaufmännische Assistentin für Datenverarbeitung und seit 2010 in der Gruppe „Zensus“ des Statistischen Bundesamtes tätig. Seit 2019 wirkt sie im Teilprojekt Personenerhebung (Konzeption und Aufbereitung) mit und ist unter anderem für die Plausibilisierung und Aufbereitung der Erhebungsdaten verantwortlich.

Tobias Tornow

hat Publizistik, Wirtschaftswissenschaften und Kommunikationswissenschaft studiert und war von 2017 bis 2020 im Statistischen Bundesamt für die methodische Vorbereitung und Koordination der Pilotstudie zur Haushaltebefragung verantwortlich. Seit 2020 leitet er als Referent im Referat „Pressestelle, Newsroom“ das Teilprojekt Öffentlichkeitsarbeit des Zensus 2022.

Sabrina Walther

hat Humangeographie sowie Umweltmanagement und Stadtplanung in Ballungsräumen studiert. Seit 2020 ist sie wissenschaftliche Mitarbeiterin im Referat „Haushaltsstichprobe (Konzeption und Aufbereitung)“ in der Gruppe „Zensus“ des Statistischen Bundesamtes. Sie beschäftigt sich bislang hauptsächlich mit den Auswertungen der Pilotstudie zur Haushaltebefragung.

GENERALPROBE FÜR DEN ZENSUS 2022: ERKENNTNISSE AUS DER PILOTSTUDIE ZUR HAUSHALTEBEFRAGUNG

Sabrina Estatico, Tobias Tornow, Sabrina Walther

➤ **Schlüsselwörter:** Zensus — Haushaltebefragung — Befragung — Pretest — Pilotstudie

ZUSAMMENFASSUNG

Im Zuge des Zensus 2022 werden grundlegende Daten über die Bevölkerung und die Wohnungssituation in Deutschland erhoben. Die Haushaltebefragung auf Stichprobenbasis ist Teil des Zensus 2022, sie wird mit vergleichbarer Methodik wie beim Zensus 2011 durchgeführt. Um das Erhebungsverfahren und die Erhebungsinstrumente zu erproben, wurde im Zeitraum von November 2019 bis März 2020 eine Pilotstudie unter möglichst realistischen Bedingungen durchgeführt. Der Aufsatz beschreibt die methodischen Hintergründe sowie den Aufbau und das methodische Vorgehen bei der Pilotstudie. Er legt ausgewählte Erkenntnisse dar und gibt Empfehlungen mit Blick auf die Haupterhebung.

➤ **Keywords:** census — household survey — survey — pretest — pilot study

ABSTRACT

Basic data on the population and housing situation in Germany will be collected as part of the 2022 Census. The sample-based household survey is part of the 2022 Census and will be conducted using comparable methodology to that of the 2011 Census. To test the survey procedure and survey instruments, a pilot study was conducted under the most realistic possible conditions in the period from November 2019 to March 2020. This paper describes the methodological background as well as the structure and methodological procedure of the pilot study. Selected findings are presented and recommendations are made concerning the main survey.

1

Einleitung

Elf Jahre nach dem ersten registergestützten Zensus ermitteln die Statistischen Ämter des Bundes und der Länder mit dem Zensus 2022 wieder, wie viele Menschen in Deutschland leben, wie sie wohnen und arbeiten.¹ Auch der Zensus 2022 nutzt Daten aus Melderegistern und kombiniert sie mit einer ergänzenden Personenerhebung sowie einer Gebäude- und Wohnungszählung. Die Personenerhebung besteht aus einer Haushaltebefragung auf Stichprobenbasis und einer Vollerhebung an Wohnheimen und Gemeinschaftsunterkünften. Weitere Informationen werden in einem maschinellen Verfahren, der sogenannten Haushaltegenerierung, aus den erhobenen Daten ermittelt.

Die Methodik des Zensus 2022 ist mit derjenigen zum Zensus 2011 vergleichbar. Anpassungen im gesamten Zensusverfahren und speziell in den Erhebungsinstrumenten machten gleichwohl eine erneute Erprobung der Feldphase unter möglichst realistischen Bedingungen erforderlich. Deswegen sollte eine Pilotstudie zur Haushaltebefragung potenzielle Problemfelder bezüglich der Haupterhebung und Optimierungspotenziale zugunsten der Erhebungsorganisation und der Datenqualität identifizieren. Die Erhebung fand im Zeitraum von November 2019 bis März 2020 statt.

Der Artikel beschreibt zunächst die Hintergründe zur Haushaltebefragung des Zensus 2022. Anschließend erläutert Kapitel 3 Ziele, Stichprobendesign und Methodik der Pilotstudie zur Haushaltebefragung. Nach einer Betrachtung ausgewählter Erkenntnisse in Kapitel 4 folgen ein Fazit zur Pilotstudie und Empfehlungen mit Blick auf die Haushaltebefragung 2022.

1 Ursprünglich war der Zensus für das Jahr 2021 geplant, wurde jedoch aufgrund der Einschränkungen durch die Corona-Pandemie in das Jahr 2022 verschoben. Alle gesetzlichen Grundlagen zum Zensus 2022 sind auf der Webseite www.zensus2022.de/gesetze zusammengestellt.

2

Hintergrund

In der Haushaltebefragung auf Stichprobenbasis des Zensus 2022 werden voraussichtlich etwa 10 Millionen Personen in Deutschland befragt. Im Kontext des Zensusmodells verfolgt die Haushaltebefragung zwei zentrale Ziele:

- › 1. Ziel: Die Haushaltebefragung ist angelegt als Korrekturstichprobe zur Feststellung und statistischen Korrektur von Über- und Untererfassung der in den Melderegistern erfassten Einwohnerinnen und Einwohner. In der sogenannten Ziel-1-Befragung werden alle 10 Millionen Personen, die an den ausgewählten Anschriften leben, befragt.
- › 2. Ziel: Mit der Haushaltebefragung werden Informationen erhoben, die nicht aus den bestehenden Registern gewonnen werden können, beispielsweise zum Bildungsstand, zur Erwerbstätigkeit sowie zum Beruf. In der sogenannten Ziel-2-Befragung werden voraussichtlich 6,7 Millionen Personen an den ausgewählten Anschriften zu diesen Merkmalen befragt (Bretsch/Lorentz, 2019).

Die Durchführung der Haushaltebefragung des Zensus 2022 obliegt den Statistischen Ämtern der Länder. Diese richten dazu vor Ort kommunale Erhebungsstellen ein, die den Einsatz von Interviewerinnen und Interviewern steuern. Diese Erhebungsbeauftragten überprüfen, welche Personen an den ausgewählten Anschriften wohnen (die sogenannte Existenzfeststellung) und befragen sie zu grundlegenden demografischen Angaben (Ziel 1). Weitergehende Fragen zu Bildungsstand, Erwerbstätigkeit sowie Beruf und Wirtschaftszweig sind in der Regel selbstständig per Online-Fragebogen zu beantworten (Ziel 2).

Die Haushaltebefragung 2022 orientiert sich dabei im Grundsatz an den im Zensus 2011 eingesetzten Methoden und Verfahren. Aufgrund wichtiger Erkenntnisse aus der letzten Zensusrunde, veränderter gesetzlicher Rahmenbedingungen sowie von Weiterentwicklungen im technischen und gesellschaftlichen Bereich wurden unter anderem folgende Aspekte der Haushaltebefragung neu justiert:

- › Der Einsatz der Korrekturstichprobe zur Ermittlung der Bevölkerungszahl wird auch in kleinen Gemeinden (Bretschi/Lorentz, 2019; Burgard und andere, 2019) erfolgen.
- › Eine Online-First-Strategie mit dem Online-Fragebogen als primärem Erhebungsweg wird umgesetzt (Freier/Mosel, 2019). Ziel der Strategie ist ein möglichst hoher Anteil an selbst ausgefüllten Online-Fragebogen bei der ergänzenden Ziel-2-Befragung. Dies soll Ressourcen sparen, zu einer höheren Datenqualität beitragen und eine schnellere Verarbeitung und Aufbereitung ermöglichen.
- › Die Evaluation des Zensus 2011 im Ganzen, aber auch der Haushaltebefragung und des Fragebogens im Speziellen gab zudem viele hilfreiche Hinweise auf methodische und organisatorische Verbesserungen der Verfahren und Erhebungsinstrumente zur Haushaltebefragung (Bretschi/Pfahl, 2018).

Mit dem Ziel, die Haushaltebefragung für alle Beteiligten so reibungslos wie möglich zu gestalten und die Datenqualität zu optimieren, wurden umfangreiche Pretests qualitativer und quantitativer Art durchgeführt. Unter einem Pretest versteht man generell „die Testung und Evaluation eines Fragebogens oder einzelner seiner Teile vor ihrem Einsatz in der Haupterhebung“ (Porst, 2014, hier: Seite 190), aber auch eine darüber hinaus gehende Erprobung der gesamten Erhebung. Der Einsatz von Pretests bei der Entwicklung von Fragebogen und Erhebungskonzepten ist gute wissenschaftliche Praxis und in der amtlichen Statistik etabliert (Blanke und andere, 2008).

Mehrere qualitative Pretests der verschiedenen Erhebungsinstrumente und Erhebungswege sollten zunächst „Mängel von Fragen und Fragebögen auf der Basis empirischer Informationen über den Befragten und sein Verhalten“ (Faulbaum und andere, 2009, hier: Seite 96) unter Laborbedingungen mit wenigen Probanden (10 bis 20 Personen) identifizieren. Mit einem anschließenden quantitativen Pretest – hier als Pilotstudie bezeichnet – wurde eine „vom Stichprobenumfang her stark verkleinerte Testerhebung eines Fragebogens am Ende der Fragebogenentwicklung unter möglichst realistischen Bedingungen der Haupterhebung“ (Faulbaum und andere, 2009, hier: Seite 98) sowie des gesamten Erhebungskonzepts durchgeführt. Dieses Verfahren wurde vergleichbar zur qualitativen und quanti-

tativen Erprobung der Haushaltebefragung des Zensus 2011 angewandt (Gauckler, 2011).

Die folgenden Kapitel stellen die Methodik sowie die Erkenntnisse aus der Pilotstudie zur Haushaltebefragung vor. Rechtsgrundlage der Erhebung ist § 6 Absatz 1 Bundesstatistikgesetz, wonach im Statistischen Verbund² zur Vorbereitung einer Bundesstatistik anordnenden Rechtsvorschrift „Fragebogen und Erhebungsverfahren auf ihre Zweckmäßigkeit“ erprobt werden können. Dabei bestand jedoch im Gegensatz zur Haupterhebung des Zensus 2022 keine Auskunftspflicht.

3

Methodischer Ansatz der Pilotstudie

3.1 Ziele der Pilotstudie

Die Pilotstudie sollte die Erhebungsorganisation, die Erhebungsinstrumente sowie die nachfolgende Aufbereitung erproben. Ihre Ziele waren im Einzelnen:

- › Erkenntnisse zur Nutzung der verschiedenen Erhebungswege und insbesondere zur Nutzung des Online-Fragebogens gewinnen.
- › Den Erhebungsablauf und die Erhebungsorganisation zur Optimierung der Prozesse für die Haupterhebung erproben.
- › Die Verständlichkeit und Handhabbarkeit der Erhebungsinstrumente (Erhebungsunterlagen des Feldpersonals und Fragebogen) testen und damit Messfehler verringern, Validität (Gültigkeit) und Reliabilität (Zuverlässigkeit und Genauigkeit) der erhobenen Daten steigern sowie bei den Befragten die Belastung reduzieren und eine größere Flexibilität erreichen.
- › Die Konzepte zur Aufbereitung und Plausibilisierung der erhobenen Daten überprüfen und modifizieren.

² Den Statistischen Verbund bilden die Statistischen Ämter des Bundes und der Länder.

3.2 Stichprobendesign

Um die zuvor genannten Ziele zu erreichen, haben die zwölf an der Pilotstudie teilnehmenden Statistischen Ämter der Länder von November 2019 bis März 2020 insgesamt 8 632 Personen befragt.¹³ Auswahlseinheiten waren analog zur Haupterhebung in einem mehrstufigen Auswahlverfahren ausgewählte Anschriften mit Wohnraum. Je Bundesland wurden städtische und ländliche Gemeinden vorgegeben, um die Organisation der Pilotstudie zu vereinfachen, und in diesen dann anteilig zur Bevölkerungszahl Anschriften mit Wohnraum und mindestens einer gemeldeten Person gezogen. Auswahlgrundlage war dabei das Steuerungsregister zur Vorbereitung und Durchführung des Zensus (§ 3 Zensusvorbereitungsgesetz 2022).

Dabei sollten 5 % der ausgewählten Anschriften sogenannte Großanschriften mit 50 oder mehr gemeldeten Personen abbilden, um Besonderheiten an diesen Anschriften testen zu können. Unabhängig von der Größe der Gemeinde wurden 10 % der ausgewählten Anschriften als sogenannte Ziel-1-Anschriften behandelt, an denen nur die kurze Ziel-1-Befragung durchgeführt werden sollte. Die Stichprobenziehung erfolgte im September 2019 im Statistischen Bundesamt.

3.3 Methodik und Durchführung der Pilotstudie

Im Zuge der Pilotstudie wurden verschiedene Methoden der Datenerhebung genutzt:

Quantitative Datenerhebung, Prozessdaten und Nachbefragung

Zentraler Untersuchungsstrang der Pilotstudie war die quantitative Datenerhebung, mit welcher vorrangig die Praxistauglichkeit der Erhebungsinstrumente sowie Aspekte der Datenqualität evaluiert werden sollten. Das Vorgehen entsprach dem der Haupterhebung: Die Erhebungsbeauftragten erfassten zunächst erhebungsorganisatorische Angaben zur Stichprobenanschrift in einem Anschriften-Mantelbogen (zum Beispiel Anmer-

kungen zum Auffinden der Anschrift und zur Anzahl der Haushalte). Anschließend überprüften sie in einer persönlichen Befragung, welche Personen in den Haushalten wohnten (Existenzfeststellung), und erfassten grundlegende demografische Daten zu Haushalt und Haushaltsmitgliedern in einem Haushaltsbogen (Ziel 1). Sofern der Haushalt für die ergänzende Ziel-2-Befragung vorgesehen war¹⁴ wurden im Anschluss entweder die Zusatzmerkmale per Selbstauskunft mittels Online-Befragung¹⁵ erhoben, eine persönliche Befragung oder auf Wunsch eine computergestützte Telefonbefragung durchgeführt.¹⁶ Die Systematik der Erhebungsinstrumente zeigt [Übersicht 1](#).

Übersicht 1

Erhebungsinstrumente der Haushaltebefragung

Erhebungsinstrumente für Erhebungsbeauftragte		
Anschriftenebene	Mantelbogen	
Haushaltsebene	Haushaltsbogen	
Erhebungsinstrumente für befragte Personen (Fragebogen)		
Personenebene	Ziel-1-Fragebogen	Ziel-2-Fragebogen (einschließlich Ziel-1-Fragen)

Auf Basis dieses Untersuchungsstrangs erfolgte unter anderem:

- › die **deskriptive Datenauswertung**, welche grundlegende Verteilungen der Merkmalsausprägungen untersuchte und potenzielle Fehlerquellen identifizieren sollte,
- › das Erarbeiten eines **Aufbereitungskonzepts zur Plausibilisierung**, welches Aufbereitungsschritte vorläufig erproben, mögliche Fehlerquellen frühzeitig aufdecken und wichtige Hinweise zur Datenqualität auch für die Haupterhebung liefern sollte,
- › die **Erhebung und Auswertung von Prozessdaten** (Kreuter/Casas-Cordero, 2010), welche die Quantifizierung von Schwierigkeiten bei den elektronisch unterstützten Befragungen und die Optimierung des Online-Fragebogens zum Ziel hatte, sowie

3 An der Pilotstudie haben die Statistischen Ämter der Länder Baden-Württemberg, Bayern, Berlin und Brandenburg, Bremen, Hamburg und Schleswig-Holstein (Nord), Niedersachsen, Nordrhein-Westfalen, Rheinland-Pfalz, Saarland, Sachsen, Sachsen-Anhalt und Thüringen teilgenommen.

4 Bei der Haushaltebefragung des Zensus 2022 werden nicht alle Haushalte auch zu den weiteren Angaben befragt, daher erfolgte diese Unterscheidung auch in der Pilotstudie. Aus erhebungsorganisatorischen Gründen ermittelt diese ergänzende Befragung auch noch einmal die demografischen Angaben aus der Ziel-1-Befragung.
 5 Der Online-Fragebogen zur Pilotstudie ist unter www.zensus2022.de verfügbar.
 6 In Bayern gab es zudem computergestützte persönliche Befragungen.

- › die **optionalen Nachbefragungen innerhalb der elektronischen Erhebungswege**. Sie sollten ergänzend zu den Erkenntnissen aus den qualitativen Pretests zum Beispiel die Handhabbarkeit und Bedienung des Online-Fragebogens sowie den Weg vom Anschreiben mit den Zugangsdaten zum Online-Fragebogen untersuchen. Falls eine computergestützte Telefonbefragung erfolgt war, sollten die Gründe für den Verzicht auf die Online-Teilnahme erfragt werden.

Debriefings der Erhebungsbeauftragten

Um Erkenntnisse und Perspektiven zur Durchführung und den Erhebungsinstrumenten aus Sicht des Feldpersonals zu erhalten, sah das Konzept strukturierte, leitfadengestützte Gruppendiskussionen mit den eingesetzten Erhebungsbeauftragten vor. Aufgrund der Einschränkungen durch die Corona-Pandemie im Durchführungszeitraum (vorgesehen war März 2020) war dies nur bedingt möglich.

Debriefing der Statistischen Ämter der Länder

Vergleichbar mit dem Austausch mit den Erhebungsbeauftragten war auch eine Debriefing-Runde mit den Statistischen Ämtern der Länder vorgesehen. Die Inhalte dieser Gruppendiskussion lehnten sich an die Evaluationsgespräche mit dem Feldpersonal an, allerdings sollten auch weiterreichende Aspekte evaluiert werden, wie technische und organisatorische Rahmenbedingungen im Zusammenhang mit der Durchführung der Pilotstudie. Die Erfahrungen der statistischen Ämter wurden mithilfe eines systematischen Evaluationsbogens gesammelt. Auch dieses Format konnte aufgrund der Corona-Pandemie nicht planmäßig als Präsenzveranstaltung durchgeführt werden.

4

Ausgewählte Erkenntnisse der Pilotstudie

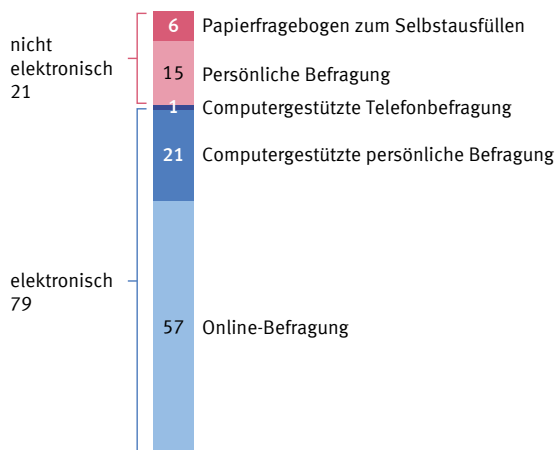
4.1 Rücklauf und Erhebungsweg

Bundesweit wurden in der Pilotstudie 8 632 Personen befragt, darunter 6 998 Personen, die auch Auskunft zu erweiterten Zensusmerkmalen, zum Beispiel zu Bildung und Erwerbstätigkeit, gaben (Ziel 2).

Rund 79 % der zusätzlichen Ziel-2-Befragungen erfolgten auf elektronischem Weg (57 % als Online-Befragung, 21 % als computergestützte persönliche Befragung und 1 % als computergestützte Telefonbefragung). Etwa 21 % der teilnehmenden Personen beantworteten die Fragen mittels eines Papierfragebogens: 6 % eigenständig, 15 % in Form einer persönlichen Befragung durch Erhebungsbeauftragte. [➤ Grafik 1](#)

Grafik 1

Rückläufe der Fragebogen der Ziel-2-Stichprobe differenziert nach Erhebungsweg
in %; Basisbezug n = 6 998



2021 - 0271

4.2 Erhebungsunterlagen des Feldpersonals

Die Evaluation der Erhebungsunterlagen des Feldpersonals (Mantel- und Haushaltsbogen) umfasste unterschiedliche Themenfelder. Grundlegende Optimierungspotenziale für die Haupterhebung zeigten sich bei folgenden Kriterien:

Identifikation von Befragungstypen

Die Zuweisung der zutreffenden Befragungstypen (Ziel 1 oder Ziel 2) gestaltete sich in der Testphase teilweise schwierig. Erhebungsbeauftragte meldeten zurück, dass aufgrund fehlender (optischer) Kennzeichnungen und unklarer Feldbezeichnungen in den Erhebungsunterlagen Unsicherheiten bezüglich des Zuordnungsprozesses bestanden. Daher war es nicht immer möglich sicherzustellen, dass jede Befragung unter den Vorgaben der jeweiligen Stichprobenzugehörigkeit erfolgte.

Frageelemente und Formulierungshinweise

Die Evaluierung des Mantel- und Haushaltsbogens zeigte, dass neben der Überarbeitung einzelner Gestaltungskomponenten (zum Beispiel Layout, Schriftart) auch eine Modifizierung von Frageelementen (zum Beispiel Antwortoptionen, missverständliche Filterführungen) notwendig war. Hier war beispielsweise festzustellen, dass verschiedene Ankreuzmuster komplexer Kombinationen nicht vollumfänglich verstanden wurden und somit die Befragungsergebnisse nicht in jedem Fall plausibel waren. Während der Anteil solcher Auffälligkeiten in der Regel unter 3 % lag, betrug der Maximalwert an formal inkorrekten Angaben 14 % aller Antworten.

Unsicherheiten im Erhebungsablauf

Die Debriefings der Erhebungsbeauftragten zeigten, dass unerwartete Situationen (zum Beispiel Abweichungen zu den Angaben der erwarteten Wohnanschrift) Unklarheiten im Befragungsablauf zur Folge haben konnten. In solchen Fällen war nicht immer ersichtlich, wie weiter zu verfahren war und wie derartige Vorkommnisse auf den entsprechenden Dokumenten zu hinterlegen waren.

Handhabbarkeit der Unterlagen

Die Handhabbarkeit der Erhebungsinstrumente stellte einige Erhebungsbeauftragte vor Herausforderungen. Dies betraf insbesondere den händischen Aufwand und die Vielzahl an papierbasierten Erhebungsunterlagen.

4.3 Fragebogen

Der Fragebogen der Pilotstudie stand sowohl in Papierform als auch elektronisch als Online-Fragebogen zur Verfügung. Beide Fragebogenvarianten unterschieden sich nicht bei Formulierungen von Fragetexten, Antwortoptionen und Hinweistexten, sondern lediglich in ihrer technischen und gestalterischen Umsetzung. Die Anforderungen an eine nutzungsfreundliche, leicht verständliche und zugängliche Gestaltung der Fragebogen galten für beide Typen gleichermaßen und wurden in der Pilotstudie umfassend untersucht:

Frageelemente und Gestaltungshinweise

Ein Ergebnis der Pilotstudie war, dass neben der Überarbeitung von einzelnen Gestaltungskomponenten wie dem Layout auch eine Modifizierung von Frageelementen (zum Beispiel Frageformulierungen, Hinweistexte, Antwortoptionen, Ergänzung fehlender Stichtagsbezüge) notwendig war. In den gedruckten Unterlagen zogen beispielsweise unauffällige Hinweistexte (zum Beispiel „Für Personen unter 15 Jahren endet die Befragung hier“) unplausible Folgeangaben nach sich. Auch war die technische Umsetzung des Online-Fragebogens teilweise verbesserungswürdig. So wurde die unterstützte Abfrage des Berufs (zum Beispiel durch eine Suchfunktion in einer hinterlegten Datenbank) nicht von allen teilnehmenden Personen wahrgenommen.

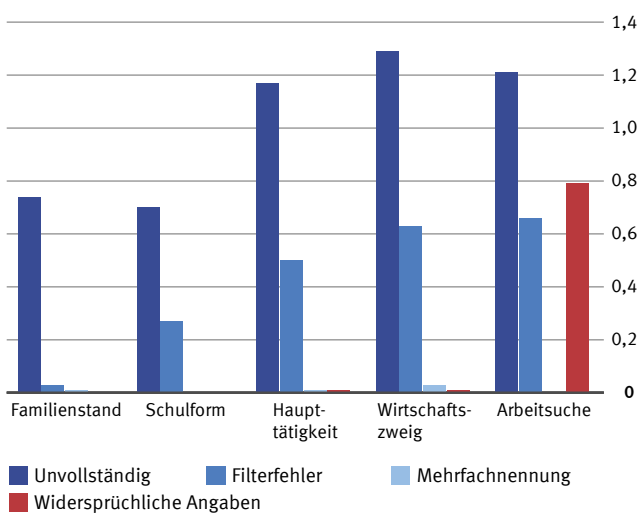
Plausibilisierung der erhobenen Daten

Eine wesentliche Rolle bei der Datenaufbereitung nahm die Ausarbeitung des Plausibilisierungskonzepts ein, das in der Pilotstudie getestet und für die Haupterhebung angepasst werden sollte. Da die Anforderungen im Zuge der Plausibilitätsprüfung bereits im Online-Fragebogen technisch eingebunden werden konnten, wurden unplausible Angaben (zum Beispiel unzulässig fehlende Angaben, Filterfehler, unzulässige Mehrfachnennungen, widersprüchliche Antworten zu zuvor getätigten Angaben) nur bei papiergebundenen Meldewegen erwartet.

Es zeigte sich, dass insgesamt wenig unplausible Angaben getätigt wurden und diese vor allem bei den papiergebundenen Erhebungswegen auftraten. Neben fehlenden Antworten waren es insbesondere Angaben mit fehlerhafter Filterführung, die von den definierten Vorgaben abwichen. Die identifizierten Unplausibilitäten am Beispiel ausgewählter Merkmale zeigt [Grafik 2](#).

Grafik 2

Exemplarische Darstellung von Unplausibilitäten in %; Basisbezug n = 6 998



2021 - 0272

Einflussfaktor Feldpersonal

Unplausible Angaben, zum Beispiel fehlerhafte Filterführungen, traten nicht ausschließlich bei den von den Befragten selbst ausgefüllten Papierfragebogen auf, sondern resultierten auch aus der persönlichen Befragung durch das Feldpersonal.

Online-Fragebogen aus Sicht der Befragten

Nach Abschluss der eigentlichen Online-Befragung konnten die Befragten eine strukturierte Rückmeldung zur Handhabbarkeit des Online-Fragebogens geben. Die wichtigsten Erkenntnisse aus diesen Rückmeldungen sind:

- › Die Handhabbarkeit des Online-Fragebogens bewerteten 96 % der Befragten sehr gut bis eher gut, weniger als 1 % als eher schlecht bis sehr schlecht.
- › Fast drei Viertel (74 %) der befragten Personen war klar, welche Angaben sie bei der Anmeldung zum Online-Fragebogen eingeben sollten.

- › Weniger als 3 % bestätigten, grundlegende Probleme bei der Anmeldung zur Online-Befragung gehabt zu haben. Als Gründe hierfür wurden insbesondere Schwierigkeiten bei der Lesbarkeit der Zugangsdaten im Anschreiben oder Serverprobleme angeführt.
- › Das Auffinden der Website zur Online-Befragung ist 68 % der Befragten sofort gelungen, 32 % waren nicht direkt erfolgreich.

Prozessdaten – hinter den Kulissen des Online-Fragebogens

Neben der Evaluierung der Erhebungsunterlagen und der Erhebungsorganisation sowie -durchführung nahm auch die Auswertung von Prozessdaten einen zentralen Stellenwert zur Identifikation von Optimierungspotenzialen ein. So war es hier unter anderem möglich, technische Informationen zu den Rahmenbedingungen der Online-Befragung (zum Beispiel Angaben zum genutzten Gerätetyp) und zum Ausfüllverhalten der befragten Person (zum Beispiel Dauer der Befragung, Identifikation von Befragungsabbrüchen) zu gewinnen.

Die Analysen ergaben unter anderem, dass

- › die durchschnittliche Ausfülldauer des Online-Fragebogens bei der reinen Ziel-1-Befragung unter fünf Minuten und bei der Ziel-2-Befragung etwas über zehn Minuten lag.
- › 75 % der Befragten den Online-Fragebogen am PC ausgefüllt haben, 17 % nutzten ein Smartphone und 8 % ein Tablet.
- › insgesamt 193 Abbrüche der Online-Befragung verzeichnet wurden. Hierzu zählen jene Fälle, bei denen sich Personen im Online-Fragebogen angemeldet, diesen jedoch vorzeitig beendet haben. Zu anderen Formen von Ausfällen, beispielsweise durch Nichtauffinden der Website, können hingegen keine Aussagen getroffen werden.

4.4 Erhebungsorganisation und -durchführung

Online-First-Strategie

In der Pilotstudie wurden vier zentrale Felder für das Gelingen der Online-First-Strategie identifiziert: die persönlichen und technischen Rahmenbedingungen der teilnehmenden Personen, die individuellen Verhaltensweisen der Erhebungsbeauftragten, länderspezifische Maßnahmen zur Umsetzung der Online-First-Strategie und die bereitgestellten Informationen hinsichtlich des Anmeldeprozesses im Onlinesystem. [➤ Grafik 3](#) Diese Einflussfaktoren können je nach Zusammenspiel eine Online-Teilnahme begünstigen oder sogar ausschließen.

Die Analyse des Themenbereichs „Rahmenbedingungen der Teilnehmenden“ ergab, dass grundsätzlich eine hohe Bereitschaft zur Online-Teilnahme bestand. Neben der diskreten Durchführung der Befragung war die Zeitersparnis einer der Hauptgründe für das Interesse an der Online-Befragung. Fehlende Internetzugänge, Barrieren für nicht online-affine Menschen und der Wunsch zur vollständigen Erhebung vor Ort (wo in der Regel mit der persönlichen Ziel-1-Befragung begonnen wurde) sind exemplarische Gründe für eine ablehnende Haltung gegenüber der Online-Befragung.

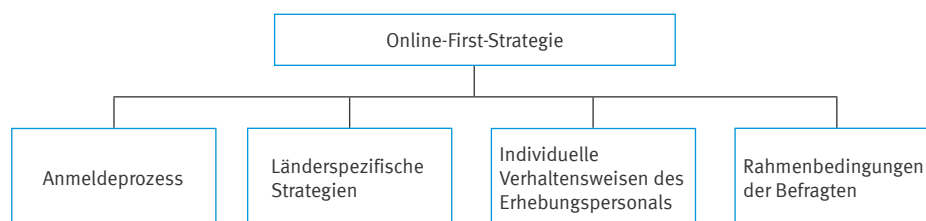
Dass die Dateneingänge allerdings über den Online-Fragebogen registriert beziehungsweise die entsprechenden Zugangsdaten von den teilnehmenden Personen entgegengenommen wurden, kann nicht ausschließlich auf Basis der Bereitwilligkeit der Befragten erklärt werden. Vielmehr wirkten sich auch der Anmeldeprozess, länderspezifische Strategien zur Priorisierung des Online-Erhebungswegs und individuelle Verhaltensweisen des Feldpersonals auf das Teilnahmeverhalten

aus. So wurde im Zuge der Debriefings beispielsweise berichtet, dass

- › die Entscheidung des Erhebungswegs häufig beim Erhebungspersonal lag. Je nach individueller Einschätzung, beispielsweise über demografische Faktoren wie das Alter, wurden Zugangsdaten zum Online-Fragebogen oder ausschließlich Papierfragebogen angeboten. In solchen Fällen wurde nicht immer ermittelt, ob eine grundlegende Bereitschaft zur Online-Teilnahme seitens der Befragten bestand.
- › einige Erhebungsbeauftragte je nach Empfinden alle oder den ausschließlich persönlich präferierten Erhebungsweg kommunizierten.
- › bei einigen Erhebungsbeauftragten Unsicherheit bezüglich des Online-Fragebogens herrschte, etwa bei Ausgabe und Zuweisung von Zugangsdaten. Sie boten in solchen Fällen präferiert papiergebundene Befragungswege an.
- › sich die Anmeldung zum Online-Fragebogen teilweise als kompliziert herausstellte. Ungünstige Druckeinstellungen beispielsweise führten in manchen Fällen dazu, dass Unterlängen von Zeichen in Bezug auf die Zugangsdaten abgeschnitten wurden. Das hatte die fehlende Differenzierbarkeit einzelner Zeichen (zum Beispiel i und j) zur Folge. Im Zusammenspiel mit ungünstigen Schriftbildern und -größen (zum Beispiel Probleme bei der Unterscheidung von 9 und g, c und e) waren mehrere Anmeldeversuche notwendig.
- › sich das Auffinden des Online-Fragebogens teilweise als schwierig erwies. So konnte der Online-Fragebogen trotz entsprechender Hinweise nicht immer aufgerufen werden. Auch weiterreichende Informationen über die Zugangsmöglichkeiten auf den Seiten der Statistischen Ämter der Länder waren in manchen Fällen nicht ersichtlich. Dies hatte zur Folge, dass

Grafik 3

Einflussfaktoren der Online-First-Strategie



2021 - 0273

einige Personen trotz einer Präferenz für die Online-Teilnahme schließlich einen anderen Erhebungsweg wählten oder eine Teilnahmeverweigerung in Betracht zogen.

- › die länderspezifischen Strategien maßgeblich Einfluss auf das Gelingen der Online-First-Strategie nahmen (zum Beispiel durch intensive Schulung des Erhebungspersonals oder den Verzicht auf das Mitführen von Papierfragebogen durch das Erhebungspersonal).

Diese Analysen zeigen, dass die Entscheidung für oder gegen eine Online-Teilnahme nicht allein den Umständen der befragten Person zuzuschreiben ist. Vielmehr resultierte sie aus einer Kombination aus erhebungsorganisatorischen Strategien, den bereitgestellten Informationen zum Anmeldeprozess und dem Verhalten des Feldpersonals. Diese Erkenntnisse erweisen sich für die Haupterhebung als ausgesprochen hilfreich, da insbesondere individuelle Befindlichkeiten des Erhebungspersonals keinen Einfluss auf den Erhebungsablauf haben sollten und die Online-First-Strategie an vorderster Stelle umgesetzt werden muss.

Herausforderungen bei der Vorbereitung und Durchführung der Pilotstudie

Im Rahmen der Debriefings galt es auch, Herausforderungen bei der Vorbereitung und Durchführung der Feldphase zu ermitteln. Neben den bereits genannten Punkten zeigte sich unter anderem, dass

- › sich die Rekrutierung des Erhebungspersonals teilweise schwierig gestaltete. So war es nicht immer möglich, die benötigte Anzahl an Erhebungsbeauftragten anzuwerben beziehungsweise auch einzusetzen.
- › krankheitsbedingte Ausfälle von Erhebungsbeauftragten aufgrund geringer Personalkapazitäten und eventuell verzögerter Bekanntgabe von Krankmeldungen kurzfristig nicht immer kompensiert werden konnten.
- › sich der Umgang mit den zu befragenden Personen nicht immer als einfach erwies. Sprachliche Barrieren und vermehrtes Misstrauen gegenüber dem Zensus beziehungsweise der Pilotstudie führten zu teilweise umfangreicher Aufklärungsarbeit und dazu, dass die Befragung nicht in jedem Fall wie geplant durchgeführt werden konnte.

- › der regelmäßige Austausch zwischen dem Erhebungspersonal und den Statistischen Ämtern der Länder als essenzielles Element der Feldphase angesehen werden kann. So war es möglich, zum Beispiel bei Unstimmigkeiten bei der Erstbegehung der Anschriften frühzeitig Probleme zu erkennen. Auch konnten Arbeitsstandkontrollen mittels fortwährender Kommunikation erfolgen.

5

Fazit und Empfehlungen für die Haushaltebefragung des Zensus 2022

Die Auswertungsergebnisse haben gezeigt, dass sich sowohl die eingesetzten Erhebungsinstrumente als auch die erhebungsspezifischen Abläufe im Feld gut bewährt haben. Die kleinteilige und methodisch vielfältige Analyse der Ergebnisse konnte allerdings Optimierungsbedarfe identifizieren, die die Weiterentwicklung des gesamten Erhebungskonzepts unterstützen und zu den nachfolgenden Empfehlungen führen:

Anpassung der Erhebungsunterlagen des Erhebungspersonals

Die Erhebungsunterlagen werden so angepasst, dass die Verständlichkeit aller Elemente gesteigert (zum Beispiel durch Anpassung von Begrifflichkeiten, Frageformulierungen, Regieanweisungen, Antwortmustern und prägnanterer Kennzeichnung des Befragungstyps) und die Handhabbarkeit verbessert wird (zum Beispiel durch die Vorbelegung statischer Angaben wie der Jahreszahl des Befragungstermins). Dies soll unter anderem dazu beitragen, dass sich das Erhebungspersonal im Umgang mit den Unterlagen sicherer fühlt. Neben der Modifizierung der Erhebungsunterlagen fließen die Erkenntnisse aus der Pilotstudie auch in die Konzeption der Schulungen des Feldpersonals ein.

Optimierungspotenziale der Fragebogen

Auch bei den Online- und Papierfragebogen ist eine Überprüfung der Umgestaltungsmöglichkeiten von Frage-, Hinweis- und Informationstexten, Filterelementen, Antwortoptionen und der Befragungsstruktur vorgesehen.

Um die Vorteile des Online-Fragebogens gut ausschöpfen zu können, werden die Oberfläche und die eingebundenen Unterstützungssysteme (zum Beispiel bei der Eingabe der Berufsbezeichnung) überarbeitet. Aufgrund entdeckter unplausibler Angaben bei persönlichen Befragungen wird zudem ein stärkerer Fokus auf den Umgang mit den Erhebungsunterlagen in den Schulungen des Feldpersonals gelegt.

Priorisierung der Online-First-Strategie

Die Pilotstudie konnte vier zentrale Elemente für das Gelingen der Online-First-Strategie identifizieren. Damit die Priorisierung dieses Leitgedankens auch in der Haupterhebung adäquat erfolgen kann, ist ein entsprechender Schwerpunkt bei der Schulung des Erhebungspersonals notwendig. Neben geeigneten Argumentationshilfen können praxisorientierte Vorgehensweisen dabei helfen, Unsicherheiten auszuräumen und die Einflussnahme individueller Befindlichkeiten auf den Erhebungsweg zu reduzieren.

Neben der Schulung des Erhebungspersonals wird auch der Anmeldeprozess angepasst: Die notwendigen Informationen werden angemessen aufbereitet (Beispiel: lesbares Schriftbild der Zugangsdaten) und können an den entsprechenden Stellen auch gefunden werden (Beispiel: vereinfachtes Auffinden des Online-Fragebogens auch auf den Webseiten der statistischen Ämter).

Öffentlichkeitsarbeit


Zur Haupterhebung wird – mit zeitlichem Vorlauf zur Feldphase – umfangreich über den anstehenden Zensus informiert werden. Die Öffentlichkeitsarbeit soll Unsicherheiten bei den befragten Personen schon im Voraus ausräumen, ein breites Verständnis zu Sinn und Zweck des Zensus erreichen, dadurch die Teilnahmebereitschaft und Motivation erhöhen und dazu beitragen, ausreichend Erhebungsbeauftragte anzuwerben.

Schulungen des Erhebungspersonals

Die Analysen der Pilotstudie zeigen zudem die Bedeutung von gut geschultem Feldpersonal für die Erhebung qualitativ hochwertiger Daten auf. Damit auch in der Haupterhebung solche Schulungen zielgerichtet erfolgen können, ist eine Optimierung der Schulungsinhalte vorgesehen. Neben den konkreten Erhebungsabläufen können dabei auch vertiefend Verhaltensstrategien ver-

mittelt werden, die neben der Priorisierung der Online-First-Strategie auch auf unerwartete Situationen, zum Beispiel mögliche Sprachbarrieren, vorbereiten.

Zensus in Zeiten der Corona-Pandemie

Die formulierten Empfehlungen beziehen sich auf den Zeitpunkt der Durchführung der Pilotstudie. Aufgrund der andauernden Corona-Pandemie müssen die Erhebungsverfahren unter Umständen angepasst und alternative Vorgehensweisen, wie etwa Maßnahmen zur Kontaktreduzierung sowie ein Hygienekonzept, erarbeitet werden. 

LITERATURVERZEICHNIS

Blanke, Karen/Gauckler, Britta/Sattelberger, Sabine. [*Fragebogen auf dem Prüfstand: Testmethoden und deren Einsatz in der amtlichen Statistik*](#). In: Wirtschaft und Statistik. Ausgabe 8/2008, Seite 641 ff.

Bretsch, Corinna/Lorentz, Kai. [*Präzisionsziele für die Ermittlung der Einwohnerzahl bei der Haushaltsstichprobe im Zensus 2021*](#). In: WISTA Wirtschaft und Statistik. Sonderheft Zensus 2021. Wiesbaden 2019, Seite 12 ff.

Bretsch, David/Pfahl, Miriam. [*Beim Fragebogen-„TÜV“ – Evaluierung des Fragebogens der Haushaltsstichprobe des Zensus 2011*](#). In: WISTA Wirtschaft und Statistik. Ausgabe 2/2018, Seite 91 ff.

Burgard, Jan Pablo/Münnich, Ralf/Rupp, Martin. [*Die Entwicklung des Stichprobenkonzepts für den Zensus 2021*](#). In: WISTA Wirtschaft und Statistik. Sonderheft Zensus 2021. Wiesbaden 2019, Seite 23 ff.

Faulbaum, Frank/Prüfer, Peter/Rexroth, Margrit. *Was ist eine gute Frage? Die systematische Evaluation der Fragenqualität*. Wiesbaden 2009.

Freier, Benjamin/Mosel, Juliane. [*Online First als Leitgedanke für effiziente Primärerhebungen beim Zensus 2021*](#). In: WISTA Wirtschaft und Statistik. Sonderheft Zensus 2021. Wiesbaden 2019, Seite 46 ff.

Gauckler, Britta. [*Die Entwicklung des Fragebogens zur Haushaltebefragung des Zensus 2011*](#). In: Wirtschaft und Statistik. Ausgabe 8/2011, Seite 718 ff.

Kreuter, Frauke/Casas-Cordero, Carolina. *Paradata. RatSWD Working Paper, No. 136*. Berlin 2010.

Porst, Rolf. *Fragebogen*. Wiesbaden 2014.

Herausgeber

Statistisches Bundesamt (Destatis), Wiesbaden

Schriftleitung

Dr. Daniel Vorgrimler

Redaktion: Ellen Römer

Ihr Kontakt zu uns

www.destatis.de/kontakt

Erscheinungsfolge

zweimonatlich, erschienen im August 2021

Ältere Ausgaben finden Sie unter www.destatis.de sowie in der [Statistischen Bibliothek](#).

Artikelnummer: 1010200-21004-4, ISSN 1619-2907

© Statistisches Bundesamt (Destatis), 2021

Vervielfältigung und Verbreitung, auch auszugsweise, mit Quellenangabe gestattet.