

WEITERENTWICKLUNG DER EUROPÄISCHEN PRODUKTIONSSTATISTIK

Luis Federico Flores, Luisa Baumgärtner

↳ **Schlüsselwörter:** Produktionsstatistik – Verwaltungsdaten – Multilevel-Modell – Geheimhaltung

ZUSAMMENFASSUNG

In einem von der europäischen Statistikbehörde (Eurostat) initiierten Projekt wurden verschiedene Methoden zur Modernisierung der europäischen Produktionsstatistik (PRODCOM) untersucht. Die Ergebnisse dieser jährlichen Statistik werden für Deutschland aus den nationalen Daten der Monatlichen und Vierteljährlichen Produktionserhebungen ermittelt. Diese Erhebungen erfassen in Deutschland die inländische Produktion von Unternehmen mit 20 oder mehr tätigen Personen. Der Beitrag beschreibt die Schätzung des Produktionswerts für Unternehmen mit weniger als 20 tätigen Personen mithilfe von Verwaltungsdaten und eines ökonometrischen Verfahrens zur Schätzung von Merkmalen neuer Einheiten (Multilevel-Modell). Zudem werden die untersuchten Möglichkeiten zur Reduzierung geheimzuhaltender Ergebnisse in der nationalen und europäischen Veröffentlichung aufgezeigt.

↳ **Keywords:** production statistics – administrative data – multilevel model – confidentiality

ABSTRACT

Various methods for modernising the European production statistics (PRODCOM) were examined in a project initiated by the Statistical Office of the European Union (Eurostat). In Germany the results of these annual statistics are based on the national data collected in the monthly and quarterly production surveys. The German surveys cover the domestic production of enterprises with 20 or more persons employed. This article describes how the production value of enterprises with less than 20 persons employed is estimated using administrative data and an econometric method for the estimation of variables of new units (multilevel model). In addition, the options are shown which were tested to reduce the scope of confidential results in the relevant national and European publications.

Luis Federico Flores

ist Volkswirt und befasste sich als wissenschaftlicher Mitarbeiter im Referat „Produktion der Industrie; Europäische „PRODCOM“-Statistik; Fachstatistiken“ des Statistischen Bundesamtes mit dem EU-Projekt „Modernisierung der europäischen Produktionsstatistik“. Im Rahmen dieser Arbeit hat er ein Modell zur Schätzung des Produktionswerts unterhalb der Abschneidegrenze entwickelt.

Luisa Baumgärtner

ist Betriebswirtin und Referentin im Referat „Produktion der Industrie; Europäische „PRODCOM“-Statistik; Fachstatistiken“ beim Statistischen Bundesamt. Aufgrund ihrer Mitarbeit in den Arbeitsgruppen der europäischen Produktionsstatistik bei Eurostat war sie beim Projekt zur „Modernisierung der europäischen Produktionsstatistik“ unterstützend tätig.

1

Einleitung

Produktionsdaten werden in Deutschland seit 1860 erhoben. Sie dienen der Beobachtung der konjunkturellen und strukturellen Entwicklung der gesamten Industrie sowie einzelner Branchen und Märkte (Bührer/Wagner, 2010, hier: Seite 110). Dies ist vor allem für die Politik (unter anderem das Bundesministerium für Wirtschaft und Energie), die Wissenschaft und die Wirtschaft selbst (zum Beispiel Wirtschaftsverbände und die Marktteilnehmenden) von Interesse. Innerhalb der amtlichen Statistik liefert die Produktionsstatistik wichtige Daten für viele Bereiche, beispielsweise für die Volkswirtschaftlichen Gesamtrechnungen, die Input-Output-Rechnung oder den monatlichen Produktionsindex. Die Produktionsstatistik erfasst die in Deutschland produzierten Güter in Menge und Wert nach Güterarten. Unter Gütern werden die von den Unternehmen im Inland produzierten Waren und industriellen Dienstleistungen verstanden. Grundlage für die Erfassung der Güter bildet das Güterverzeichnis für Produktionsstatistiken (GP, siehe Abschnitt 2.2) mit insgesamt etwa 5 100 verschiedenen Güterkategorien. Die heutige Produktionserhebung wird dezentral von den Statistischen Ämtern der Länder monatlich und vierteljährlich durchgeführt. Die Daten werden auf nationaler Ebene monatlich, vierteljährlich und jährlich veröffentlicht. Zudem werden die aggregierten Jahresdaten an das Statistische Amt der Europäischen Union (Eurostat) zur Veröffentlichung jährlicher, europaweiter Produktionsdaten (PRODCOM) übermittelt.

Eurostat strebt neben einer Harmonisierung der Erhebungs- und Berechnungsmethoden innerhalb der Mitgliedstaaten auch eine Optimierung der PRODCOM-Statistik an. Daher hat das Statistische Bundesamt im Rahmen eines von Eurostat ausgeschriebenem Projekts Methoden zur Verbesserung der Qualität und des Nutzens der PRODCOM-Statistik untersucht. Die Ergebnisse dieser Analysen und die daraus resultierenden Auswirkungen auf die nationale und europäische Produktionsstatistik sind Gegenstand dieses Aufsatzes.

Das Projekt umfasste zwei unterschiedliche Themenbereiche. Ein erster Teil des Projekts zielte darauf ab, auf der Grundlage von Verwaltungsdaten die Gesamtproduktion aller produzierenden Unternehmen zu

ermitteln. Sowohl in Deutschland als auch in zahlreichen anderen Mitgliedstaaten der Europäischen Union (EU) ist die Produktionserhebung eine Vollerhebung mit Abschneidegrenze. Das bedeutet, dass die kleinen Unternehmen von der gesetzlichen Meldepflicht zur Produktionsstatistik befreit sind. In Deutschland trifft dies auf die Unternehmen mit weniger als 20 tätigen Personen zu. Um die vollständige Abdeckung der inländischen Produktion ohne zusätzliche Belastung der Unternehmen zu erreichen, wurde vom Statistischen Bundesamt ein Schätzverfahren entwickelt. Mit dessen Hilfe wird der Produktionswert von Unternehmen unterhalb der gegenwärtigen Abschneidegrenze geschätzt. Das Schätzverfahren schließt eine Regressionsanalyse ein, die auf vorhandenen Erhebungsdaten basiert. Mithilfe von Verwaltungsdaten, die für die kleinen Unternehmen vorliegen, und des Regressionsmodells ist es möglich, den Produktionswert zu ermitteln.

Der zweite Teil des EU-Projekts befasste sich mit den Methoden zur Geheimhaltung von Ergebnissen in den Veröffentlichungen der nationalen und der europäischen Produktionsstatistik. Aktuell werden in der Ergebnislieferung an Eurostat nur rund 70% der Ergebnisse zur Veröffentlichung freigegeben. Die restlichen Ergebnisse sind aufgrund bestehender Geheimhaltungsregeln zur Wahrung des Schutzes von Einzelangaben gesperrt. Den gesetzlichen Rahmen für die Geheimhaltung von Einzelangaben und die daraus resultierenden Anwendungsregeln liefert §16 Bundesstatistikgesetz. Jedoch benötigen politische Entscheidungsträger, Verbände sowie andere Nutzerinnen und Nutzer auf nationaler und europäischer Ebene möglichst detaillierte Produktionsdaten. Ziel des zweiten Projektteils war, die aktuellen Geheimhaltungsmethoden im Rahmen der gesetzlichen Möglichkeiten zu überprüfen, um den Nutzerinnen und Nutzern der Produktionsstatistik mehr Daten zur Verfügung zu stellen.

Im Folgenden wird zunächst der grundlegende Prozess zur Erstellung der nationalen und europäischen Produktionsstatistik erläutert. Anschließend werden die zwei Teile des Projekts detailliert beschrieben und die Ergebnisse der Analysen dargestellt. Der Aufsatz schließt mit einem Ausblick, der die möglichen Änderungen für die nationale und europäische Produktionsstatistik aus deutscher Sicht aufzeigt.

2

Erstellung der nationalen und europäischen Produktionsstatistik

2.1 Berichtskreis der Erhebung

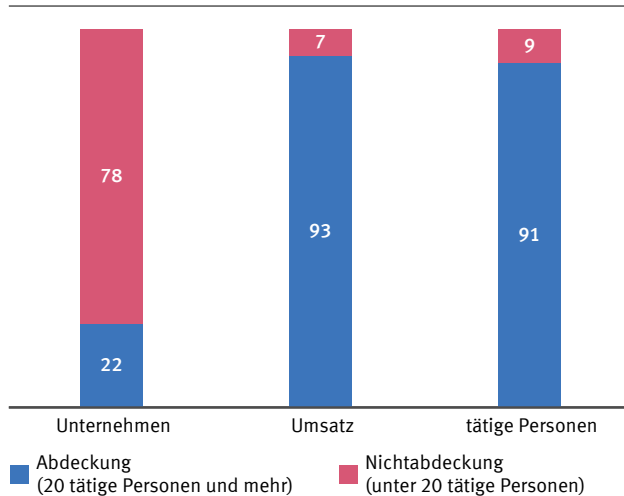
Die aktuelle Datenquelle der deutschen und der europäischen Produktionsstatistik ist die nationale Monatliche und Vierteljährliche Produktionserhebung. Die Grundgesamtheit dieser Erhebung umfasst alle Betriebe, die zu einem Unternehmen mit 20 oder mehr tätigen Personen gehören und in den Wirtschaftsabschnitten Bergbau und Gewinnung von Steinen und Erden (Abschnitt B der WZ 2008¹) oder Verarbeitendes Gewerbe (Abschnitt C der WZ 2008) tätig sind. Einbezogen sind ferner die produzierenden Betriebe mit 20 oder mehr tätigen Personen, die Unternehmen außerhalb der genannten Wirtschaftsabschnitte B und C angehören. Befragungs- und Beobachtungseinheit ist demnach der Betrieb, hingegen bildet das Unternehmen als rechtliche Einheit die Darstellungseinheit. Die Abschneidegrenze von 20 tätigen Personen wurde in der Bundesrepublik Deutschland 1975 mit dem neuen Gesetz über die Statistik im Produzierenden Gewerbe festgelegt (Bührer/Wagner, 2010, hier: Seite 114). Hauptgrund für die Festlegung dieses Schwellenwerts war, eine möglichst hohe Abdeckung der Industrieproduktion zu erreichen, ohne die kleinen Unternehmen zu belasten. Diese Vorgehensweise schadet der Aussagekraft der Statistik kaum, denn durch dieses Erhebungskonzept wird mithilfe einer vergleichsweise geringen Anzahl an Befragungseinheiten eine Abdeckung der Erhebungsmerkmale von mehr als 90% erreicht. Deshalb sind solche sogenannten Cut-off-Erhebungen in der Wirtschaft und in der amtlichen Statistik gängige Praxis.

↳ **Grafik 1** zeigt die Grundgesamtheit der Unternehmen in den WZ-Abschnitten B und C in Deutschland für das Berichtsjahr 2016. Nur 22% der 168 188 Unternehmen

¹ Die Klassifikation der Wirtschaftszweige, Ausgabe 2008 (WZ 2008), dient der Einordnung von Daten und Ergebnissen der amtlichen Statistik, die sich auf statistische Einheiten, beispielsweise Unternehmen, beziehen. Die Gliederung der WZ 2008 berücksichtigt die Vorgaben der statistischen Systematik der Wirtschaftszweige in der Europäischen Gemeinschaft (NACE Rev. 2).

wurden im Rahmen der Produktionsstatistik befragt (beziehungsweise die Betriebe dieser Unternehmen). Diese beschäftigten allerdings 91% aller tätigen Personen und generierten 93% des Gesamtumsatzes.

Grafik 1
Abdeckung von ausgewählten Merkmalen in der Produktionserhebung¹
in %



Auswertung des statistischen Unternehmensregisters des Berichtsjahres 2016, Stand April 2019.

¹ Abschnitte B "Bergbau und Gewinnung von Steinen und Erden" und C "Verarbeitendes Gewerbe" der WZ 2008.

2019 - 01 - 0597

2.2 Güterklassifikation

Die Monatliche und Vierteljährliche Produktionserhebung werden nach der Güterliste des Güterverzeichnisses für Produktionsstatistiken (GP; Statistisches Bundesamt, 2009) erhoben. Den Güterarten sind neunstellige Nummern zugeordnet, nach denen die Betriebe ihre Produktionsdaten melden. Die von Eurostat entwickelte und jährlich überarbeitete Liste von Produkten für die europäische Produktionsstatistik (PRODCOM-Liste) enthält achtstellige Produktnummern und ist demnach in den meisten Fällen weniger detailliert gegliedert als die nationale Güterklassifikation. Der Zusammenhang zwischen beiden Klassifikationen besteht darin, dass die ersten acht Stellen der GP-Nummern in der Regel der achtstelligen Nummer der PRODCOM-Liste entsprechen, die neunte Stelle ist den nationalen Unterteilungen vorbehalten. Das Güterverzeichnis wird – wenn auch in größeren Abständen – weitestgehend an die geänderte PRODCOM-Liste angepasst.

Die europäische Produktionsstatistik PRODCOM ist keine eigene Erhebung auf Basis der achtstelligen Meldenummern der PRODCOM-Liste. Vielmehr werden die erhobenen Einzeldaten, die auf der nationalen Güterklassifikation basieren, in die PRODCOM-Liste „übersetzt“. Dies erfolgt anhand einer Gegenüberstellung der Güternummern beider Klassifikationen. Somit ist es möglich, das Ergebnis auf Basis der PRODCOM-Liste zu erstellen ohne eine zusätzliche Erhebung durchzuführen.

2.3 Geheimhaltungsregeln

Sowohl die nationale Veröffentlichung als auch die Ergebniserstellung für die PRODCOM-Statistik unterliegen dem Geheimhaltungsgrundsatz gemäß § 16 Bundesstatistikgesetz. In Deutschland gilt die aktive Geheimhaltung. Das heißt, wenn bei der Veröffentlichung Daten eines einzelnen Unternehmens von Dritten identifiziert werden können, dürfen diese nicht veröffentlicht werden. „Aktiv“ in diesem Zusammenhang bedeutet, dass einzelne Daten eines Unternehmens automatisch von den statistischen Ämtern geschützt werden, ohne dass das Unternehmen dies ausdrücklich angeben muss. Um die aktive Geheimhaltung zu gewährleisten, sind bei der Zusammenstellung der Ergebnisse zur Veröffentlichung bestimmte Geheimhaltungsregeln anzuwenden. Diese Geheimhaltungsregeln sind im Detail nicht gesetzlich vorgegeben. Vielmehr handelt es sich um Regeln innerhalb des Verbunds der Statistischen Ämter des Bundes und der Länder in Deutschland (Statistischer Verbund), die in anderen Mitgliedstaaten der Europäischen Union abweichen können.

In Bezug auf die nationale Veröffentlichung werden die Geheimhaltungsregeln für jedes Merkmal einer Meldenummer durchgeführt. Hierbei können bis zu drei Merkmale je Güterkategorie eine Rolle spielen:

1. Wert der Absatzproduktion,
2. Menge der Absatzproduktion,
3. Menge der Gesamtproduktion (Absatzproduktion zuzüglich der Produktion, die zur Weiterverarbeitung innerhalb des gleichen Unternehmens bestimmt ist).

Auf nationaler Ebene ist zwischen der primären und der sekundären Geheimhaltung zu differenzieren. Die primären Geheimhaltungsregeln identifizieren die kritischen Ergebnisse, die direkt einem einzelnen Unternehmen zugerechnet werden können. Hingegen gewährleistet die

sekundäre Geheimhaltung, dass primär gesperrte Felder nicht indirekt über Rückrechnung (in der Regel Differenzbildung) aufgedeckt werden können. Diese Prüfung ist notwendig, da neben den neunstelligen Meldenummern zusätzliche Aggregate veröffentlicht werden (zwei- und vierstellige Produktaggregate und Jahressummen).

Um die primäre Geheimhaltung zu gewährleisten, werden in der deutschen amtlichen Statistik die Fallzahlregel und die Dominanzregel angewendet. Ein Ergebnis – zum Beispiel der Absatzproduktionswert einer Meldenummer – darf nicht veröffentlicht werden, wenn weniger als drei Unternehmen (Fallzahlregel) oder ein oder zwei dominierende Unternehmen (Dominanzregel) hierzu Angaben in der Erhebung getätigt haben. Im Jahr 2001 haben die Statistischen Ämter des Bundes und der Länder beschlossen, dass die p%-Regel gegenüber anderen Regeln zur Identifizierung dominanter Unternehmen vorzuziehen ist (Gießing/Dittrich, 2006, hier: Seite 806). Die p%-Regel besagt: Ein Tabellenwert (X) ist geheimzuhalten, wenn er nach Abzug des zweitgrößten Einzelwertes (x_2) den größten Einzelwert (x_1) um weniger als p% übersteigt (Gießing/Dittrich, 2006, hier: Seite 806). Die Formel der Dominanzregel lautet:

$$(1) \quad X - x_2 < \left(1 + \frac{p}{100}\right) x_1$$

Trifft dies zu, wird der Produktionswert und/oder die Produktionsmenge dieser Meldenummer von ein oder zwei Unternehmen dominiert und darf demnach nicht veröffentlicht werden.

Die jährliche PRODCOM-Statistik unterliegt ebenfalls dem Prozess zur Überprüfung der Geheimhaltungsregeln. Nach der „Übersetzung“ der Einzeldaten von den nationalen auf die europäischen Güternummern finden die gleichen nationalen primären Geheimhaltungsregeln Anwendung. Eine sekundäre Geheimhaltungsprüfung ist nicht notwendig, da die PRODCOM-Statistik keine zusätzlichen Aggregate veröffentlicht, sondern lediglich Produktionsdaten auf Ebene der achtstelligen Güternummern der PRODCOM-Liste. Jedoch muss eine sogenannte Crossover-Geheimhaltung durchgeführt werden: Die jährlichen PRODCOM-Ergebnisse werden zu einem Zeitpunkt erstellt, zu dem die vierteljährlichen und jährlichen Ergebnisse der Produktionserhebung bereits veröffentlicht sind. Daher müssen alle PRODCOM-Ergebnisse verdeckt bleiben, die ein nationales Ergebnis offenlegen würden.

3

Schätzung des Produktionswerts unterhalb der Abschneidegrenze

3.1 Das Schätzmodell im Überblick

Aufgrund der Abschneidegrenze von 20 tätigen Personen kann das Gesamtbild der inländischen Produktion nicht vollständig erreicht werden. Auf der Grundlage von Verwaltungsdaten aus dem statistischen Unternehmensregister² kann – neben der Vollabdeckung – eine höhere Datenverfügbarkeit und Repräsentativität sowie eine bessere Kenntnis der Struktur der Industriestatistik erzielt werden. Hierfür muss der Produktionswert aller Unternehmen der WZ-Abschnitte B und C, die weniger als 20 tätige Personen aufweisen, anhand anderer für diese Unternehmen verfügbaren Merkmale aus den Verwaltungsdaten geschätzt werden.

Der entwickelte methodische Ansatz basiert auf einem Schätzprozess und besteht aus zwei Phasen:

1. Regressionsphase

Mithilfe der Daten aller in der Produktionserhebung befragten Unternehmen wird zunächst der Zusammenhang zwischen dem Produktionswert und anderen Merkmalen, wie dem Umsatz und der Anzahl der tätigen Personen, anhand eines Multilevel-Modells identifiziert.³ Die Angaben zum Produktionswert⁴ entstammen der Produktionserhebung. Die Merk-

male Umsatz⁵, Anzahl der tätigen Personen⁶ und vierstelliger Wirtschaftszweig stammen aus der Investitionserhebung⁷.

2. Schätzphase

Sobald die Beziehung zwischen den in Phase 1 genannten Variablen für Unternehmen mit mindestens 20 tätigen Personen mathematisch modelliert wurde, wird dieser Zusammenhang auf Unternehmen unterhalb der Abschneidegrenze anhand von Verwaltungsdaten übertragen. Diese Verwaltungsdaten entstammen dem statistischen Unternehmensregister (siehe auch Fußnote 2) und beziehen sich auf die Merkmale Umsatz, vierstellige Wirtschaftsklasse und Anzahl der sozialversicherungspflichtig Beschäftigten. Die Schätzung des Produktionswerts auf Ebene der vierstelligen Produktnummer basiert auf einem Simulationsansatz (siehe Abschnitt 3.2). Das Ergebnis der Schätzung wird auf Basis der Simulation mehrerer Modellparameter anhand des Medians und des dazugehörigen 95%-Schätzintervalls⁸ dargestellt.

➤ Grafik 2 stellt den Schätzprozess grafisch dar.

In der Schätzphase werden nur Unternehmen berücksichtigt, die im statistischen Unternehmensregister zu den Wirtschaftszweigen 05 bis 33 gehören und weniger als 20 sozialversicherungspflichtig Beschäftigte ausweisen. Beobachtungen ohne Angaben oder ohne Wert für

2 Das statistische Unternehmensregister enthält alle Unternehmen (rechtliche Einheiten), die entweder einen steuerbaren Jahresumsatz von mehr als 17 500 Euro erzielen oder über mindestens eine(n) sozialversicherungspflichtig Beschäftigte(n) verfügen. Das statistische Unternehmensregister speist sich aus Umsatzdaten der Finanzverwaltungen und Beschäftigtendaten der Bundesagentur für Arbeit. Es enthält für diese Unternehmen und deren Betriebe Angaben zu Hilfsmerkmalen (Name, Adresse), Ordnungsmerkmalen (Wirtschaftszweig, Rechtsform und andere) sowie Größe (Umsatz, Beschäftigte).

3 Die Regressionsphase erfolgt nach einem im Bericht des EU-Projekts beschriebenen Verfahren zur Ausreißerbereinigung. Der Prozess basiert auf dem Leverage- und Cook-Abstand-Konzept.

4 Der Wert der zum Absatz bestimmten Produktion (Produktionswert) wird unter Zugrundelegung des im Berichtszeitraum erzielten oder zum Zeitpunkt des Absatzes erzielbaren Verkaufspreises ab Werk berechnet (Verkaufswert). Der Verkaufswert enthält auch die Kosten der Verpackung, auch wenn sie gesondert in Rechnung gestellt werden. Steuern, Rabatte oder Transportkosten sind im Verkaufswert nicht enthalten (Statistisches Bundesamt, 2016a).

5 Der Umsatz ist der Wert der abgerechneten Lieferungen und Leistungen an Dritte ohne die in Rechnung gestellte Umsatzsteuer. Der Gesamtumsatz umfasst den Umsatz aus eigenen Erzeugnissen und Leistungen sowie den Umsatz aus Handelsware und sonstigen nicht industriellen, nicht handwerklichen Tätigkeiten (Statistisches Bundesamt, 2017b).

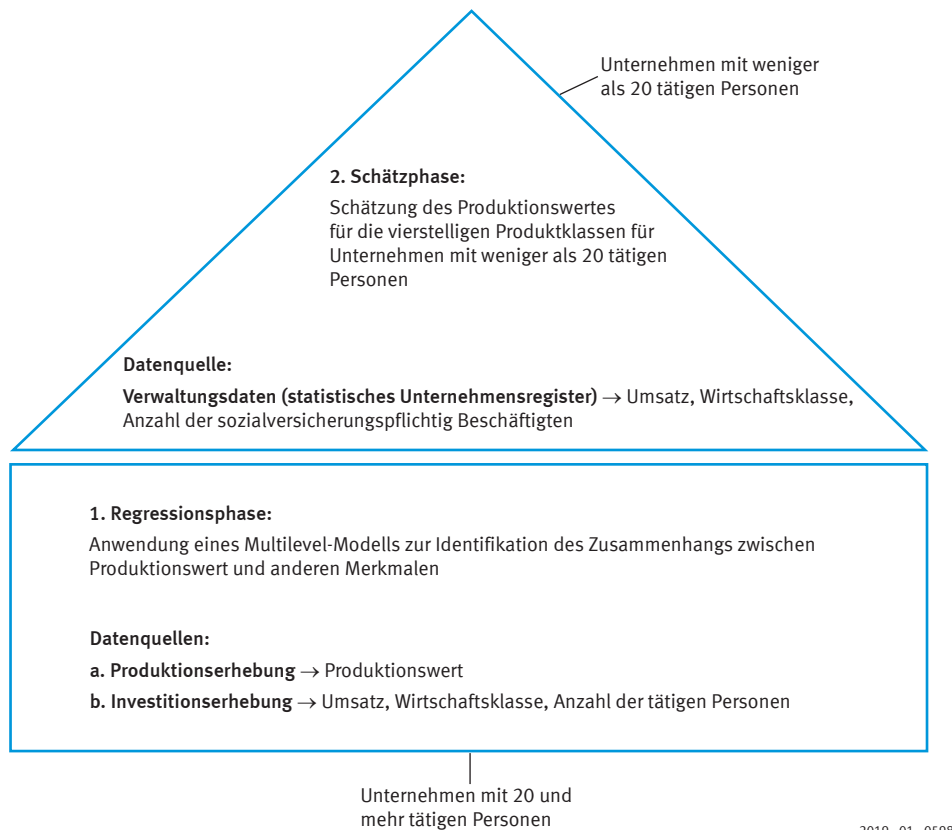
6 Die tätigen Personen umfassen alle im Unternehmen oder Betrieb tätigen Personen einschließlich der tätigen Inhaberinnen/Inhaber, mithelfenden Familienangehörigen, an andere Unternehmen überlassenen Mitarbeiterinnen/Mitarbeiter und Heimarbeiterinnen/Heimarbeiter, jedoch nicht die Leiharbeiterinnen/Leiharbeiter. Die tätigen Personen umfassen auch die kaufmännischen Auszubildenden und die gewerblich Auszubildenden (Statistisches Bundesamt, 2017b).

7 Die Investitionserhebung liefert grundsätzlich Informationen zu Investitionen nach Arten, gemieteten und gepachteten Sachanlagen und Verkaufserlösen aus dem Abgang von Sachanlagen (Statistisches Bundesamt, 2017a). Die Investitionserhebung ist für die Analyse des Zusammenhangs in der Regressionsphase gut geeignet, da beide Erhebungen exakt den gleichen Berichtskreis haben und somit gut miteinander zu verknüpfen sind.

8 Das Konzept der Schätzintervalle darf nicht mit Konfidenzintervallen verwechselt werden. Während ein Konfidenzintervall einen Bereich für $E(y|x)$ anzeigt, gibt ein Schätzintervall einen Bereich für $y|x$ selbst an. Die Berücksichtigung der Standardabweichung ist von Bedeutung, da dadurch die Unsicherheit (Fehlerterm) in die Schätzung mit einbezogen werden kann. Somit ist ein Schätzintervall immer breiter als ein Konfidenzintervall.

Grafik 2

Schätzmodell zur Schätzung des Produktionswerts



Umsatz oder Anzahl an sozialversicherungspflichtig Beschäftigten werden nicht berücksichtigt. Zur Vermeidung von Doppelzählungen werden nur Unternehmen berücksichtigt, die alle vorherigen Bedingungen erfüllen und nicht zur nationalen Produktionserhebung gehören. Die bereinigten Daten enthalten 130 745 Beobachtungen von Unternehmen, für die der Produktionswert geschätzt wird.

Normalerweise wird der Produktionswert auf der Ebene der neunstelligen Meldenummer des Güterverzeichnis ausgewiesen. Da keinerlei Kenntnisse über die Art der hergestellten Güter für bestimmte Unternehmen vorhanden sind, basiert die Schätzung des Produktionswerts auf den vierstelligen Produktklassen. Die vierstellige Produktklasse ist eine gröbere Hierarchieebene innerhalb der Güterklassifikation. Im Schätzmodell wird diese Produktklasse direkt aus dem vierstelligen Wirtschaftszweig des Unternehmens abgeleitet. Darüber hinaus ist eine Abschätzung von Produktionsmengen nicht möglich, da auf dieser Hierarchieebene keine ein-

heitliche Maßeinheit aller darin enthaltenen Produktarten existiert.

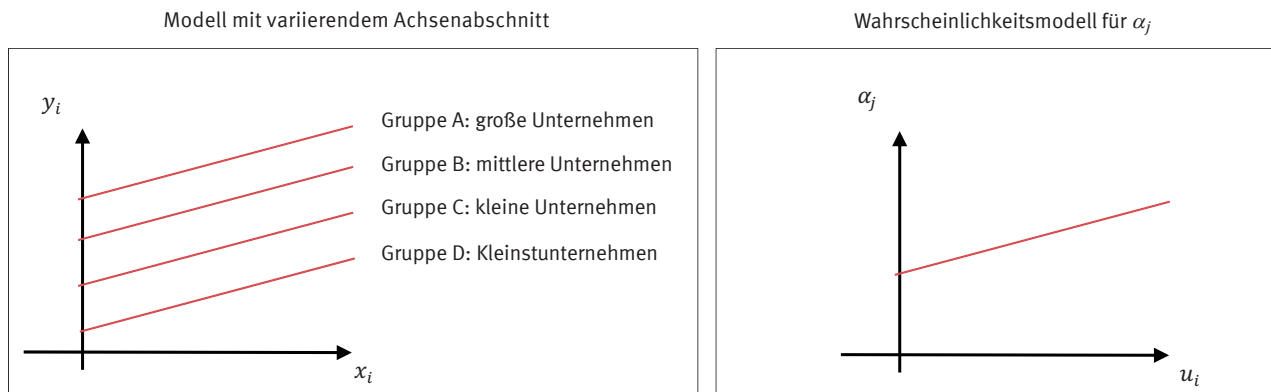
3.2 Das Multilevel-Modell

Multilevel-Modelle (auch als hierarchisch lineare Modelle oder Mehrebenenmodelle bezeichnet) sind statistische Verfahren, die insbesondere für Daten geeignet sind, die hierarchisch in mehreren Gruppen oder Ebenen strukturiert sind.

Es gibt mehrere Gründe, warum dieses Modell für die Schätzung des Produktionswerts verwendet wird. Zum einen ermöglicht es die Untersuchung von Zusammenhängen, die zwischen Gruppen variieren. Für die Schätzung wird die Annahme getroffen, dass der Zusammenhang zwischen den untersuchten Merkmalen von der Unternehmensgröße abhängt. Zum anderen ist die Schätzung von Merkmalen von Einheiten neuer, noch unbekannter Gruppen innerhalb dieses Modellansatzes

Grafik 3

Varying-intercept-Modell



2019 - 01 - 0599

möglich. Letzteres ist in der klassischen linearen Regression schwierig, da sie dazu neigt, den Fehler bei den Vorhersagen für neue Gruppen zu unterschätzen (für weitere Einzelheiten siehe Gelman/Hill, 2007). Die unbekannte Gruppe, für die eine Schätzung des Produktionswerts vorgenommen werden soll, sind die Kleinstunternehmen⁹.

Das vorgeschlagene Modell zur Analyse der gruppierten Daten ist ein Varying-intercept-Modell, bei dem der Achsenabschnitt α gruppenspezifisch ist und der Steigungsparameter β über die Gruppen hinweg gleich ist. [↪ Grafik 3](#)

Der erste Schritt der Multilevel-Modellierung besteht darin, eine Regression mit unterschiedlichen gruppenspezifischen Achsenabschnitten durchzuführen (linker Teil der Grafik 3). Im zweiten Schritt wird diesem Achsenabschnitt ein Wahrscheinlichkeitsmodell zugrunde gelegt (rechter Teil der Grafik 3). Das ist das entschei-

dende Modellierungsmerkmal des Multilevel-Modells. Eine Variante, das Multilevel-Modell darzustellen, ist die Verknüpfung lokaler Regressionen in jeder Gruppe. Innerhalb jeder Gruppe j wird eine Regression mit der erklärenden Variable x_i auf Unternehmensebene und mit dem variierenden Achsenabschnittsparameter α durchgeführt, der nach Gruppen indiziert ist. Der Achsenabschnitt erhält dann ein Wahrscheinlichkeitsmodell. Somit kann das Modell durch die folgenden Gleichungen vollständig dargestellt werden:

- (2) für jede Gruppe j : $y_i \sim N(\alpha_{j(i)} + \beta x_i, \sigma^2_y)$
für Unternehmen $i = 1, \dots, n$
- (3) $\alpha_j \sim N(\gamma_0 + \gamma_1 u_j, \sigma^2_\alpha)$
für die Größe $j = \text{Klein, Mittel, Groß}$

Bei der Gleichung (2) stellt x_i die erklärende Variable (Regressor) auf Unternehmensebene dar und bezeichnet den logarithmischen Umsatz in 1 000 Euro. Für den gruppenweise variierenden Achsenabschnitt gilt $\alpha_{j(i)}$. Die Indizes $j(i)$ beschreiben die Größe j des Unternehmens i . Darüber hinaus stellt β den Steigungsparameter auf der Unternehmensebene und σ^2_y den ungeklärten Fehler des gesamten Modells dar. Die erste Regressionsgleichung repräsentiert die Unternehmensebene des Modells. Die Gleichung (3) stellt ein Regressionsmodell für den variierenden Parameter α_j dar. Es enthält die erklärende Variable u_j auf der Gruppenebene (Anzahl der sozialversicherungspflichtig Beschäftigten in jeder Gruppe) mit den Parametern γ_0 und γ_1 . Darüber hinaus ist σ^2_α die Varianz der ungeklärten Fehler auf Gruppenebene.

⁹ Die Bestimmung der Unternehmensgrößenklassen erfolgt weitgehend anhand der Empfehlung der Europäischen Kommission vom 6. Mai 2003 betreffend die Definition der Kleinstunternehmen sowie der kleinen und mittleren Unternehmen (Amtsblatt der EU Nr. L 124, Seite 36) (2003/2361/EG). Demnach haben große Unternehmen 250 oder mehr Beschäftigte und mittlere Unternehmen 50 oder mehr Beschäftigte. Bei der Abgrenzung zwischen kleinen Unternehmen und Kleinstunternehmen weicht die hier getätigte Definition etwas ab. Die EU-Empfehlung definiert Kleinstunternehmen mit weniger als 10 Beschäftigten. Im Rahmen der Untersuchungen zur Schätzung des Produktionswerts werden Kleinstunternehmen jedoch als Unternehmen mit weniger als 20 tätigen Personen definiert, da diese Grenze zur Definition des Berichtskreises in der Erhebung festgesetzt ist. Demnach sind kleine Unternehmen hier als Unternehmen mit 20 oder mehr und weniger als 50 Beschäftigten definiert.

3.3 Der Simulationsansatz

Die Schätzung des Produktionswerts der unbekanntenen Gruppe erfolgt aus einem Simulationsansatz wie von Gelman und Hill (2007) erläutert. Mit Simulation ist gemeint, dass Schlussfolgerungen durch Zufallszahlen und nicht durch Punktschätzungen und Standardfehler dargestellt werden. Zu diesem Zweck wird das in den Gleichungen (2) und (3) beschriebene Modell geschätzt, wobei die Schätzungen $\hat{\alpha}_j, \hat{\beta}, \hat{\sigma}_y^2, \hat{\sigma}_\alpha^2, \hat{\gamma}_0$ und $\hat{\gamma}_1$ enthalten sind. Auf intuitive Weise simuliert der Simulationsprozess α_{ss} (Achsenabschnitt der Kleinstunternehmen) β und σ_y^2 unter Angabe von x_i auf der Basis der geschätzten Parameter. Somit ist es möglich, so viele Simulationen für $y_i|x_i$ zu erhalten, wie Zufallsziehungen ausgeführt werden. Nach einer vernünftig großen Anzahl an Ziehungen wird eine Punktschätzung für $y_i|x_i$ zusammen mit Schätzintervallen unter Verwendung der 50 %- , 2,5 %- und 97,5 %-Quantile über alle simulierten Werte für $y_i|x_i$ durchgeführt. Diese Vorgehensweise ist im Exkurs „Simulationsansatz für Multilevel-Modelle“ beschrieben.

Simulationsansatz für Multilevel-Modelle

1. Schätzung des in den Gleichungen (2) und (3) definierten Multilevel-Modells und Berechnung von $\hat{\alpha}_{ss}$ [das durch $\hat{\gamma}_0$ und $\hat{\gamma}_1$ in Gleichung (3) bestimmt wird], $\hat{\beta}, \hat{V}$ (die unskalierte Kovarianzmatrix der Koeffizienten mit den Komponenten $\hat{\sigma}_\alpha$ und $\hat{\sigma}_\beta$), und die Störvarianz $\hat{\sigma}_y^2$ unter Verwendung von n Datenpunkten auf Prädiktoren.
2. Zufallssimulation für α_{ss}, β und σ_y . Für jede Simulationsziehung:
 - a) Simulierung von $\sigma_y = \hat{\sigma}_y \sqrt{(n-k)/x}$, wobei x eine zufällige Ziehung aus der X^2 -Verteilung mit $n-k$ Freiheitsgraden ist.
 - b) Simulierung von $B = (\alpha_{ss}, \beta)$ bei der Zufallsziehung von σ_y , aus einer multivariaten Normalverteilung mit dem mittleren Wert $\hat{B} = (\hat{\alpha}_{ss}, \hat{\beta})$ mit $\hat{\alpha}_{ss} = \hat{\gamma}_0 + \hat{\gamma}_1 \hat{u}_{ss}$ wie in der Gleichung (3) dargestellt ist und der Varianzmatrix $\sigma_y^2 \hat{V}$. \hat{u}_{ss} gilt für die durchschnittliche Anzahl der sozialversicherungspflichtig Beschäftigten bei den Kleinstunternehmen.
3. Zufallsziehung von $Y_{sim,j}$ aus $N(X_{ss}B, \sigma_y^2)$, wobei j für jede Simulation steht und X_{ss} die Umsatzdaten der Kleinstunternehmen aus dem Unternehmensregister beinhaltet.

4. Wiederholung von Schritten 1 bis 3 N_{sim} -mal. Für diese Anwendung ist es ausreichend, $N_{sim} = 3000$ festzulegen.
5. Um die Punktschätzung zusammen mit den Schätzintervallen des Produktionswerts der kleinsten Unternehmen zu erhalten, müssen die 50 %- , 2,5 %- und 97,5 %-Quantile über $Y_{sim,1} \dots Y_{sim,3000}$ berechnet werden.

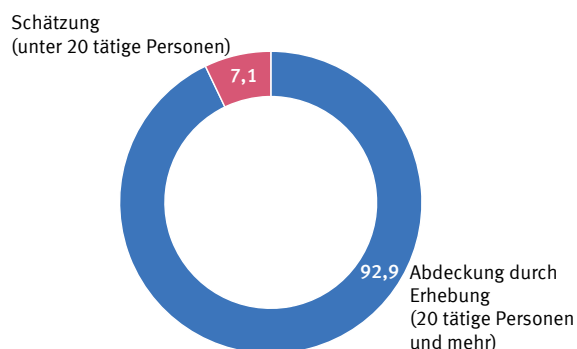
Der Simulationsansatz ist nützlich, um sowohl die Schätzgenauigkeit (gemessen am Fehlerterm im Regressionsmodell) als auch die Inferenzunsicherheit (gemessen an den Standardfehlern der Parameter und der Unsicherheit über den Restfehler) in die Ergebnisdarstellung mit einzubeziehen. Dieses Verfahren basiert auf einer informellen Bayesianischen Inferenz¹⁰, die sich auf statistische Methoden bezieht und die unbekannte Parameter als Zufallsvariablen modelliert.

3.4 Empirische Ergebnisse

Mit dem Schätzverfahren kann der Gesamtproduktionswert aller Unternehmen der Wirtschaftsabschnitte B und C berechnet werden. Rund 7 % der gesamten Industrieproduktion in Deutschland wird durch das entwickelte Schätzverfahren im Berichtsjahr 2016 abgedeckt. Dies

¹⁰ Bayes-Statistik ist ein Zweig der Statistik, der mit dem bayesschen Wahrscheinlichkeitsbegriff und dem Satz von Bayes Fragestellungen der Stochastik untersucht. Charakteristisch für die bayessche Statistik ist die konsequente Verwendung von Wahrscheinlichkeitsverteilungen beziehungsweise Randverteilungen, deren Form die Genauigkeit der Verfahren beziehungsweise Verlässlichkeit der Daten und des Verfahrens transportiert.

Grafik 4
Abdeckung und Schätzung des inländischen Gesamtproduktionswerts für Berichtsjahr 2016 in %



2019 - 01 - 0600

entspricht dem Anteil der Produktion der Kleinstunternehmen. Entsprechend repräsentiert die aktuelle Erhebung und Veröffentlichung bereits etwa 93 % der inländischen Produktion. [↘ Grafik 4](#)

Durch dieses Schätzverfahren wird eine vollständige Abdeckung erreicht, ohne die Kleinstunternehmen durch eine Erhebung zu belasten. Im Rahmen des EU-Projekts wurde der dadurch vermiedene Aufwand der Unternehmen berechnet,¹¹. Die Belastung, die durch eine vollständige Erhebung der Kleinstunternehmen entstehen würde, würde demnach 22,2 Millionen Euro jährlich betragen.

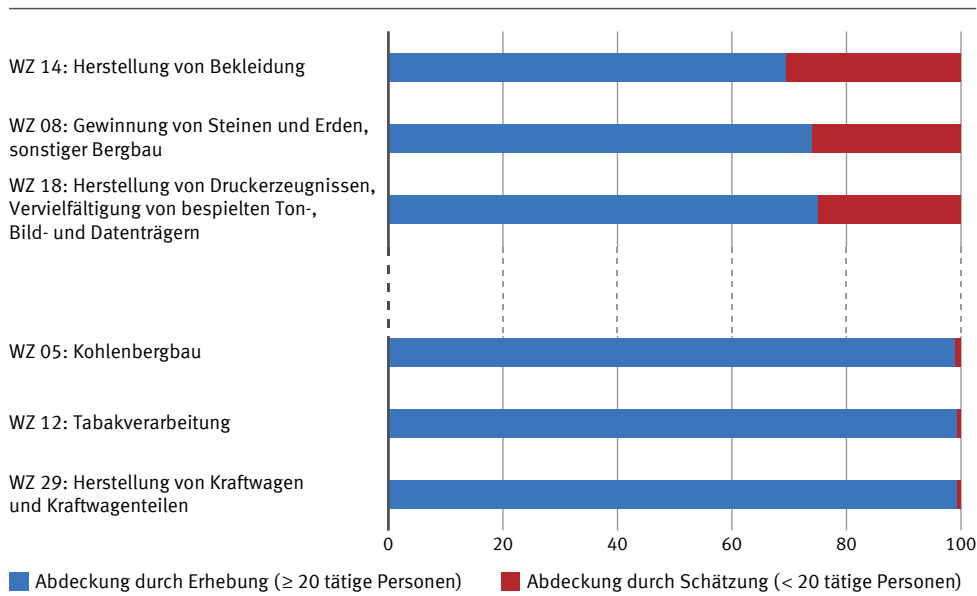
11 In Deutschland hat die Regierung das Arbeitsprogramm Bessere Rechtsetzung und Bürokratieabbau aufgelegt. Das Statistische Bundesamt hat im Rahmen dieses Programms den monetären Aufwand der Wirtschaft für alle Gesetze des Bundes gemessen. Dies schließt auch den Aufwand aus den Informationspflichten der amtlichen Statistik ein. Die Ergebnisse dieser Messungen sind über eine Online-Datenbank zu Bürokratiekosten öffentlich zugänglich. Die darin aufgeführte Belastung der Unternehmen der Produktionserhebung wurde auf die Anzahl der Kleinstunternehmen übertragen, um die Bürokratiekosten für diese Unternehmen im Falle einer Erhebung abschätzen zu können.

[↘ Grafik 5](#) zeigt die so geschätzten Anteile des Produktionswerts der Kleinstunternehmen (rot) gegliedert nach zweistelligen Wirtschaftsabteilungen. Dargestellt sind die jeweils drei Wirtschaftsabteilungen, in denen die Kleinstunternehmen den größten beziehungsweise kleinsten Anteil am Produktionswert der gesamten Wirtschaftsabteilung haben. Es zeigt sich, dass diese Anteile zwischen den Wirtschaftsabteilungen stark variieren. Beispielsweise haben die Kleinstunternehmen in der Wirtschaftsabteilung 14 „Herstellung von Bekleidung“ einen Anteil von 30 % an der Gesamtproduktion dieser Abteilung. Hingegen gibt es in der Abteilung 29 „Herstellung von Kraftwagen und Kraftwagenteilen“ kaum Kleinstunternehmen, die in diesem Bereich Güter produzieren. Der Anteil beträgt hier lediglich 0,6 % an der Gesamtproduktion dieser Abteilung.

[↘ Grafik 6](#) bekräftigt die Annahme bei der Regressionsanalyse, dass der Zusammenhang zwischen den untersuchten Merkmalen von der Unternehmensgröße abhängt. In dieser Grafik ist die beobachtete Struktur der untersuchten Unternehmensdaten dargestellt; sie enthält die empirischen Dichtefunktionen des logarith-

Grafik 5

Abdeckung und Schätzung des inländischen Gesamtproduktionswerts nach Wirtschaftsabteilungen für Berichtsjahr 2016 in %

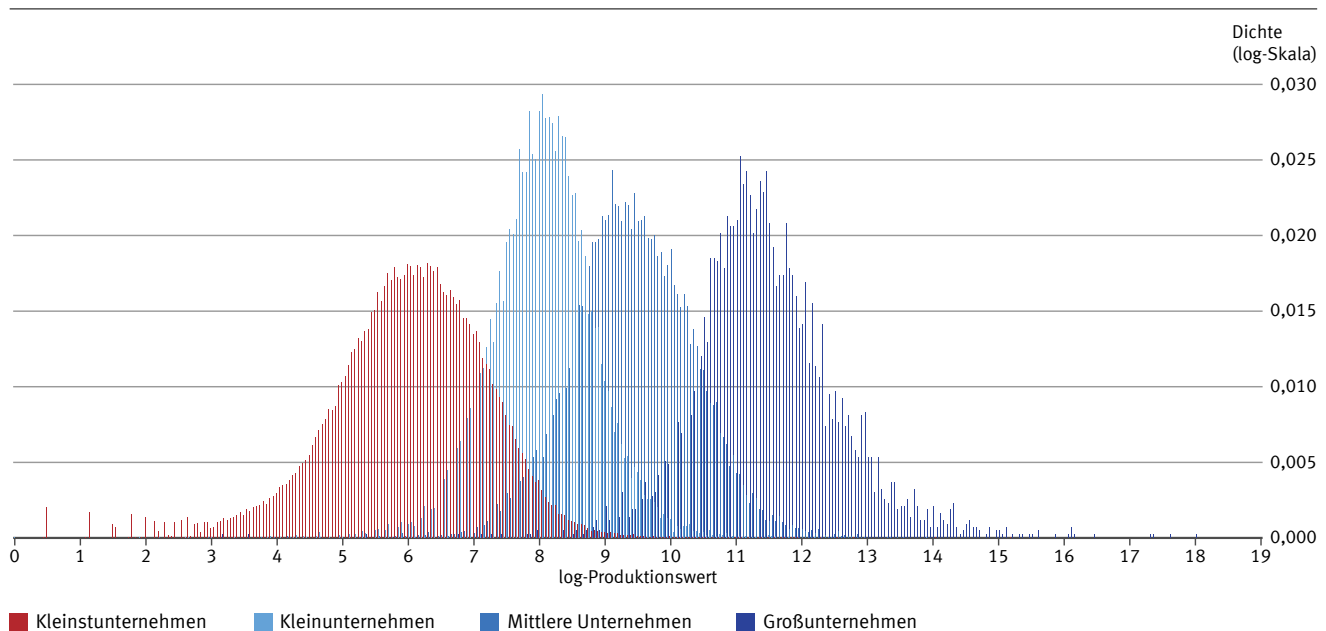


Dargestellt sind die jeweils 3 Wirtschaftsabteilungen der Klassifikation der Wirtschaftszweige, Ausgabe 2008 (WZ 2008) mit den größten beziehungsweise geringsten Schätzanteilen.

2019 - 01 - 0601

Grafik 6

Empirische Dichtefunktion des Produktionswerts nach Unternehmensgrößenklassen



2019 - 01 - 0602

mischen Produktionswerts¹² der Kleinstunternehmen und der drei Unternehmensgrößenklassen der erhobenen Unternehmen: große, mittlere und kleine Unternehmen (siehe Fußnote 9).

Schließlich zeigt [Grafik 7](#) die Schätzung des Produktionswerts bei gegebenem Umsatz zusammen mit den Schätzintervallen (Unter- und Obergrenzen). Mit diesen Intervallen kann festgestellt werden, dass der Produktionswert bei gegebenem Umsatz mit einer bestimmten Wahrscheinlichkeit zwischen den genannten Grenzwerten liegt. Beispielsweise liegt der Produktionswert bei einem Umsatzwert von 1 578 000 mit einer Wahrscheinlichkeit von 95 % zwischen 708 000 und 3 392 000 Euro.

3.5 Bewertung des Multilevel-Modells

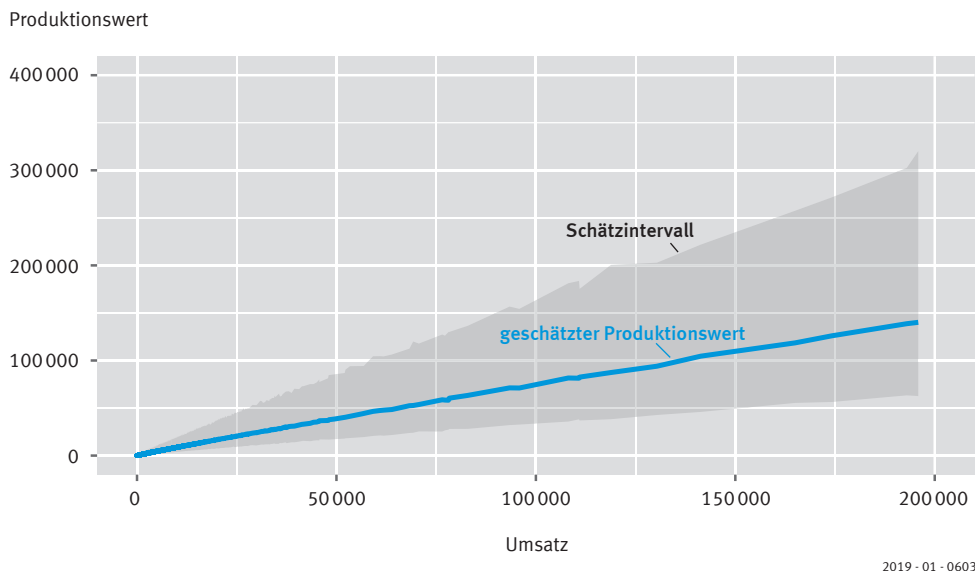
Die traditionellen Alternativen zum Multilevel-Modell sind das Ignorieren von Gruppenunterschieden (sogenannte Pooling-Modelle) oder die separate Analyse der

jeweiligen Gruppen (No-pooling-Modelle) im Rahmen der einfachen linearen Regression. Beide Varianten wurden ebenso anhand der vorhandenen Daten durchgeführt und getestet, um die Modelle vergleichen zu können. Beide Alternativen weisen jedoch Nachteile gegenüber dem Multilevel-Modell auf. Die Pooling-Analyse ignoriert die Gruppenstruktur der Unternehmensdaten und dadurch werden Abweichungen im Durchschnitt der Produktion zwischen den Gruppen nicht beachtet. Das Ziel besteht aber genau darin, die Gruppeneffekte zu berücksichtigen und in die Analyse einfließen zu lassen, um den bedingten Produktionswert der unbekanntenen Gruppe der kleinsten Unternehmen $y_i|x_i$ zu schätzen. Bei der No-pooling-Analyse gibt es zwei Modellierungsmöglichkeiten. Eine erste Alternative kann Größenindikatoren beziehungsweise Dummy Variablen in die Regression einbeziehen, um die Unterschiede zwischen den Gruppen zu berücksichtigen. Dieses Modell weist ebenfalls Probleme auf, da die Schätzung der Parameter nicht die unterschiedliche Anzahl von Beobachtungen zwischen den Gruppen berücksichtigt. Das bedeutet, die Schätzung für jede Gruppe wird nicht mit der Menge der in jeder Gruppe enthaltenen Beobachtungen gewichtet. Eine zweite Alternative beinhaltet ein Regressionsmodell basierend auf den Daten der Kleinunternehmen, welches dann eins zu eins auf die Kleinstunternehmen übertra-

12 Die logarithmische Transformation einer Variablen erzeugt symmetrische Daten und erlaubt den Vergleich verschiedener Gruppen von Beobachtungen, sodass die Variabilität in jeder Gruppe ungefähr gleich bleibt. Diese zwei Eigenschaften sind im Rahmen der Regressionsanalyse wünschenswert und in der praktischen Anwendung weit verbreitet.

Grafik 7

Schätzung des Produktionswerts 1 000 EUR



gen wird. Die offensichtliche Einschränkung dieser zweiten Alternative besteht im Informationsverlust aufgrund der Missachtung der mittleren und großen Unternehmen bei der Regressionsanalyse.

Trotz der theoretischen Vorteile des Multilevel-Modells wurde die Qualität der Ergebnisse mit denen der einfachen linearen Regressionsmethoden (Pooling- und No-pooling-Modelle) verglichen. Anhand des sogenannten Diebold-Mariano-Tests (Diebold/Mariano, 1995) wird die Schätzung des Produktionswerts der verschiedenen Modelle mit dem Produktionswert der erhobenen Unternehmen verglichen. Auf Basis einer Verlustfunktion überprüft der Test, ob der mittlere quadratische Fehler (Root Mean Square Error – RMSE¹³) der alternativen Modelle signifikant höher ist als der RMSE des Multilevel-Modells (weitere Einzelheiten siehe Diebold/Mariano, 1995). Das Ergebnis des Tests zeigt, dass sich bei der Verwendung des Multilevel-Modells ein offensichtlicher Gewinn in Bezug auf die Genauigkeit der Schätzung ergibt. Es bestätigt die beschriebenen Vorteile und die Eignung des Multilevel-Modells in Bezug auf die Schätzung des Produktionswerts für Unternehmen unterhalb der Abschneidegrenze von 20 tätigen Personen.

13 Der RMSE ist ein Maß für die Abweichung der Schätzungen vom tatsächlichen Wert (berechnet als Quadratwurzel der durchschnittlichen Quadratabweichung).

4

Reduzierung der Geheimhaltungsfälle in der Produktionsstatistik

4.1 Änderung des p%-Werts bei der Dominanzregel

Die Bedeutung des p%-Werts bei der Anwendung der Dominanzregel wurde bereits in Kapitel 2 erläutert. Ziel der folgenden Abschnitte ist es, die Auswirkungen eines niedrigeren p%-Werts auf die Ergebnisse der PRODCOM-Statistik zu analysieren und abzuwägen. Allgemein bewirkt ein höherer p%-Wert mehr Sperrungen von Ergebnissen als ein niedrigerer. Die Höhe des p%-Werts bedeutet jedoch immer eine Abwägung (Trade-off) zwischen dem Datenschutz und dem Nutzerinteresse.

Wie in Kapitel 2 beschrieben ist die Geheimhaltung von Ergebnissen der PRODCOM-Statistik in Deutschland abhängig von der Geheimhaltung der Ergebnisse in der nationalen vierteljährlichen und jährlichen Veröffentlichung (Crossover-Geheimhaltung). Daher ist zunächst der p%-Wert für die Erstellung der nationalen Veröffentlichung zu betrachten. Für die Festlegung des p%-Werts bei

der Erstellung der nationalen Produktionsstatistik gibt es bestimmte Einschränkungen, die beachtet werden müssen. Die Rahmenbedingungen in Deutschland sind zum einen, dass die Produktionserhebung eine eigene nationale Güterklassifikation hat, auf deren Basis Ergebnisse veröffentlicht werden. Zum anderen gibt es innerhalb der nationalen Veröffentlichung mehrere Aggregationsebenen (Aggregation von Berichtszeiträumen und Meldenummern der Güterklassifikation), die sekundäre Geheimhaltungsprüfungen erfordern. Durch die Crossover-Geheimhaltung werden die gesperrten Positionen auf die PRODCOM-Statistik übertragen, auch wenn diese selbst keine zusätzlichen Aggregationsebenen umfasst (und demnach keine sekundäre Geheimhaltung notwendig wäre).

Zudem ist eine Änderung des p%-Werts in der deutschen amtlichen Statistik nicht ohne Weiteres möglich. Um den ausreichenden Schutz von Einzelangaben zu gewährleisten, sollten die Geheimhaltungsregeln stets für alle Veröffentlichungen und über alle Anwendungsbereiche des Statistischen Verbunds hinweg konsistent umgesetzt werden (Rohde und andere, 2018, hier: Seite 93). Das heißt, dass die Ergebnisse auf Länderebene ebenso eine Rolle spielen und den gleichen p%-Wert verwenden. Daher ist die Höhe des p%-Werts eine Vereinbarung der Statistischen Ämter des Bundes und der Länder, die den hohen Stellenwert des Datenschutzes und das Vertrauen in die amtliche Statistik in Deutschland mit einbezieht. Deshalb kann eine Änderung des p%-Werts nur in Absprache im Statistischen Verbund erfolgen. Zudem ist die Produktionsstatistik in ein System mehrerer Unternehmensstatistiken eingebettet. Beispielsweise haben der Monatsbericht oder auch die Kostenstrukturerhebung im Verarbeitenden Gewerbe den gleichen p%-Wert wie die Produktionsstatistik. Alle diese Statistiken weisen ähnliche Berichtskreise auf und veröffentlichen ähnliche oder miteinander korrelierende Merkmale.

Für die Berechnung der Anzahl der Ergebnisse, die zusätzlich offengelegt werden können, wird daher lediglich der p%-Wert zur Erstellung der PRODCOM-Statistik reduziert – jedoch nicht der p%-Wert der nationalen Produktionsstatistik. Aufgrund der Crossover-Geheimhaltung werden die meisten geheimzuhaltenden Ergebnisse von der nationalen Veröffentlichung auf die PRODCOM-Ergebnisse übertragen. Daher ist der Effekt eines niedrigeren p%-Werts nur sehr gering. Im Rahmen des Eurostat-Projekts wurden noch andere Varianten getestet, die unterschiedliche Einschränkungen voraussetzen. Mit diesem Vorgehen soll gezeigt werden, welchen Effekt die Senkung des p%-Werts hätte, wenn die nationale Produktionsstatistik den genannten Einschränkungen nicht unterliegen würde.

Ergebnisse für die Reduzierung des p%-Werts

Für die Reduzierung des p%-Werts wurden zwei niedrigere p%-Werte getestet und die Anzahl der gesperrten Ergebnisse mit dem aktuellen Wert verglichen. Der Referenzwert ist daher der aktuelle p%-Wert und im Folgenden mit „p%-Wert hoch“ bezeichnet. Aufgrund gegebener Rahmenbedingungen gilt für Deutschland die in [Tabelle 1](#) dargestellte Variante 2. Ihr werden noch zwei andere Varianten zum Vergleich gegenübergestellt.

Die in Tabelle 1 angegebenen Prozentwerte beziehen sich auf insgesamt 8 055 Ergebnisse. Diese Zahl berücksichtigt, dass es für die 3 833 PRODCOM-Nummern bis zu drei Ergebnisse je Güternummer geben kann: Absatzproduktionswert, Absatzproduktionsmenge und Gesamtmenge.¹⁴ Die Anzahl zusätzlicher Ergebnisse

14 Nicht alle PRODCOM-Nummern schließen alle drei genannten Merkmale ein. Bei manchen Güterarten der PRODCOM-Statistik gibt es beispielsweise keine Gesamtmenge, es wird lediglich die Absatzproduktion in Menge und Wert erhoben und veröffentlicht. Bei anderen Güterarten hingegen sind überhaupt keine Mengenangaben vorgesehen.

Tabelle 1
Varianten zur Änderung des p%-Werts im Berichtsjahr 2016

	Crossover-Geheimhaltung aufgrund einer nationalen Veröffentlichung mit anderer Klassifikation und/oder zusätzlichen Aggregationsstufen	Einschränkung zur Änderung des p%-Werts	Anzahl und Anteile der gesperrten PRODCOM-Ergebnisse gemessen an insgesamt 8 055 Ergebnissen					
			p%-Wert					
			hoch		mittel		niedrig	
			Anzahl	%	Anzahl	%	Anzahl	%
Variante 1	nein	nein	1 955	24	1 764	22	1 522	19
Variante 2	ja	ja	2 458	31	2 454	30	2 440	30
Variante 3	ja	nein	2 458	31	2 225	28	1 969	24

ist bei Variante 2 marginal. Selbst bei einem niedrigen p%-Wert können nur 18 Ergebnisse zusätzlich veröffentlicht werden. Dies entspricht einer Erhöhung der offenen Ergebnisse um lediglich 0,2%. Grund für den nur sehr geringen Nutzen ist die Übernahme der Geheimhaltung aus der nationalen Veröffentlichung, bei der der p%-Wert unverändert bleibt. Eine zusätzliche Offenlegung eines Ergebnisses – wenn überhaupt – kann nur bei PRODCOM-Güterarten erfolgen, bei denen GP-Meldenummer und PRODCOM-Nummer inhaltlich nicht exakt identisch sind (keine „1:1“-Güternummer).

Hingegen könnten bei Variante 1 bei einem niedrigen p%-Wert 5% mehr Ergebnisse veröffentlicht werden (Rückgang von 1955 auf 1522 gesperrte Ergebnisse). Variante 1 würde zutreffen, wenn es keinerlei Einschränkungen aufgrund einer eigenen nationalen Veröffentlichung und aufgrund unveränderbarer nationaler Geheimhaltungsregeln gäbe. Hier würden den Nutzerinnen und Nutzern mehr als 400 zusätzliche Ergebnisse zur Verfügung stehen. Variante 3 kann aufgrund der flexiblen Anpassung des p%-Werts bereits bei der Erstellung der nationalen Veröffentlichung eine geringere Anzahl an Geheimhaltungspositionen aufweisen. Dies wirkt sich insgesamt positiv auf die Anzahl an zusätzlichen Ergebnissen aus: Hier könnten 6% mehr Ergebnisse bei einem niedrigen p%-Wert für die Nutzerinnen und Nutzer zugänglich sein (Rückgang von 2458 auf 1969 gesperrte Ergebnisse). Auch wenn es die Einschränkungen bei der nationalen Produktionserhebung und der Ergebniserstellung nicht gäbe und hierbei durchaus eine lohnenswerte Anzahl an zusätzlichen Ergebnissen verfügbar wäre, bleiben die Risiken – nämlich die Gefahr der Aufdeckung von Einzelergebnissen – bei einem niedrigeren p%-Wert bestehen. Das heißt auch bei den Varianten 1 und 3 wäre eine Senkung des p%-Werts zu überdenken und abzuwägen.

4.2 Unternehmensumfrage im Rahmen der aktiven Geheimhaltungspraxis

Wie bereits beschrieben, ist in Deutschland die aktive Geheimhaltung bei der Bereitstellung von Ergebnissen gesetzlich festgelegt. Die Veröffentlichung von Einzeldaten ist nach §16 Absatz 1 Bundesstatistikgesetz jedoch dann erlaubt, wenn das betroffene Unternehmen schriftlich dazu eingewilligt hat. Mithilfe einer Unternehmensumfrage sollte im Rahmen des EU-Projekts getestet werden, ob die Unternehmen bereit wären, in die Veröffentlichung ihrer eigentlich geheimzuhaltenden Daten einzuwilligen. Acht Statistische Landesämter beteiligten sich an der Durchführung der Umfrage. Da eine schriftliche Einwilligung gesetzlich vorgeschrieben ist (§16 Absatz 1 Bundesstatistikgesetz), erfolgte die Umfrage postalisch mit einer beigefügten, vorgefertigten Einwilligungserklärung.

Bei der Unternehmensumfrage wurden nur diejenigen Unternehmen kontaktiert, deren Daten aktiv geheim gehalten werden. Dies kann zum einen zutreffen, wenn ein Unternehmen das einzige oder eines von zwei Unternehmen in Deutschland ist, das eine bestimmte Güterart herstellt – hier greift die Fallzahlregel. Zum anderen kann die Geheimhaltung nach der Dominanzregel zutreffen: Das Unternehmen ist ein dominierendes oder eines von zwei dominierenden Unternehmen bei der Produktion einer bestimmten Güterart. Um zu vermeiden, dass Unternehmen angeschrieben werden, die nur

Tabelle 2

Analyse der von Geheimhaltungsregeln betroffenen Unternehmen in der Monatlichen und Vierteljährlichen Produktionsstatistik für das Berichtsjahr 2017

	Unternehmen	
	Anzahl	%
Alle GP-Nummern mit einer Produktion > 0 ¹	4 906	100
darunter: dauerhaft gesperrte GP-Nummern ²	668	13,6
Gesamtzahl der Unternehmen in der Produktionserhebung ³	39 361	100
darunter: Anzahl der von Geheimhaltung betroffenen Unternehmen ⁴	1 152	2,9
darunter: Anzahl der Unternehmen in der Umfrage ⁵	550	1,4

GP: Güterverzeichnis für Produktionsstatistiken, Ausgabe 2009.

1 Enthalten sind alle Meldenummern (GP 2009), für die im Berichtsjahr 2017 mindestens ein Unternehmen Produktionsdaten gemeldet hat. Meldenummern ohne Produktion werden nicht zum 100%-Referenzwert gezählt.

2 „Dauerhaft“ heißt, dass die Meldenummer in mindestens 6 von 8 Quartalen aufgrund primärer Geheimhaltungsregeln (Fallzahl- und Dominanzregel) gesperrt war. Die Auswertung bezieht sich auf das GP 2009. Für die Anschreiben wurden die GP-Nummern ins GP 2019 „übersetzt“. Durch eine GP-Umstellung können sich durch Aufsplittungen oder Zusammenführungen Geheimhaltungsfälle neu bilden oder aufheben. Durch die erst kürzlich stattgefundene Umstellung war jedoch eine Auswertung auf Basis des neuen GP nicht möglich.

3 Anzahl der Unternehmen, die im Berichtsjahr 2017 Betriebe hatten, die im Rahmen der Produktionserhebung gemeldet haben.

4 Die Zahl bezieht sich auf die Unternehmen, die hinter den 668 geheim gehaltenen GP-Positionen stehen. Im Falle der Dominanzregel sind dies nur die dominanten Unternehmen.

5 Die Zahl der 550 Unternehmen bezieht sich auf die Zahl der betroffenen Unternehmen in den neun Bundesländern, die sich an der Umfrage beteiligt haben.

zufällig in einem bestimmten Auswertungsquartal von der Geheimhaltung betroffen waren, erfolgte eine Auswertung von acht Quartalen. Das Ergebnis war, dass 668 Meldenummern dauerhaft aufgrund der Dominanzregel oder der Fallzahlregel geheimzuhalten sind. Gemessen an 4906 Meldenummern des Güterverzeichnisses, die einen Produktionswert größer 0 haben, sind dies 13,6%. „Dauerhaft“ bedeutet in diesem Zusammenhang, dass die Ergebnisse dieser Meldenummern in mindestens sechs von acht (75%) untersuchten Quartalen (2017 und 2018) für die Nutzerin oder den Nutzer aufgrund der Geheimhaltungsregeln nicht einsehbar waren. Zu diesen 668 Güterarten meldeten insgesamt 1152 von der Geheimhaltung betroffene Unternehmen. Dies sind 3,0% aller Unternehmen, die für die Monatliche und Vierteljährliche Produktionserhebung melden. Bezogen auf die neun Bundesländer konnte etwa die Hälfte der 1152 betroffenen Unternehmen (550 Unternehmen) angeschrieben werden. [↘ Tabelle 2](#)

Ergebnis der Unternehmensumfrage

[↘ Tabelle 3](#) zeigt die Anzahl der zustimmenden Unternehmen nach Bundesländern. Von 550 in der Umfrage angeschriebenen Unternehmen haben 92 Unternehmen (16,7%) eine Einwilligungserklärung zurückgeschickt.

Tabelle 3

Zustimmende Unternehmen nach Bundesländern

	Unternehmen in der Umfrage	Unternehmen mit Einwilligungserklärung	
	Anzahl		%
Berlin	13	1	7,7
Brandenburg	21	1	4,8
Bremen	10	2	20,0
Niedersachsen	104	18	17,3
Nordrhein-Westfalen	269	36	13,4
Saarland	14	6	42,9
Sachsen	56	14	25,0
Sachsen-Anhalt	32	6	18,8
Thüringen	31	8	25,8
Insgesamt bezogen auf Unternehmen in der Umfrage	550	92	16,7

Mit diesem Anteil an zustimmenden Unternehmen, bei denen die Daten (oder ein Teil der Daten) veröffentlicht werden dürfen, könnten im Berichtsjahr 2018 insgesamt 54 zusätzliche Ergebnisse in der PRODCOM-Statistik veröffentlicht werden. Diese 54 zusätzlichen Ergebnisse beziehen sich auf insgesamt 8055 Ergebnisse, also

weniger als 1%. Die geringe Anzahl an zusätzlich verfügbaren Ergebnissen resultiert daraus, dass aufgrund der Begrenzung auf nur neun Bundesländer auch nur ein Teil der betroffenen Unternehmen angeschrieben wurde. In mehreren Fällen blieb eine Zustimmung nutzlos, weil – etwa bei einer gesperrten Position aufgrund der Fallzahlregel mit zwei Unternehmen – das andere Unternehmen der Veröffentlichung nicht zugestimmt hat oder nicht befragt werden konnte.


5

Ausblick

Die Beurteilung der im Rahmen des Eurostat-Projekts getesteten Maßnahmen zur Verbesserung der PRODCOM-Statistik fällt eher unterschiedlich aus. Auf der einen Seite konnte ein qualitativ hochwertiges Schätzverfahren zur vollständigen Ermittlung der inländischen Produktion entwickelt werden. Auf der anderen Seite besteht aktuell keine realistische Möglichkeit, die Zahl der Geheimhaltungsfälle in den nationalen und europäischen Veröffentlichungen merklich zu reduzieren. Der hohe Stellenwert, den die statistische Geheimhaltung für die deutsche Industrie hat, wurde durch die Umfrage bei den Unternehmen bestätigt.

Es wurde gezeigt, dass das Multilevel-Modell als Regressionsmodell sehr gut geeignet ist, um den Produktionswert unterhalb der Abschneidegrenze von 20 tätigen Personen zu schätzen. Dies ist auf Ebene der vierstelligen Produktnummer und folglich auch der zweistelligen Produktnummer realisierbar und kann demnach die aktuelle Veröffentlichung ergänzen und bereichern – ohne die Kleinstunternehmen zusätzlich zu belasten. Das Ergebnis für das analysierte Berichtsjahr 2016 war, dass 7% des gesamten Produktionswerts von den nicht erhobenen Unternehmen erwirtschaftet wird. Da eine solche Schätzung auf Ebene der neunstelligen Güternummern nicht möglich ist, müssen die Nutzerinnen und Nutzer in einer solchen Veröffentlichung ausreichend über die Unterschiede informiert werden. Die Umsetzung wird auf nationaler Ebene durch das Abwägen der Vor- und Nachteile noch geprüft. Aufgrund der Nutzung bereits vorliegender Datenquellen in Form von Verwaltungsdaten stellt dieses Vorgehen einen Weg in Richtung einer zukunftsorientierten, belastungsarmen Statistikproduktion dar. Auf europäischer Ebene ist abzuwarten,

ob Eurostat aufgrund der Projektarbeit künftig Produktionszahlen basierend auf der Zuschätzung ergänzend zur Erhebung veröffentlichen möchte. Mittelfristig strebt Eurostat die Veröffentlichung zusätzlicher Aggregate (4-Steller und 2-Steller) bei der PRODCOM-Statistik an. Aus dieser Sicht wäre das Schätzverfahren auf der vierstelligen Hierarchieebene der Güterklassifikation auch für die PRODCOM-Statistik realisierbar.

Die in Deutschland bestehenden Rahmenbedingungen begrenzen die Möglichkeiten der Reduzierung von Geheimhaltungsfällen in der Produktionsstatistik. Die strenge Harmonisierung des p%-Werts im Statistischen Verbund und die hohe Bedeutung des Schutzes von Einzeldaten in Deutschland schränken Änderungen von Geheimhaltungsregeln ein. Auch ohne diese Einschränkung würde die Senkung des p%-Werts nur begrenzt weniger Geheimhaltungsfälle erzeugen. Das große Gewicht der Geheimhaltung der Einzelangaben bei den Meldepflichtigen hat letztlich auch die Unternehmensumfrage bestätigt. Nur 17 % der angeschriebenen Unternehmen erteilten ihr Einverständnis zur Veröffentlichung ihrer eigentlich geheimzuhaltenden Daten. Angesichts des im Ergebnis nur marginalen Effekts (54 zusätzliche Ergebnisse – weniger als 1 %) kann durch dieses Verfahren eine spürbare Minderung der Geheimhaltungsfälle in den Produktionsstatistiken nicht realisiert werden. Dies gilt insbesondere vor dem Hintergrund eines unverhältnismäßig hohen organisatorischen und technischen Aufwands, den eine laufende Implementierung eines solchen Verfahrens in den statistischen Ämtern verursachen würde. Darüber hinaus bedeutet eine gegebenenfalls regelmäßig erforderliche Befragung eine zusätzliche Belastung der Unternehmen. 

LITERATURVERZEICHNIS

- Bührer, Wilhelm/Wagner, Ingo. [150 Jahre Produktionsstatistik im Bergbau und Verarbeitenden Gewerbe](#). In: *Wirtschaft und Statistik*. Ausgabe 2/2010, Seite 109 ff.
- Diebold, Francis X./Mariano, Roberto S. *Comparing Predictive Accuracy*. In: *Journal of Business and Economic Statistics*. Jahrgang 13. 1995, Seite 253 ff.
- Diebold, Francis X. *Comparing Predictive Accuracy, Twenty Years Later: A Personal Perspective on the Use and Abuse of Diebold-Mariano Tests*. In: *Journal of Business and Economic Statistics*. 2014.
- Dinges, Gerlinde/Haitzmann, Martin. *Modellbasierte Ergänzung der Konjunkturstatistik im Produzierenden Bereich*. In: *Austrian Journal of Statistics, Statistische Nachrichten*. Ausgabe 12/2009, Seite 1153.
- Fahrmeir, Ludwig/Kneib, Thomas/Lang, Stefan. *Regression. Modelle, Methoden und Anwendungen*. 2. Auflage. Heidelberg 2009.
- Harvey, David/Leybourne, Stephen/Newbold, Paul. *Testing the equality of prediction mean squared errors*. In: *International Journal of Forecasting*. Jahrgang 13. Ausgabe 2/1997, Seite 281 ff.
- Gelman, Andrew/Hill, Jennifer. *Data Analysis Using Regression and Multilevel/Hierarchical Models*. New York 2006.
- Giessing, Sarah/Dittrich, Stefan. [Tabellengeheimhaltung im statistischen Verbund – ein Verfahrensvergleich am Beispiel der Umsatzsteuerstatistik](#). In: *Wirtschaft und Statistik*. Ausgabe 8/2006, Seite 805 ff.
- Koenker, Roger/Hallock, Kevin. *Quantile Regression: An Introduction*. In: *Journal of Economic Perspectives*. Symposium on Econometric Tools. 2001.
- Koenker Roger/Basset Jr., Gilbert. *Regression Quantiles*. In: *Econometrica*. Jahrgang 46. Ausgabe 1/1978, Seite 33 ff.
- Rohde, Johannes/Seifert, Christiane/Gießing, Sarah. [Entscheidungskriterien für die Auswahl eines Geheimhaltungsverfahrens](#). In: *WISTA Wirtschaft und Statistik*. Ausgabe 3/2018, Seite 90 ff.
- Statistisches Bundesamt. *Güterverzeichnis für Produktionsstatistiken*. 2009. Verfügbar unter: www.destatis.de
- Statistisches Bundesamt. *Beschäftigte, Umsatz und Investitionen der Unternehmen und Betriebe des Verarbeitenden Gewerbes sowie des Bergbaus und der Gewinnung von Steinen und Erden*. 2017a. Fachserie 4 Reihe 4.2.1. Verfügbar unter: www.destatis.de
- Statistisches Bundesamt. *Qualitätsbericht der Investitionserhebung bei Unternehmen und Betrieben des Verarbeitenden Gewerbes sowie des Bergbaus und der Gewinnung von Steinen und Erden*. 2017b. Verfügbar unter: www.destatis.de

LITERATURVERZEICHNIS

Statistisches Bundesamt. *Qualitätsbericht der Produktionserhebungen*. 2016a. Verfügbar unter: www.destatis.de

Statistisches Bundesamt. *Qualitätsbericht des Unternehmensregister-Systems*. 2016b. Verfügbar unter: www.destatis.de

RECHTSGRUNDLAGEN

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

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

Herausgeber

Statistisches Bundesamt (Destatis), Wiesbaden

Schriftleitung

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

Ihr Kontakt zu uns

www.destatis.de/kontakt

Erscheinungsfolge

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

Print

Einzelpreis: EUR 19,- (zzgl. Versand)
Jahresbezugspreis: EUR 114,- (zzgl. Versand)
Bestellnummer: 1010200-19005-1
ISSN 0043-6143
ISBN 978-3-8246-1084-6

Download (PDF)

Artikelnummer: 1010200-19005-4, ISSN 1619-2907

Vertriebspartner

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

Papier: Design Offset, FSC-zertifiziert

© Statistisches Bundesamt (Destatis), 2019

Vervielfältigung und Verbreitung, auch auszugsweise, mit Quellenangabe gestattet.