

Claudia Finke

hat Sozialwissenschaften mit dem Schwerpunkt angewandte Sozialforschung an der Ruhr-Universität Bochum studiert und ist im Statistischen Bundesamt als Referentin im Referat „Verdienste, Überschuldung“ tätig. Schwerpunktmäßig betreut sie als Product Owner die Entwicklung einer Webanwendung zur Aufbereitung der Daten der Verdiensterhebung.

Susanna Geisler

ist M.Sc. Economics und hat an den Universitäten Köln, Bayreuth und Santiago de Chile studiert. Im Statistischen Bundesamt ist sie seit 2017 und derzeit als Referentin im Referat „Verdienste, Überschuldung“ tätig. Ihre Aufgabenschwerpunkte liegen in der methodischen Weiterentwicklung sowie Erstellung und Verbreitung von Ergebnissen der Verdienst- und Überschuldungsstatistik.

Dr. Anja Überschaer

ist Referentin im Referat „Verdienste, Überschuldung“ des Statistischen Bundesamtes. Sie hat in Gießen und Milwaukee im Rahmen eines Doppelmasterprogramms Betriebswirtschaftslehre und Volkswirtschaftslehre studiert. Ihr Tätigkeitsbereich umfasst methodische Fragestellungen rund um die Verdienststatistik und sie hat die interaktive Anwendung „Gehaltsvergleich“ entwickelt.

AUS DREI MACH EINS: DIE NEUE VERDIENSTERHEBUNG

Claudia Finke, Susanna Geisler, Anja Überschaer

📌 **Schlüsselwörter:** Verdiensterhebung – Reallöhne – Mindestlohn – Niedriglohn – Gender Pay Gap

ZUSAMMENFASSUNG

Im Jahr 2022 hat die neu eingeführte Verdiensterhebung die drei ehemaligen Verdienststatistiken – die Verdienststrukturerhebung, die Vierteljährliche Verdiensterhebung sowie die freiwilligen Verdiensterhebungen – abgelöst. Neben den Hintergründen und der Ausgangslage für diese Umgestaltung beschreibt der Beitrag die moderne Datengewinnung, die Datenaufbereitung sowie die dafür verwendeten statistischen Methoden. Ebenso steht das modernisierte Daten- und Veröffentlichungsangebot der neuen Verdienststatistik im Fokus.

📌 **Keywords:** *earnings survey – real earnings – minimum wage – low wage – gender pay gap*

ABSTRACT

In 2022, the new earnings survey was introduced which replaced the three previous earnings surveys – the structure of earnings survey, the quarterly survey of earnings and the voluntary earnings surveys. In addition to discussing the background and the situation at the start of the reorganisation, the article describes the modern way of data collection, the data processing and the statistical methods used. Furthermore, the article highlights the modernised range of data and of publications available on the new earnings statistics.

1

Ausgangslage: geänderte Anforderungen an die Verdienststatistik

In der amtlichen Verdienststatistik wurden bis zum Jahr 2018 alle vier Jahre die Verdienststrukturerhebung durchgeführt und bis 2022 die Vierteljährliche Verdiensterhebung. Den Schwerpunkt der Verdienststrukturerhebung bildeten Analysen zur Verdienstverteilung der Arbeitnehmerinnen und Arbeitnehmer. Um entsprechende Auswertungen zu ermöglichen, übermittelten die berichtspflichtigen Betriebe Einzelangaben zu den Beschäftigten an die Statistischen Ämter der Länder. Die Vierteljährliche Verdiensterhebung ermöglichte demgegenüber auf Basis der erhobenen Daten Aussagen zur Entwicklung der Verdienste. Da hierfür keine Einzelangaben je Beschäftigten notwendig waren, wurden in der Vierteljährlichen Verdiensterhebung lediglich aufsummierte Angaben zu Vollzeit-, Teilzeit- und geringfügig Beschäftigten bei den Betrieben erfragt. Sowohl bei der Verdienststrukturerhebung als auch bei der Vierteljährlichen Verdiensterhebung waren die Betriebe gesetzlich zur Auskunft verpflichtet. Um Datenlücken rund um das Thema „Mindestlohn“ zu schließen, fand zudem in den Jahren 2015, 2016, 2017 und 2019 eine freiwillige Verdiensterhebung nach §7 Absatz 1 Bundesstatistikgesetz statt (Schymura, 2020). Solche freiwilligen Sondererhebungen können laut Bundesstatistikgesetz zur Erfüllung eines kurzfristig auftretenden Datenbedarfs oberster Bundesbehörden, aber nicht dauerhaft, durchgeführt werden.

Ausgangspunkt für die Neukonzeption der Verdienststatistik waren verschiedene neue Datenbedarfe. So äußerte die Konferenz der Gleichstellungs- und Frauenministerinnen und -minister, -senatorinnen und -senatoren der Länder ihren Bedarf an jährlichen Zahlen zum bereinigten und unbereinigten Gender Pay Gap. Die Geschäftsstelle der Mindestlohnkommission benötigte zudem in kürzeren Abständen Daten für die Mindestlohnforschung, da sie alle zwei Jahre über die Anpassung der Höhe des 2015 eingeführten gesetzlichen Mindestlohns entscheidet. Beide Anforderungen sind nur mit jährlichen Angaben je Beschäftigten zu erfüllen. Weiterhin sollte die neue Verdienststatistik sicherstellen, dass unterjährige Verdienstangaben zur Verfügung

stehen, die wiederum Aussagen zur Verdienstentwicklung ermöglichen.

Der Artikel beschreibt in Kapitel 2 die Konzeption und den Aufbau der neuen Verdiensterhebung. Welche Daten einfließen, wie sie aufbereitet und hochgerechnet werden, ist Thema der Kapitel 3, 4 und 5. Informationen zur Geheimhaltung und Fehlerrechnung sowie zum Veröffentlichungsangebot enthalten die Kapitel 6 und 7. Der Beitrag schließt mit einem Ausblick auf geplante weitere Nutzungsmöglichkeiten der Daten der neuen Verdiensterhebung.

2

Verschmelzung der bisher durchgeführten Verdienststatistiken zur neuen Verdiensterhebung

Das Bundesministerium für Wirtschaft und Energie (2023: Bundesministerium für Wirtschaft und Klimaschutz) hat das Statistische Bundesamt Ende 2018 beauftragt, ein Konzept zu entwickeln, das die oben aufgeführten neuen Nutzungsanforderungen erfüllt und dabei keine zusätzlichen Bürokratiekosten verursacht. Daraufhin haben die Statistischen Ämter des Bundes und der Länder gemeinsam die Idee entwickelt, die bisher durchgeführten Verdiensterhebungen zu einer monatlichen Statistik zu verschmelzen. Der Grundgedanke war, die bisher bestehenden Statistiken durch eine Erhebung zu ersetzen, die hinsichtlich der abgedeckten Beschäftigtengruppen und Wirtschaftszweige so vollständig wie die Verdienststrukturerhebung ist und es zugleich ermöglicht, die Verdienstentwicklung am aktuellen Rand zu beobachten. Die neu konzipierte Verdiensterhebung ist eine monatliche Befragung bei Betrieben ab einem sozialversicherungspflichtigen Beschäftigten in der Land- und Forstwirtschaft, im Produzierenden Gewerbe sowie im Dienstleistungsbereich. Sie basiert auf einer repräsentativen Stichprobe und erfasst Verdienste und Arbeitsstunden sämtlicher Arbeitnehmerinnen und Arbeitnehmer des jeweiligen Betriebs. Darüber hinaus sind lohdeterminierende Merkmale der Beschäftigten, wie etwa Geschlecht, Alter, Staatsangehörigkeit, höchster Bildungsabschluss und Beruf, Bestandteil der Erhebung. Die Betriebe sind gesetzlich zur Auskunft verpflichtet.

Um die Belastung der Betriebe zu minimieren, wurde bei der Konzeption der Verdiensterhebung das Digitalisierungs- und Automatisierungspotenzial ausgeschöpft. Konkret betrifft dies die in den folgenden Abschnitten dargestellten Aspekte.

2.1 Nutzung von bereits digital vorliegenden Angaben

Bei der neuen Verdiensterhebung wurden die zu erhebenden Merkmale und deren Definition konsequent daran ausgerichtet, welche Angaben in den Betrieben digital vorliegen.

Sowohl in der Verdienststrukturerhebung als auch in der Vierteljährlichen Verdiensterhebung war die Leistungsgruppe der Beschäftigten ein erfragtes Merkmal. Dabei handelte es sich um eine ausschließlich für die Statistik durch den Betrieb zu pflegende Angabe, die Rückschlüsse zum Anforderungsniveau des Arbeitsplatzes und zur Qualifikation der Beschäftigten zuließ. Dagegen erfasst die neue Verdiensterhebung den Tätigkeitsschlüssel der Beschäftigten. Dieser liegt bei den Betrieben für sozialversicherungspflichtig Beschäftigte bereits vor, da er bei der Meldung zur Sozialversicherung anzugeben ist. Der neunstellige Schlüssel enthält verschiedene Informationen, beispielsweise die Tätigkeit, die Vertragsform sowie die Schul- und Ausbildung der Beschäftigten.

In der Verdiensterhebung ist es anders als in der Vierteljährlichen Verdiensterhebung nicht notwendig, Verdienst- und Stundenangaben über Arbeitnehmergruppen und über drei Berichtsmonate hinweg aufzusummieren. Die Übermittlung der Daten erfolgt für die Verdiensterhebung so, wie die Angaben in den Betrieben digital bereits vorliegen: Monatlich werden Einzelangaben zu den Beschäftigten an die statistischen Ämter geliefert.

Weil sich die Verdienstdefinition an der Entgeltbescheinigungsverordnung orientiert, kann der Betrieb beispielsweise das Gesamtbruttoentgelt melden, das auch in der Lohnbescheinigung der Beschäftigten als zentrale Angabe enthalten ist.

Schließlich wird – sofern möglich – auf die Nutzung von Sekundärquellen zurückgegriffen. So wird beispielsweise der Wirtschaftszweig des Betriebes nicht bei den Auskunftspflichtigen erfragt, sondern aus dem

statistischen Unternehmensregister an die Erhebung herangespielt. Auch die Betriebsgröße, die Form der wirtschaftlichen und finanziellen Kontrolle sowie der Urlaubsanspruch der Beschäftigten sind nicht von den Betrieben zu melden, da die entsprechenden Angaben ebenfalls in Sekundärquellen vorliegen.

2.2 Abgabe von Online-Meldungen

Nach §11a Absatz 2 Bundesstatistikgesetz sind die Angaben von den Betrieben online zu übermitteln. Hierfür existieren zwei Verfahren:

- › Die Betriebe können das Online-Meldeverfahren eSTATISTIK.core nutzen, um die Merkmale an die statistischen Ämter zu übermitteln. Bei diesem Verfahren werden Statistikdaten automatisiert aus der Lohnbuchhaltung der Betriebe gewonnen und online an die zentrale Annahmestelle der statistischen Ämter übermittelt.
- › Alternativ steht den Betrieben ein Online-Fragebogen (IDEV) zur Übermittlung der Daten zur Verfügung.

Rund drei Viertel der Auskunftspflichtigen meldeten die Angaben für den Berichtsmonat Dezember 2022 über eSTATISTIK.core und ein Viertel über das Online-Formular. Gerade große Betriebe verfügen über Lohnsoftware und übermitteln die Angaben „per Knopfdruck“ über eSTATISTIK.core.

2.3 Meldung aller Beschäftigten

Die Betriebe melden sämtliche Beschäftigte zur Verdiensterhebung. Sie sind nicht gezwungen, erhebungsrelevante Arbeitnehmerinnen und Arbeitnehmer zu filtern. Angaben zu Beschäftigten, die für die Erhebung nicht relevant sind – hierzu zählen beispielsweise Personen, die ein freiwilliges soziales oder ökologisches Jahr absolvieren –, werden automatisiert ausgeschlossen.

2.4 Imputation von fehlenden oder fehlerhaften Angaben

Bei der neuen Verdiensterhebung wird weitgehend auf Rückfragen bei den Betrieben verzichtet. Fehlende oder unplausible Angaben werden mithilfe eines automatisierten Imputationsverfahrens ersetzt (siehe Kapitel 4).

3

Die Bestandteile der Verdiensterhebung

Die Datenbasis der Verdiensterhebung setzt sich aus drei Teilen zusammen. Die Erhebungsdaten der Stichprobe sind das Herzstück und haben den größten Anteil. Fortgeschätzte Daten aus der Personalstandstatistik des öffentlichen Dienstes und Imputationen für Betriebe, die keine sozialversicherungspflichtig Beschäftigten haben, ergänzen die Erhebungsdaten.

Die Imputation der Betriebe ohne sozialversicherungspflichtig Beschäftigte sowie die gebundene Hochrechnung erfolgen regulär einmal im Jahr für den Erhebungsmonat April, der bereits bei der Verdienststrukturerhebung als repräsentativer Erhebungsmonat genutzt wurde. Für einige Merkmale der Verdiensterhebung wird das statistische Unternehmensregister als Sekundärquelle genutzt (siehe Kapitel 2.1).

3.1 Primär erhobene Daten

Die Auswahlgrundlage umfasst alle Betriebe des statistischen Unternehmensregisters im aktuellsten verfügbaren Berichtsjahr ($t-1$) zum Zeitpunkt der Stichprobenziehung (t), die

- › einer der Wirtschaftsabteilungen (sogenannte Zweisteller) der Abschnitte A bis S der Klassifikation der Wirtschaftszweige 2008 (WZ 2008) angehören,
- › wirtschaftlich aktiv und nicht in einem Insolvenzverfahren sind und
- › mindestens eine(n) sozialversicherungspflichtig Beschäftigte(n) haben.

Als Stichprobendesign dient eine geschichtete Zufallsstichprobe. Dazu wird die Auswahlgrundlage in rund 9 500 Schichten unterteilt, die sich als Kreuzkombination der Merkmale Bundesland, Wirtschaftsabteilung plus Extra-Schichten für Krankenhäuser und sieben Beschäftigtengrößenklassen ergeben. Der Gesamtumfang der Stichprobe beträgt 58 000 Betriebe (Obergrenze laut § 4 Absatz 3 Verdienststatistikgesetz). Die Stichprobe wird rollierend durchgeführt, das heißt es wird jährlich ein Sechstel der Betriebe aus der Berichtspflicht entlassen

und durch neue Betriebe ersetzt. Zudem wird die Stichprobe dabei auf die gesetzlich zulässige Höchstgrenze von 58 000 Betrieben aufgefüllt. Das ist erforderlich, weil jedes Jahr ein kleiner Teil der Betriebe seine Tätigkeit aufgibt. Dieses Vorgehen erhöht die Datenqualität, da eine komplett neue Stichprobenziehung die Vergleichbarkeit der Daten über längere Zeiträume hinweg beeinträchtigen würde. Darüber hinaus verteilt dieses Vorgehen die Belastung insbesondere bei den kleineren Betrieben breiter über die Wirtschaft. Jährlich rotiert ein Sechstel der Stichprobenbetriebe, künftig steht den Datennutzenden somit in jedem Jahr ein 5/6-Panel der Berichtsbetriebe für zwei Jahre zur Verfügung.

3.2 Personalstandstatistik

Die Personalstandstatistik „liefert Daten über die Beschäftigten der öffentlichen Arbeitgeber, die in einem unmittelbaren Dienst- oder Arbeitsvertragsverhältnis mit der jeweiligen Einrichtung stehen“ (Statistisches Bundesamt, 2021, hier: Seite 3). Diese Daten enthalten auch Informationen über den Verdienst der Beschäftigten. Für die Wirtschaftsabschnitte „Erziehung und Unterricht“ sowie „Öffentliche Verwaltung, Verteidigung; Sozialversicherung“ wird daher zum überwiegenden Teil auf eine (erneute) Befragung verzichtet, stattdessen werden die Daten der Personalstandstatistik verwendet. Diese werden mithilfe von Tarifvertragsinformationen auf die jeweils aktuellen Berichtsmonate fortgeschrieben.

3.3 Imputation der Betriebe ohne sozialversicherungspflichtig Beschäftigte

Angaben von Betrieben, die ausschließlich geringfügig Beschäftigte haben und somit keine sozialversicherungspflichtigen Personen beschäftigen, werden imputiert. Bei diesen oftmals sehr kleinen Betrieben wäre eine Befragung im Vergleich zum Erkenntnisgewinn unverhältnismäßig.

Die Imputation der Angaben zu den Betrieben und ihren Beschäftigten erfolgt durch ein Nearest-Neighbour-Verfahren (CANCEIS, Beschreibung siehe Kapitel 4).¹

¹ Für frühere Beschreibungen der Methode siehe Frentzen/Günther (2017) sowie Statistisches Bundesamt (2017).

Dafür wird zunächst aus dem Verwaltungsdatenspeicher² eine Stichprobe von deutschlandweit 10 000 Betrieben mit nur (ausschließlich) geringfügig entlohnt Beschäftigten gezogen. Die Stichprobe wird dabei nach Bundesland, Zahl der Beschäftigten im Betrieb und Wirtschaftsabteilungen geschichtet. Um die Anzahl an Schichten zu reduzieren, werden Zweisteller, in denen nur wenige Beschäftigte tätig sind, zusammengefasst. Der Verwaltungsdatenspeicher enthält auch Informationen über Wirtschaftszweig, Bundesland, Regierungsbezirk, Kreis und Gemeinde für jeden gezogenen Betrieb. Anhand dieser Merkmale und der Größenklasse des Betriebs werden die Spender aus der aktuellen Erhebung heraus ermittelt. Dabei werden die Betriebe als Spender ausgewählt, die dem zu imputierenden Betrieb in Bezug auf Wirtschaftszweig, regionale Lage und Betriebsgröße am ähnlichsten sind. Der Verwaltungsdatenspeicher liefert auch Informationen über die Anzahl der (ausschließlich) geringfügig entlohnt Beschäftigten in den Betrieben. Um die endgültige Anzahl an geringfügig entlohnt Beschäftigten (ausschließlich geringfügig entlohnt Beschäftigte und in Nebentätigkeit geringfügig entlohnt Beschäftigte) zu ermitteln, werden die ausschließlich geringfügig entlohnt Beschäftigten mit dem Faktor der in Nebentätigkeit geringfügig entlohnt Beschäftigten multipliziert.³ Die Angaben zu den Beschäftigten werden anschließend vollständig imputiert. Als Spender werden die geringfügig entlohnt Beschäftigten der aktuellen Erhebung genutzt. Insgesamt werden so Angaben für 10 000 Betriebe und etwa 45 000 Beschäftigte imputiert.

4

Datenaufbereitung

Den Kern der Aufbereitung der bei den Betrieben erhobenen Daten bilden umfangreiche Prüfungen der gemeldeten Daten. Bei fehlenden Angaben oder Unstimmigkeiten erfolgt in der Regel eine automatisierte Imputation unter

2 Der Verwaltungsdatenspeicher entspricht dem kompletten Bestand der Betriebe mit Meldungen an die Bundesagentur für Arbeit (Statistisches Bundesamt, 2023a).

3 Der Faktor ergibt sich aus dem Verhältnis der geringfügig entlohnt Beschäftigten der Verdiensterhebung zu den ausschließlich geringfügig entlohnt Beschäftigten aus dem Verwaltungsdatenspeicher. Dieser Quotient wird für jeden Wirtschaftszweig in jedem Bundesland ermittelt.

Einsatz von CANCEIS (CANadian Census Edit and Imputation System). Bei dieser vom kanadischen Statistikamt entwickelten Software handelt es sich um ein Imputationsverfahren nach dem Nearest-Neighbour-Prinzip. Es kategorisiert die erfassten Datensätze in unplausible und plausible Datensätze. Dabei dienen die plausiblen Datensätze als Spender. Bei den Datensätzen, die Fehler aufweisen, handelt es sich um „Empfänger“. Das Ziel des Imputationsverfahrens besteht darin, dass unplausible Datensätze keine Unstimmigkeiten mehr aufweisen. Dies wird erreicht, indem die Ausprägung eines Merkmals eines plausiblen Datensatzes das fehlerhafte oder fehlende Merkmal eines unplausiblen Datensatzes ersetzt. Dabei wird beim Nearest-Neighbour-Verfahren der Spenderdatensatz ausgewählt, der sich am wenigsten vom Empfängerdatensatz unterscheidet. Sollten mehrere „Nächste Nachbarn“ als Spender infrage kommen, wird ein Datensatz zufällig aus den potenziell geeigneten gezogen (CANCEIS Development Team, 2020). Dieser Datensatz „spendet“ dem unplausiblen Datensatz anschließend Kopien seiner Merkmalsausprägungen. Da sowohl Spender als auch Empfänger aus derselben Datenmatrix stammen, ist CANCEIS den sogenannten Hot-Deck-Verfahren zuzuordnen.

Nur in Einzelfällen kommt das beschriebene Imputationsverfahren nicht zum Einsatz. In diesen Einzelfällen kontaktieren die Mitarbeitenden in den Statistischen Ämtern der Länder die Betriebe bei Unstimmigkeiten in den gemeldeten Angaben und korrigieren diese bei Bedarf manuell. Das trifft insbesondere auf Betriebe zu, die Besonderheiten in Bezug auf die Arbeitszeit oder Verdienstangaben aufweisen (zum Beispiel extrem hohe Verdienste oder niedrige Wochenarbeitszeit). Hier korrigiert die Bearbeiterin oder der Bearbeiter fehlerhafte Werte nach Rücksprache mit dem Betrieb.

Die Tatsache, dass die Angaben in der Regel aus der Lohnabrechnung der Betriebe stammen, wirkt sich positiv auf die Datenqualität aus. Beispielsweise wurde der Bruttomonatsverdienst im Oktober 2022 lediglich in weit unter 1 % der gemeldeten Datensätze als unplausibel gewertet. Bei zwei Dritteln der betroffenen Datensätze hat CANCEIS plausible Verdienstangaben imputiert, die übrigen Fällen wurden manuell korrigiert.

5

Gebundene Hochrechnung

Um von der Stichprobe der Betriebe (und somit der Beschäftigten) wieder auf alle Betriebe und Beschäftigte der Grundgesamtheit schließen zu können, werden die Daten hochgerechnet. Im einfachsten Fall erfolgt dies durch das Horvitz-Thompson-Verfahren (Horvitz/Thompson, 1952). Hierbei wird der Kehrwert der Auswahlwahrscheinlichkeit der Stichprobeneinheit als Hochrechnungsfaktor angesetzt. Dieser Hochrechnungsfaktor wird auch als Design-Hochrechnungsfaktor bezeichnet.

Bei der gebundenen Hochrechnung werden zusätzlich externe Informationen über die Grundgesamtheit genutzt. Im Falle der Verdiensterhebung werden die gebundenen Hochrechnungsgewichte so bestimmt, dass sie kohärent zu den Informationen aus dem Verwaltungsdatenspeicher und den dort vorliegenden Meldungen aller Betriebe an die Bundesagentur für Arbeit sind.¹⁴ Dabei wird der Design-Hochrechnungsfaktor mit einem zusätzlichen Gewicht so angepasst, dass bestimmte Eckwerte der externen Quelle getroffen werden. Die Kombination aus dem Design-Hochrechnungsfaktor und dem zusätzlichen Gewicht ergibt somit den gebundenen Hochrechnungsfaktor.

Mit diesem Verfahren werden zwei Ziele verfolgt:

1. die Qualität zu verbessern
2. die Kohärenz zur externen Quelle zu steigern

Durch die Einbindung von Informationen aus der externen Quelle wird der stichprobenbedingte Zufallsfehler reduziert und somit ein Genauigkeitsgewinn erzielt. Zusätzlich kann der Verzerrung aufgrund von systematischen Antwortausfällen entgegengewirkt werden, was die Qualität ebenfalls erhöht. Durch die gebundene Hochrechnung gelingt es zudem, die hochgerechneten Werte der Stichprobe mit den Totalwerten der externen Quelle in Übereinstimmung zu bringen.

In der Verdiensterhebung wird das Verfahren Generalised Regression Estimator (GREG) genutzt¹⁵, welches zur

¹⁴ Für frühere Beschreibungen der Methode siehe Frentzen/Günther (2017) sowie Statistisches Bundesamt (2017).

¹⁵ Für einen Vergleich von GREG mit anderen Hochrechnungsverfahren siehe Frentzen/Günther (2017).

Schätzung von endlichen Grundgesamtheiten auf Basis einer Stichprobe unter der Zuhilfenahme von multivariaten Hilfsmerkmalen dient. Die zusätzlichen Gewichte ergeben sich dabei aus einem Regressionsmodell der abhängigen Variablen auf die Hilfsmerkmale (Deville/Särndal, 1992). Das Regressionsmodell wird dabei aus der Stichprobe selbst geschätzt. GREG benötigt dazu die Eckwerte der Hilfsmerkmale sowie die Angaben zu den Hilfsmerkmalen für alle Einheiten der Stichprobe. Wichtig ist dabei, dass die Hilfsmerkmale einen starken Zusammenhang mit der Zielvariable haben und diese gut erklären können. Jedoch erfüllen potenzielle Hilfsmerkmale die an sie gestellten Bedingungen (Gleichartigkeit der Merkmale, hohe Korrelation) häufig nicht in ausreichendem Maße, was das Verfahren beeinträchtigt (Böltkén, 2013). Für die Verdiensterhebung kann allerdings auf Hilfsmerkmale zurückgegriffen werden, die in der externen Quelle vorliegen und mit den Erhebungsdaten eindeutig verknüpfbar sind.

Hilfsmerkmale sind:

- › Zahl der Betriebe mit Beschäftigten
- › Zahl der sozialversicherungspflichtig Beschäftigten
- › Zahl der ausschließlich geringfügig Beschäftigten

Die Hilfsmerkmale sind untergliedert nach:

- › Bundesland (48 Eckwerte)
- › Größe des Betriebs (7 Klassen anhand der Zahl der sozialversicherungspflichtig Beschäftigten, 21 Eckwerte)
- › Wirtschaftszweig (Wirtschaftsabschnitte der WZ 2008, wobei die Wirtschaftsabschnitte O und P zusammengefasst werden, da hier die Daten überwiegend aus der Personalstandstatistik kommen, 54 Eckwerte)

Insgesamt stehen somit 123 Eckwerte für die Anpassung zur Verfügung.

6

Aufbereitung für Veröffentlichungen

Um bei Veröffentlichungen (siehe Kapitel 7) zu verhindern, dass Rückschlüsse auf einzelne Beschäftigte oder Betriebe möglich sind, werden verschiedene Maßnahmen ergriffen. So werden Ergebnisse, denen nicht eine

bestimmte Mindestzahl an Beschäftigten und Betrieben zugrunde liegt, geheim gehalten. Darüber hinaus wird die hochgerechnete Anzahl der Beschäftigten ausschließlich auf 1 000 gerundet ausgegeben. Dies verhindert eine potentielle Berechnung einzelner gesperrter Zellen anhand der veröffentlichten Werte anderer Zellen und eine weitere Sperrung von Werten durch eine sogenannte sekundäre Geheimhaltung kann entfallen. Auch beziehen sich die angegebenen Anzahlen (mit Bezug auf die Beschäftigten) und Mittelwerte beziehungsweise Mediane auf die hochgerechneten Daten. Da Informationen über die Hochrechnungsfaktoren nicht veröffentlicht werden, ist eine Ermittlung der nicht hochgerechneten Anzahl und folglich eine Summenbildung auf diese Weise kaum möglich.

Zusätzlich zu Maßnahmen zur Geheimhaltung erfolgen Vorkehrungen zur Qualitätssicherung. Die Ergebnisse der Verdiensterhebung basieren auf einer repräsentativen Zufallsstichprobe, die zu anderen Ergebnissen führen kann, wenn wiederholt Stichproben gezogen und die Betriebe befragt würden. Diese Ergebnisschwankungen werden als Stichprobenzufallsfehler bezeichnet und durch anerkannte Stichprobenmethoden (Schichtung und gebundene Hochrechnung) reduziert. Das Ausmaß dieser Schwankungen kann mithilfe des relativen Standardfehlers geschätzt werden. Ergebnisse mit einem relativen Standardfehler zwischen 5 und 10% werden in Klammern dargestellt, um ein erhöhtes Maß an Unsicherheit zu kennzeichnen. Bei einem relativen Standardfehler über 10 % wird ein Wert nicht veröffentlicht. Die Methodik zur Fehlerrechnung mit dem SAS-Makropaket CLAN von Statistics Sweden wurde aus der Verdienststrukturhebung für die neue Verdiensterhebung übernommen. Im April 2022 lag der geschätzte relative Standardfehler für den durchschnittlichen Bruttomonatsverdienst je Beschäftigungsverhältnis bei 0,23 %⁶.

⁶ Die geschätzten relativen Standardfehler für andere Merkmale enthält der Qualitätsbericht der Verdiensterhebung 2023 (Statistisches Bundesamt, 2023a).

7

Veröffentlichungsangebot der Verdiensterhebung

Die ehemaligen Verdiensterhebungen wurden von vielen Akteuren sowie Statistiken als wichtige Datenquelle genutzt (Klemt/Droßard, 2013). All diese Bedarfe werden seit dem Berichtsjahr 2022 durch die neue Verdiensterhebung gedeckt. Auf der Internetseite des Statistischen Bundesamtes unter www.destatis.de sowie in der Datenbank GENESIS-Online steht ein umfangreiches Veröffentlichungsangebot für die Allgemeinheit kostenfrei und digital zur Verfügung.

7.1 Die Entwicklung von Verdiensten im Zeitverlauf: Verdienstindizes

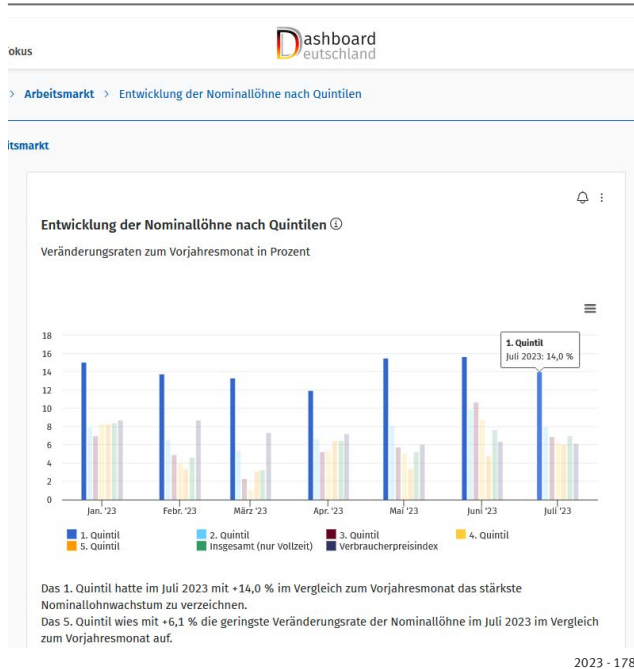
Die Verdiensterhebung ist als monatliche Erhebung auch eine Konjunkturstatistik, die kurzfristige Entwicklungen von Durchschnittsverdiensten abbildet. Wie haben sich die Verdienste real verändert, das heißt unter Berücksichtigung der Entwicklung der Verbraucherpreise? Der dafür konzipierte Reallohnindex beantwortet diese häufig gestellte Frage. Der Nominallohnindex stellt dazu ab dem Berichtsjahr 2022 die Entwicklung der durchschnittlichen Bruttomonatsverdienste einschließlich Sonderzahlungen vollständig für alle Beschäftigten der Gesamtwirtschaft in Deutschland bereit. Bisher wurden diese Indizes mit einer vierteljährlichen und jährlichen Periodizität berechnet. Seit 2022 stehen alle Verdienstindizes monatlich auch als Schnellindizes zur Verfügung. Weiter werden Indizes der Bruttomonats- und Bruttostundenverdienste (jeweils ohne Sonderzahlungen) von Vollzeitbeschäftigten berechnet, welche vor allem bei Wertsicherungsklauseln in längerfristigen Verträgen Verwendung finden.

Das Angebot der Indizes erstreckt sich über verschiedene Gliederungsebenen, die eine sehr genaue Analyse der Verdienstentwicklung in Deutschland ermöglichen – nach Geschlecht, Branche, beruflichem Ausbildungsabschluss und Gebietsstand. Seit 2022 existiert dazu noch die Darstellung der Verdienstentwicklung nach Quintilen, welche die Bruttomonatsverdienste einschließlich Sonderzahlungen von Vollzeitbeschäftigten nach fünf Verdienstgrößenklassen misst. Dieses neue und innova-

tive Datenangebot ist als Visualisierung im [Dashboard Deutschland](#) im Themenbereich „Arbeitsmarkt“ sowie in der Datenbank [GENESIS-Online](#) zu finden. Es ermöglicht erstmals die Analyse, welche Beschäftigtengruppe (Gering-, Mittel- oder Hochverdienende) am stärksten oder schwächsten von der allgemeinen Lohnentwicklung profitiert beziehungsweise welche Beschäftigtengruppen eine Verdienstentwicklung oberhalb der Inflation aufweisen. [↪ Grafik 1](#)

Grafik 1

Datenangebot der neuen Verdiensterhebung im Dashboard Deutschland



7.2 Niveaus und Verteilung der durchschnittlichen Verdienste – arithmetisches Mittel und Median

Um eine Einschätzung über das Verdienstniveau zu erhalten, werden die Erhebungsmerkmale zu verschiedenen Lagemaßen aggregiert und veröffentlicht. Das neue Erhebungsdesign der Verdiensterhebung erlaubt dabei, neben dem gängigen arithmetischen Mittel auch in nun jährlichem Turnus den Median anzugeben. Die hierfür benötigten Verdienstinformationen auf Ebene der Beschäftigten liegen nun monatlich vor. Da die Verdienstverteilung rechtsschief ist – zwei Drittel der Vollzeitbeschäftigten verdienen weniger als das arithme-

tische Mittel –, kann durch den Median somit ergänzend ein gegen Ausreißer robuster Wert angegeben werden (Günther, 2013).

7.3 Der interaktive Gehaltsvergleich: individuelle Schätzung von Verdiensten

Das individuelle Gehalt weicht von diesen Durchschnittsverdiensten in der Regel ab, denn es ist von sehr vielen Faktoren abhängig. Dazu gehören beispielsweise Berufserfahrung und Bildungsabschluss, aber auch unternehmensbezogene Merkmale wie Branche oder Unternehmensgröße. Seit dem Jahr 2020 besteht mit dem [interaktiven Gehaltsrechner](#) die Möglichkeit, sich auf Basis von individuellen verdienstrelevanten Angaben ein geschätztes monatliches Bruttogehalt berechnen zu lassen. So können die Nutzerinnen und Nutzer zum Beispiel sehen, ob sie mit ihrer Ausbildung in einer anderen Branche mehr verdienen würden, oder ob es sich lohnt, eine Meisterprüfung anzustreben (Überschaer, 2021).

7.4 Geschlechterspezifische Verdienstunterschiede: Gender Pay Gap und Gender Gap Arbeitsmarkt

Sowohl die Gleichstellungsstrategie als auch der Koalitionsvertrag 2021 bis 2025 der Bundesregierung fordern, die Lohnlücke zwischen Frauen und Männern zu schließen (Bundesregierung, 2021/2022). Der zentrale Indikator zur Messung dieser Lohnlücke ist der Gender Pay Gap, der den Verdienstabstand je Stunde zwischen Frauen und Männern beschreibt. Für eine weitere Differenzierung wird zwischen dem unbereinigten und dem bereinigten Gender Pay Gap unterschieden.

Der unbereinigte Gender Pay Gap wird als Differenz zwischen den durchschnittlichen Bruttostundenverdiensten von Frauen und Männern in Prozent des durchschnittlichen Bruttostundenverdiensts männlicher Beschäftigter definiert. Es handelt sich hierbei um die Lohnlücke ohne Anpassungen. Dagegen wird beim bereinigten Gender Pay Gap jener Teil des Verdienstunterschieds herausgerechnet, der auf strukturelle Unterschiede zwischen den Geschlechtern zurückzuführen ist: Unter-

schiede im Hinblick auf Beruf, Branche, Beschäftigungsumfang, Qualifikation oder Karrierelevel. Zu berücksichtigen ist hier, dass nicht über alle lohnrelevanten Einflussfaktoren Informationen zur Verfügung stehen. Beispielsweise fehlen Angaben zu Erwerbsunterbrechungen (zum Beispiel durch Elternzeit), mit denen der bereinigte Gender Pay Gap geringer ausfallen würde. Daher darf der Wert nicht mit einer Verdienstdiskriminierung gleichgesetzt werden, sondern gilt eher als „Obergrenze“ für Verdienstdiskriminierung (Finke und andere, 2017; Beck, 2018; Mischler, 2021).

Das Thema Verdienstungleichheit zwischen den Geschlechtern hat noch weitere Dimensionen: Frauen nehmen zum einen seltener am Erwerbsleben teil als Männer, zum anderen arbeiten sie häufiger in Teilzeit. Der Gender Gap Arbeitsmarkt vereint neben dem Gender Pay Gap auch Unterschiede in der bezahlten Arbeitszeit (Gender Hours Gap) und in der Erwerbstätigenquote (Gender Employment Gap). Auf diese Weise erweitert der neue Indikator die Perspektive auf die Verdienstsituation von Frauen und Männern seit dem Berichtsjahr 2022. Der Gender Gap Arbeitsmarkt kann über den [Gender Gap Simulator](#) auf spielerische Art und Weise kennengelernt werden (Statistisches Bundesamt, 2023b).

7.5 Verdienste im Mindestlohnbereich und Niedriglohnsektor

In Deutschland gilt seit dem 1. Januar 2015 ein gesetzlicher Mindestlohn. Er gilt als unterste Lohngrenze für nahezu alle Arbeitnehmerinnen und Arbeitnehmer (mit wenigen Ausnahmen, wie Auszubildenden). Nach dem Mindestlohngesetz beschließt die Mindestlohnkommission, in der Gewerkschaften und Arbeitgeber vertreten sind, alle zwei Jahre die Mindestlohnentwicklung, die daraufhin per Rechtsverordnung verbindlich wird. Sie orientiert sich dabei nachlaufend an der Tariflohnentwicklung, gemessen am Tarifindex des Statistischen Bundesamtes. Die Verdiensterhebung bildet seit 2022 die Datengrundlage, welche die Auswirkung des Mindestlohns auf die allgemeine Lohn- und die Beschäftigungsentwicklung messen kann. Im Fokus steht vor allem die Fragestellung, ob die Mindestloohnerhöhungen bei den Beschäftigten ankommen (Frentzen und andere, 2018; Kann, 2018; Schymura, 2020).


Jobs im Mindestlohnbereich und der sogenannte Niedriglohnsektor sind naturgemäß eng miteinander verbunden. Darunter fallen laut internationaler Definition alle Beschäftigungsverhältnisse, die mit weniger als zwei Drittel des mittleren Verdienstes (Median) entlohnt werden. Die Messung und Darstellung des Niedriglohnsektors wird ebenfalls mit der neuen Verdiensterhebung bedient.

7.6 Tarifbindung in Deutschland

Der Indikator Tarifbindung ist als der Anteil der Beschäftigung in tarifgebundenen Betrieben an allen Beschäftigungsverhältnissen definiert. Als tarifgebunden gelten Betriebe, in denen ein Firmen- oder Branchentarifvertrag zur Anwendung kommt. Für Beschäftigte in Betrieben mit Tarifvertrag gelten Mindeststandards, allerdings gibt es in Branchentarifverträgen auch teilweise Öffnungsklauseln. Inwiefern sich die Anteile der Beschäftigung in tarifgebundenen Betrieben nach Branchen und Bundesländern unterscheiden, wird seit April 2022 mit der neuen Verdiensterhebung ermittelt und steht seit dem Sommer 2023 nun als jährliches (anstatt vierjährliches) Datenangebot zur Verfügung.

8

Ausblick

Da die Verdiensterhebung monatlich durchgeführt wird und sich ein Großteil der Stichprobe nicht verändert, soll auf Grundlage der erhobenen Daten ein Betriebspanel in Kombination mit einem Beschäftigtenpanel erstellt und der Forschung über das Forschungsdatenzentrum zugänglich gemacht werden. Mithilfe eines solchen Panels ließen sich Fragestellungen wie „Welche Beschäftigten(-gruppen) verlassen die Betriebe?“ oder „Welcher Midi-Job war im Vormonat ein Mini-Job?“ empirisch untersuchen. Darüber hinaus bietet dieses Panel aufgrund des Alleinstellungsmerkmals „Arbeitszeit“ Forschungsmöglichkeiten insbesondere zum Thema Mindestlohn, die keine andere Datengrundlage ermöglicht. 

LITERATURVERZEICHNIS

Alter, Hannah. *Vier Jahre Neukonzeption der Verdienststatistik: Ein Fazit aus Sicht der Vierteljährlichen Verdiensterhebung*. In: *Wirtschaft und Statistik*. Ausgabe 12/2010, Seite 1110 ff.

Beck, Martin. *Verdienstunterschiede zwischen Männern und Frauen nach Bundesländern*. In: *WISTA Wirtschaft und Statistik*. Ausgabe 4/2018, Seite 26 ff.

Die Bundesregierung. *MEHR FORTSCHRITT WAGEN – Bündnis für Freiheit, Gerechtigkeit und Nachhaltigkeit*. Koalitionsvertrag zwischen SPD, BÜNDNIS 90/DIE GRÜNEN UND FDP. 2021. [Zugriff am 27. Juni 2023]. Verfügbar unter: www.bundesregierung.de

Die Bundesregierung. *Die Gleichstellungsstrategie der Bundesregierung*. 2022. [Zugriff am 27. Juni 2023]. Verfügbar unter: www.bmbsfj.de

Böltken, Ferdinand. *Auswahlverfahren: Eine Einführung für Sozialwissenschaftler*. 1. Auflage. Stuttgart 1976.

CANCEIS Development Team. *CANCEIS User's Guide Version 5.4, Statistical Integration Methods Division*. 2020. Statistics Canada.

Deville, Jean-Claude/Särndal, Carl-Erik. *Calibration Estimators in Survey Sampling*. In: *Journal of the American Statistical Association*. 1992. Ausgabe 87/418, Seite 376 ff.

Finke, Claudia/Dumpert, Florian/Beck, Martin. *Verdienstunterschiede zwischen Männern und Frauen*. In: *WISTA Wirtschaft und Statistik*. Ausgabe 2/2017, Seite 43 ff.

Frentzen, Kathrin/Beck, Martin/Stelzer, Jonas. *Beschäftigungswirkungen des Mindestlohns*. In: *WISTA Wirtschaft und Statistik*. Ausgabe 1/2018, Seite 35 ff.

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

Günther, Roland. *Methodik der Verdienststrukturerhebung 2010*. In: *Wirtschaft und Statistik*. Ausgabe 2/2013, Seite 127 ff.

Horvitz, Daniel G./Thompson, Donovan J. *A Generalization of Sampling without Replacement from a finite Universe*. In: *Journal of the American Statistical Association*. 1952. Ausgabe 47/260, Seite 663 ff.

Kann, Kathrin. *Der Einfluss des Mindestlohns auf die Verdienststrukturen*. In: *WISTA Wirtschaft und Statistik*. Ausgabe 5/2018, Seite 44 ff.

Klemt, Sandra/Droßard, Ralf. *Vierteljährliche Verdiensterhebung: neue Nutzer, neue Indizes, die neuesten Ergebnisse*. In: *Wirtschaft und Statistik*. Ausgabe 8/2013, Seite 603 ff.

Mischler, Frauke. *Verdienstunterschiede zwischen Männern und Frauen*. In: *WISTA Wirtschaft und Statistik*. Ausgabe 4/2021, Seite 110 ff.

LITERATURVERZEICHNIS

Schymura, Sandra. *Beschäftigte und ihre Verdienste nach der zweiten Erhöhung des Mindestlohns*. In: WISTA Wirtschaft und Statistik. Ausgabe 6/2020, Seite 58 ff.

Statistisches Bundesamt. *Klassifikation der Wirtschaftszweige, Ausgabe 2008*. 2008.

Statistisches Bundesamt. *Qualitätsbericht Personalstandstatistik 2021*. 2021.

Statistisches Bundesamt. *Qualitätsbericht Verdiensterhebung – Erhebung der Arbeitsverdienste nach § 4 Verdienststatistikgesetz*. 2023a.

Statistisches Bundesamt. *Neuer Indikator „Gender Gap Arbeitsmarkt“ erweitert den Blickwinkel auf Verdienungleichheit*. Pressemitteilung Nr. 084 vom 6. März 2023. 2023b. [Zugriff am 21. August 2023]. Verfügbar unter: www.destatis.de

Statistisches Bundesamt. *Verdiensterhebung 2015 – Abschlussbericht einer Erhebung über die Wirkung des gesetzlichen Mindestlohns auf die Verdienste und Arbeitszeiten der abhängig Beschäftigten*. 2017.

Überschaer, Anja. *Der interaktive Gehaltsvergleich – eine profilbasierte Schätzung von Verdiensten*. In: WISTA Wirtschaft und Statistik. Ausgabe 6/2021, Seite 67 ff.

RECHTSGRUNDLAGEN

Gesetz über die Statistik für Bundeszwecke (Bundesstatistikgesetz – BStatG) in der Fassung der Bekanntmachung vom 20. Oktober 2016 (BGBl. I Seite 2394), das zuletzt durch Artikel 5 des Gesetzes vom 20. Dezember 2022 (BGBl. I Seite 2727) geändert worden ist.

Gesetz über die Statistik der Verdienste und Arbeitskosten (Verdienststatistikgesetz – VerdStatG) vom 21. Dezember 2006 (BGBl. I Seite 3291), das zuletzt durch Artikel 1 des Gesetzes vom 12. August 2020 (BGBl. I Seite 1872) geändert worden ist.

Gesetz zur Regelung eines allgemeinen Mindestlohns (Mindestlohngesetz – MiLoG) vom 11. August 2014 (BGBl. I Seite 1348), das zuletzt durch Artikel 2 des Gesetzes vom 28. Juni 2023 (BGBl. I Nr. 172) geändert worden ist.

Herausgeber

Statistisches Bundesamt (Destatis), Wiesbaden

Schriftleitung

Dr. Daniel Vorgrimler

Redaktion: Ellen Römer

Ihr Kontakt zu uns

www.destatis.de/kontakt

Erscheinungsfolge

zweimonatlich, erschienen im Oktober 2023

Ältere Ausgaben finden Sie unter www.destatis.de sowie in der [Statistischen Bibliothek](#).

Artikelnummer: 1010200-23005-4, ISSN 1619-2907

© Statistisches Bundesamt (Destatis), 2023

Vervielfältigung und Verbreitung, auch auszugsweise, mit Quellenangabe gestattet.