

HERAUSFORDERUNGEN UND POTENZIALE DER EINZELDATEN-VERKNÜPFUNG IN DER UNTERNEHMENSSTATISTIK

Dr. Sandra Jung, Stefanie Käuser

↳ **Schlüsselwörter:** Einzeldaten – Micro-data linking – kalibrierte Hochrechnung – Unternehmensstatistik – Außenhandelsstatistik

ZUSAMMENFASSUNG

Die Verknüpfung von Einzeldaten unterschiedlicher Statistiken wird in Zukunft zunehmend an Relevanz gewinnen: Auf diese Weise können neue Informationen gewonnen werden, ohne Auskunftspflichtige zusätzlich zu belasten. Dieser Beitrag beschreibt das Vorgehen zur Erstellung eines verknüpften Einzeldatensatzes, bestehend aus den Unternehmensstrukturstatistiken, der Unternehmensdemografie, der Außenhandelsstatistik sowie des Unternehmensregisters. Neben dem Verknüpfungsprozess werden grundlegende methodische Validierungsschritte vorgestellt sowie die kalibrierte Hochrechnung der verknüpften Einzeldaten erläutert. Die Auswertungsmöglichkeiten dieser Daten werden exemplarisch anhand der Beschäftigungsentwicklung außenhandeltreibender und nicht außenhandeltreibender Unternehmen dargestellt.

↳ **Keywords:** microdata – microdata linking – calibration weighting – business statistics – foreign trade statistics

ABSTRACT

Linking microdata from different statistics will be of increasing importance as a tool for obtaining new information without increasing the respondent burden. This article describes the methodology of creating a linked set of microdata from structural business statistics, business demography, foreign trade statistics and the business register. In addition to the data linking process, the article presents the basic validation procedures and discusses methods for calibrated weighting of linked microdata. The analytical potential of the data is examined using the example of employment trends in enterprises engaged and not engaged in foreign trade.



Dr. Sandra Jung

ist Volkswirtin und leitet das Referat „Weiterentwicklung der Unternehmensstatistik“ des Statistischen Bundesamtes. Ihre Arbeitsschwerpunkte sind kleine und mittlere Unternehmen, Inward-FATS und Globalisierung.



Stefanie Käuser

ist Volkswirtin und seit 2014 wissenschaftliche Mitarbeiterin im Referat „Weiterentwicklung der Unternehmensstatistik“ des Statistischen Bundesamtes. Sie beschäftigt sich im Rahmen eines EU-Projektes mit der Verknüpfung und Auswertung von Einzeldaten aus verschiedenen Bereichen der Unternehmensstatistik.

1

Einleitung

Die Verknüpfung von Einzeldaten ist für die amtliche Statistik eine Möglichkeit, auf neue Nutzeranforderungen angemessen zu reagieren. Sie stellt die amtliche Statistik aber auch vor neue Herausforderungen. Die Unternehmensstatistik tut sich mitunter schwer, auf neue Fragestellungen – beispielsweise im Zusammenhang mit Globalisierungsaspekten oder auf dem Gebiet der Nutzung neuer Informationstechnologien – mit einem entsprechenden Datenangebot zu reagieren. Die meisten Erhebungen der Unternehmensstatistik sind nach Wirtschaftsbereichen getrennt ausgerichtet und unterscheiden sich zuweilen auch im Erhebungsinhalt. Querschnittsfragestellungen sind damit häufig nur schwer zu beantworten. Beispiele sind die Auswirkungen von Außenhandelsaktivität oder Nutzung moderner Informationstechnologie auf Unternehmenserfolg und Beschäftigung. Einen belastungsarmen Ausweg bietet hier die Verknüpfung von Einzeldaten unterschiedlicher statistischer Erhebungen.

Die wissenschaftliche Forschung hat das Potenzial von Datenverknüpfungen schon lange erkannt. Unterstützt durch die Forschungsdatenzentren der statistischen Ämter erschließen externe Wissenschaftler auf Basis von verknüpften Mikrodaten der amtlichen Statistik inzwischen vielfältige Analysefelder und -methoden (Voshage und andere, 2015, hier: Seite 14).

Auch auf europäischer Ebene werden die Vorteile der Datenverknüpfung erkannt. Eurostat, das Statistische Amt der Europäischen Union, unterstützt schon seit einigen Jahren Projekte, die sich mit der Zusammenführung und der Analyse amtlicher Einzeldaten befassen.¹ Auch die Erstellung des hier beschriebenen verknüpften Datensatzes wurde von Eurostat gefördert. In dem EU-Projekt “Micro-data linking of Structural Business Statistics and other business statistics” (MDL) wurden Einzeldaten aus verschiedenen Bereichen der Unternehmensstatistik miteinander verknüpft. Neben dem Statistischen Bundesamt waren acht weitere nationale statistische Ämter in der Europäischen Union (Dänemark,

Österreich, Lettland, Finnland, Irland, Niederlande, Portugal, Schweden) sowie Statistics Norway, das nationale statistische Amt Norwegens, an der Durchführung dieses Projektes in den Jahren 2014 und 2015 beteiligt. Hierzu wurden die Einzeldaten des statistischen Unternehmensregisters, der Unternehmensstrukturstatistiken, der Unternehmensdemografie, des Außenhandels, der Statistik auslandskontrollierter Unternehmen (Inward-FATS) und der Erhebung über die Nutzung von Informations- und Kommunikationstechnologien in Unternehmen (IKT-Erhebung) zunächst für jedes einzelne der fünf Berichtsjahre 2008 bis 2012 sowie darüber hinaus für den gesamten Beobachtungszeitraum zu einem Längsschnittdatensatz zusammengeführt und ausgewertet.

Die mit der Verknüpfung von Einzeldaten gewonnenen neuen Informationsangebote stellen die amtliche Statistik aber auch vor neue Aufgaben. Nicht zuletzt sind die Vorgaben des Bundesstatistikgesetzes eine wichtige Rahmenbedingung, die hinsichtlich der Durchführung sowie der Berichtspflicht bei Zusammenführungen von Datensätzen von unterschiedlichen Erhebungen zu berücksichtigen ist.

➔ § 13a Bundesstatistikgesetz: Zusammenführung von Daten

Soweit es zur Gewinnung von statistischen Informationen ohne zusätzliche statistische Erhebungen erforderlich ist, dürfen Daten aus Statistiken nach § 13 Abs. 1, Daten aus dem Statistikregister, Daten nach dem Verwaltungsdatenverwendungsgesetz und Daten, die die statistischen Ämter des Bundes und der Länder aus allgemein zugänglichen Quellen gewinnen, zusammengeführt werden.

Aber auch methodisch ist die Einzeldatenverknüpfung eine neue Herausforderung. Im Folgenden werden grundlegende Aspekte der Validierung des erstellten MDL-Datensatzes erläutert sowie Möglichkeiten aufgezeigt, wie größtenteils stichprobenbasierte Erhebungen angemessen berücksichtigt werden können. Als Beispiel für die Auswertungsmöglichkeiten des Datensatzes wird die Beschäftigungsentwicklung außenhandeltreibender Unternehmen dargestellt.

¹ Beispielhaft sind hier die folgenden EU-Projekte zu nennen: ICT-Impact-Project, ESSLimit-Project, ESSLait-Project. Für einen Bericht über das ICT-Impact-Projekt siehe Rauland/Bauer, 2010.

2

Datenbasis und Verknüpfungsprozess

Für die Erstellung des MDL-Datensatzes wurde eine Vielzahl unterschiedlich ausgestalteter Erhebungen auf Einzeldatenebene miteinander verknüpft. Dazu muss ein eindeutiger Identifikator vorhanden sein, mit dem die Einzeldaten der jeweiligen Statistik im Querschnitt sowie im Längsschnitt miteinander verknüpft werden können. Nachfolgend werden die verwendeten Datenquellen kurz erläutert sowie eine Übersicht der Verknüpfungsquoten präsentiert.

2.1 Verwendete Datenquellen

Das Grundgerüst des MDL-Datensatzes bilden die Unternehmensstrukturstatistiken der Berichtsjahre 2008 bis 2012. Diese jährlichen Erhebungen umfassen das Produzierende Gewerbe, das Baugewerbe, den Handel sowie die wesentlichen Teile der Dienstleistungen. Die Erhebungsmerkmale decken die Hauptbereiche der wirtschaftlichen Tätigkeit von Unternehmen ab, wie den Umsatz und die tätigen Personen. Darüber hinaus steht eine Vielzahl weiterer Merkmale zur Verfügung (zum Beispiel Bruttowertschöpfung, Investitionen, Löhne und Gehälter). Die meisten Strukturstatistiken sind Stichprobenerhebungen. ↘ Übersicht 1

Das statistische Unternehmensregister (URS) dient als weitere wichtige Datenquelle und spielt insbesondere bei der Verknüpfung der einzelnen Erhebungen eine zentrale Rolle. Das Unternehmensregister ist eine laufend aktualisierte Datenbank und umfasst alle aktiven rechtlichen Unternehmenseinheiten und deren Betriebe/Niederlassungen. Im URS sind unternehmensrelevante Merkmale und administrative Daten zu steuerbaren Umsätzen und sozialversicherungspflichtig Beschäftigten verfügbar. Auch Angaben über die Zugehörigkeit eines Unternehmens zu einer Unternehmensgruppe sind enthalten. Jede rechtliche Unternehmenseinheit im URS besitzt eine Identifikationsnummer (URS-ID). Die URS-ID dient als Identifikator für die Verknüpfung der Strukturstatistiken (einschließlich Unternehmensdemografie, Inward-FATS und IKT-Erhebung) untereinander und mit dem URS.

Eine weitere verwendete Datenquelle ist die Außenhandelsstatistik, welche den grenzüberschreitenden Warenverkehr mit dem Ausland darstellt und aus der Extrahandels- sowie der Intrahandelsstatistik besteht. Während die Intrahandelsstatistik den Warenhandel zwischen den EU-Mitgliedstaaten dokumentiert, erfasst der Extrahandel den Warenverkehr mit sogenannten Drittländern.¹² Als Identifikator zur Verknüpfung der Außenhandelsstatistik mit dem URS dient die Umsatzsteuernummer.

Die Unternehmensdemografie liefert Angaben über Unternehmensgründungen und -schließungen und den Bestand aktiver Unternehmen. Damit lassen sich Strukturindikatoren wie die Gründungsrate und die Überlebensquote von neu gegründeten Unternehmen berechnen. Wesentliche Quelle für die Ermittlung von Informationen zur Unternehmensdemografie ist das statistische Unternehmensregister (Rink und andere, 2013).

Daneben wurden Angaben der Statistik über Auslandsunternehmenseinheiten (sogenannte Inward-FATS; Nahm/Söllner, 2014) und der Erhebung über die Nutzung von Informations- und Kommunikationstechnologien in Unternehmen (IKT) in die Analyse einbezogen.¹³

2.2 Verknüpfungsprozess

Die Datenverknüpfung wurde in zwei Stufen durchgeführt. In der ersten Stufe wurden die Einzeldaten der Unternehmensstrukturstatistiken, der Unternehmensdemografie, der Inward-FATS und der IKT-Erhebung mithilfe der URS-ID zunächst mit dem URS verknüpft und anschließend zu einem Querschnittsdatensatz für jedes Berichtsjahr zusammengeführt. Die Verknüpfung der Außenhandelsstatistik mit dem URS erfolgte anhand der Umsatzsteuernummer. In der zweiten Stufe wurden die Daten aller Berichtsjahre mittels URS-ID zu einem finalen Längsschnittsdatensatz verknüpft. Der Längsschnitt-

2 Detaillierte Informationen zu den verwendeten Methoden und Definitionen sowie zur Qualität der statistischen Ergebnisse der Außenhandelsstatistik enthält der Qualitätsbericht „Außenhandel“ (www.destatis.de).

3 Detaillierte Informationen zu den verwendeten Methoden und Definitionen sowie zur Qualität der statistischen Ergebnisse der IKT-Erhebung enthält der Qualitätsbericht „Nutzung von Informations- und Kommunikationstechnologien (IKT) in Unternehmen 2015“ (www.destatis.de).

Übersicht 1

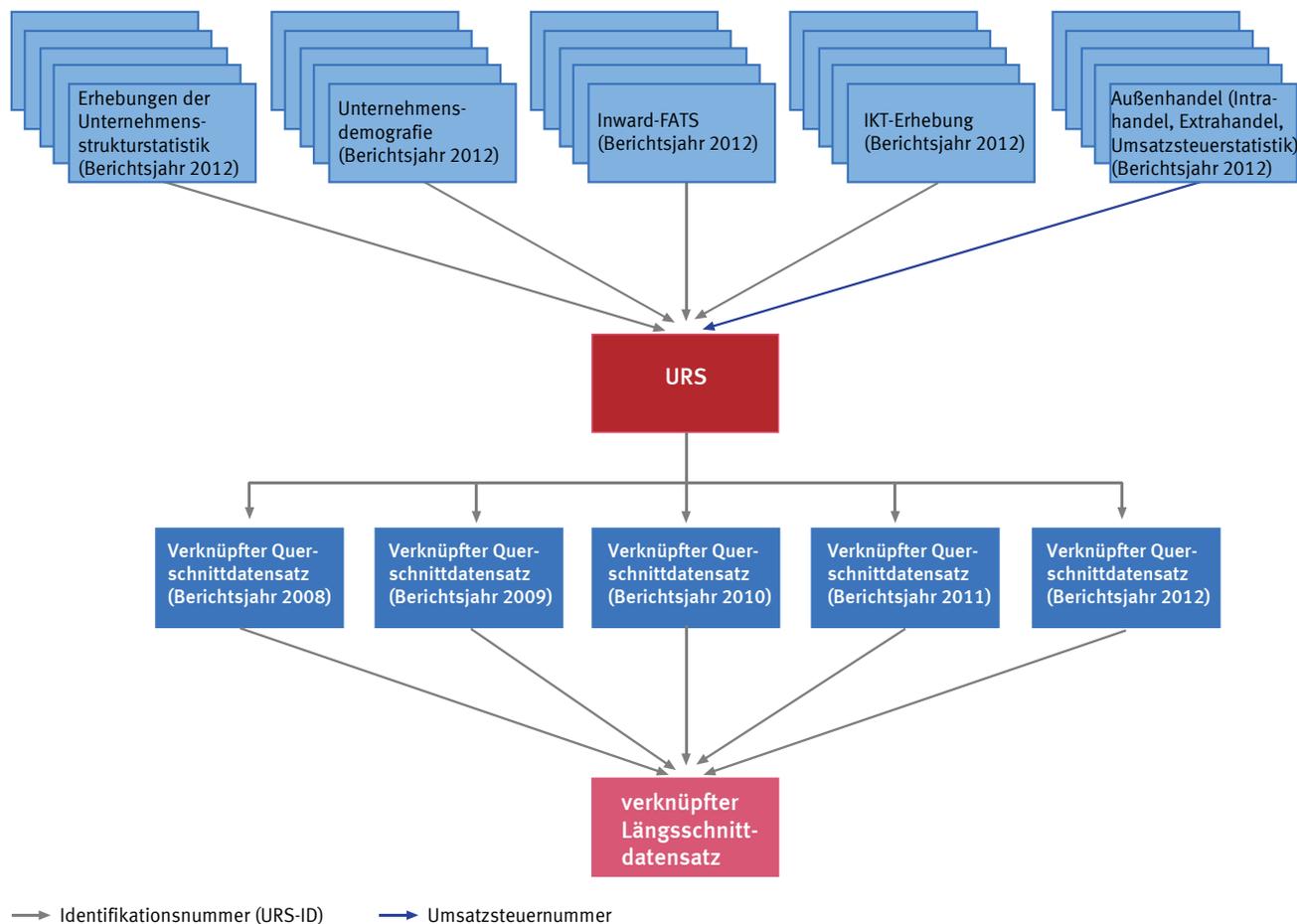
Datenbasis des MDL-Datensatzes

Wirtschaftsabschnitt ¹	Erhebung	Art der Erhebung	Abschneidegrenze	Erhebungsumfang
Jährliche Unternehmensstrukturstatistiken (SBS)				
B, C	Kostenstrukturerhebung im Bereich Verarbeitendes Gewerbe, Bergbau und Gewinnung von Steinen und Erden	Stichprobe	Unternehmen mit 20 und mehr Beschäftigten	etwa 18 000 Unternehmen
	Investitionserhebung im Bereich Verarbeitendes Gewerbe, Bergbau und Gewinnung von Steinen und Erden	Totalerhebung	Unternehmen mit 20 und mehr Beschäftigten	etwa 50 000 Unternehmen
	Strukturerhebung für kleine Unternehmen im Bereich Verarbeitendes Gewerbe, Bergbau und Gewinnung von Steinen und Erden	Stichprobe	Unternehmen mit weniger als 20 Beschäftigten	etwa 6 000 Unternehmen
D, E	Kostenstrukturerhebung im Bereich der Energie- und Wasserversorgung	Totalerhebung	Unternehmen der Wasserversorgung mit einer jährlichen Wasserabgabe von 200 000 m ³ und mehr	3 000 Energieversorgungsunternehmen; 7 000 Wasserversorgungsunternehmen
	Investitionserhebung im Bereich der Energie- und Wasserversorgung	Totalerhebung	Unternehmen der Wasserversorgung mit einer jährlichen Wasserabgabe von 200 000 m ³ und mehr	3 000 Energieversorgungsunternehmen; 7 000 Wasserversorgungsunternehmen
F	Kostenstrukturerhebung im Baugewerbe	Stichprobe	Unternehmen mit 20 und mehr Beschäftigten	6 000 Unternehmen
	Jahreserhebung einschließlich Investitionserhebung im Bauhauptgewerbe	Totalerhebung	Unternehmen mit 20 und mehr Beschäftigten	etwa 7 500 Unternehmen
	Jahreserhebung einschließlich Investitionserhebung im Ausbaugewerbe	Totalerhebung	Unternehmen mit 20 und mehr Beschäftigten	etwa 6 500 Unternehmen
	Strukturerhebung für kleine Unternehmen im Baugewerbe	Stichprobe	Unternehmen mit weniger als 20 Beschäftigten	6 000 Unternehmen
G	Jahreserhebung im Handel sowie in der Instandhaltung und Reparatur von Kraftfahrzeugen und Gebrauchsgütern	Stichprobe	keine	etwa 47 000 Unternehmen
I	Jahreserhebung im Gastgewerbe	Stichprobe	keine	etwa 10 000 Unternehmen
H, J, L, M, N, S95	Strukturerhebung im Dienstleistungsbereich	Stichprobe	keine	etwa 220 000 Unternehmen
Unternehmensdemografie				
B bis N, P bis S	Unternehmensneugründungen und -aufgaben	Registerbasierte Auswertung	mindestens 17 500 EUR steuerbarer Umsatz oder mindestens 1 sozialversicherungspflichtig Beschäftigter	etwa 3 700 000 Unternehmen (einschließlich überlebende Unternehmen)
IKT-Erhebung				
C bis N	Nutzung von Informations- und Kommunikationstechnologien in Unternehmen	Stichprobe (freiwillige Befragung)	Kleinstunternehmen (weniger als 10 Beschäftigte) erhalten reduzierten Fragebogen	etwa 20 000 Unternehmen (Anteil Kleinstunternehmen zwischen 25 und 30%)
Inward-FATS				
B bis N, P bis S	Auslandskontrollierte Unternehmen	Extraktion aller auslandskontrollierten Unternehmen aus dem URS, Verknüpfung mit den SBS-Daten	mindestens 17 500 EUR steuerbarer Umsatz oder mindestens 1 sozialversicherungspflichtig Beschäftigter	etwa 21 000 bis 27 000 Unternehmen
Außenhandelsstatistik				
Alle	Güterexporte und -importe deutscher Unternehmen	Totalerhebung	Intrahandel: 500 000 EUR Warenwert je Verkehrsrichtung im Berichtsjahr (Meldepflicht) (400 000 EUR bis 2011), Extrahandel: keine	etwa 550 000 bis 577 000 Unternehmen (einschließlich Unternehmen unterhalb der Meldeschwelle im Intrahandel)

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

Grafik 1

Darstellung des Verknüpfungsprozesses von Einzeldaten aus unterschiedlichen Unternehmensstatistiken



2016 - 01 - 0153

datensatz enthält Einzeldaten von Unternehmen, die in mindestens einer Erhebung (einschließlich URS) in mindestens einem Berichtsjahr vorhanden sind und umfasst rund 7,7 Millionen Einheiten. [↘ Grafik 1](#)

Auswahlgrundlage für die Stichprobenziehung der Strukturstatistiken dient, sind die entsprechenden Verknüpfungsquoten mit rund 99% relativ hoch. In Ausnahmefällen kann es allerdings sein, dass eine Ein-

2.3 Verknüpfungsquoten

In dem zweistufigen Verknüpfungsprozess wurden unterschiedlich hohe Verknüpfungsquoten erreicht. Eine möglichst hohe Verknüpfungsquote ist wünschenswert, da dies die Belastbarkeit der Auswertungsergebnisse erhöht.

[↘ Tabelle 1](#) stellt die Verknüpfungsquoten der jeweiligen Erhebungen mit dem URS dar. Da das URS als

Tabelle 1
Verknüpfungsquoten¹ der einzelnen Erhebungen mit dem statistischen Unternehmensregister (URS)

	2008	2009	2010	2011	2012
	%				
Unternehmensstrukturstatistiken	99,4	99,5	99,6	99,6	99,6
Außenhandelsstatistik	99,6	99,6	99,0	100	99,7
Inward-FATS	98,6	100	100	100	100
Unternehmensdemografie	100	100	100	100	100
IKT-Erhebung	98,8	99,0	99,1	98,9	99,1

¹ Anteil der Einheiten in der Datenquelle, die mit dem URS des jeweiligen Berichtsjahres verknüpft werden können.

heit der Strukturstatistiken nicht im URS des gleichen Berichtsjahres aufzufinden ist (siehe ausführlicher auch Abschnitt 3.1).⁴

Da die Unternehmensstrukturstatistiken das Grundgerüst des MDL-Datensatzes sind, ist es interessant zu sehen, wie hoch die Verknüpfungsquoten der Strukturstatistiken mit den übrigen Erhebungen sind. Für die Außenhandelsstatistik sind die Verknüpfungsquoten relativ niedrig: Sie liegen zwischen 11,2% und 13,5%. Der Hauptgrund dafür ist, dass die Außenhandelsstatistik eine Totalerhebung mit Abschneidegrenze ist, die Strukturstatistiken dagegen Stichprobenerhebungen (siehe auch Abschnitt 3.1). Beispielsweise ist ein in der Außenhandelsstatistik erfasstes Unternehmen möglicherweise nicht in den Strukturstatistiken enthalten, da es nicht für die jeweilige Stichprobe gezogen wurde. Oder der Außenhandelsumsatz eines in den Strukturstatistiken enthaltenen Unternehmens liegt unterhalb der Abschneidegrenze der Außenhandelsstatistik und das Unternehmen wird somit in der Außenhandelsstatistik nicht erfasst. In beiden Fällen ist eine Verknüpfung nicht möglich. Bei der IKT-Erhebung liegen die Verknüpfungsquoten zwischen 40,1% und 45,4% und damit am höchsten. [↘ Tabelle 2](#)

Tabelle 2
Verknüpfungsquoten¹ mit den Unternehmensstrukturstatistiken

	2008	2009	2010	2011	2012
	%				
Außenhandelsstatistik	11,2	12,6	12,8	13,4	13,5
Inward-FATS	30,4	31,5	30,6	31,1	35,6
IKT-Erhebung	43,0	45,4	40,6	40,1	45,4

¹ Anteil der Einheiten in der jeweiligen Erhebung, die im jeweiligen Berichtsjahr in den Unternehmensstrukturstatistiken befragt wurden.

Die zweite Stufe des Verknüpfungsprozesses war die Bildung eines Längsschnittdatensatzes. Interessant sind hier die Verknüpfungsquoten der Strukturstatistiken über alle Berichtsjahre, da sie angeben, welcher Anteil der Einheiten des Berichtsjahres 2008 in den Folgejahren noch beobachtet werden kann. [↘ Tabelle 3](#)

⁴ Die recht hohen Verknüpfungsquoten des Außenhandels mit dem URS beziehen sich auf bereits aufbereitete Außenhandelsdaten. Bei einer erstmaligen Verknüpfung der Außenhandelsdaten mit dem URS sind diese Quoten etwas niedriger (Allafi, 2011).

Tabelle 3
Verknüpfungsquoten¹ der Einzeldaten aus den Unternehmensstrukturstatistiken im Längsschnitt²

	2009	2010	2011	2012
	%			
Unternehmensstrukturstatistiken	72,1	65,0	31,3	25,9

¹ Anteil der 2008 in den Unternehmensstrukturstatistiken befragten Einheiten, die in den Folgejahren weiterhin beobachtet wurden.

² Bezogen auf die Kohorte 2008 der Unternehmensstrukturstatistiken.

Von den 2008 in den Unternehmensstrukturstatistiken befragten Einheiten waren im Jahr 2012 noch 25,9% im MDL-Datensatz vorhanden. Der Rückgang der Verknüpfungsquoten im Zeitablauf hat unterschiedliche Gründe. Hauptursache hierfür ist die Stichprobenrotation: Befragte Einheiten werden in unterschiedlichen zeitlichen Abständen gegen neue Einheiten ausgetauscht. Daneben können Unternehmensschließungen ein Grund für den Rückgang sein (siehe auch Abschnitt 3.1).

3

Validierung und Hochrechnung

3.1 Validierung der verknüpften Daten

Um die Konsistenz und Qualität des MDL-Datensatzes im Hinblick auf spätere Auswertungen sicherzustellen, wurden die verknüpften Daten einem umfangreichen Validierungsprozess unterzogen. Ein Schwerpunkt dabei war die Untersuchung der Ursachen von nicht verknüpfbaren Einheiten zwischen den Datenquellen und innerhalb der Datenquellen im Zeitverlauf. Ein zweiter Aspekt bei der Validierung war, die Meldungen steuerlicher Organschaften⁵ in der Außenhandelsstatistik angemessen zu berücksichtigen. Schließlich wurde eine Vorgehensweise erarbeitet, um sogenannte demografische Ereignisse im MDL-Datensatz zu identifizieren. Die drei genannten Aspekte werden nachfolgend erläutert.

Nicht verknüpfbare Einheiten

Bei der Vielzahl der verwendeten Erhebungen ist eine hohe Anzahl von nicht verknüpfbaren Einheiten nicht

⁵ Eine ausführliche Erläuterung zu steuerlichen Organschaften findet sich in Wagner, 2004.

ungewöhnlich. Bei der Untersuchung der Ursachen wurde unterteilt in Ursachen, die auf das Design und die Erhebungsmethodik der Statistiken zurückzuführen sind, und Ursachen, die aufgrund technischer Fehler bei der Verknüpfung entstehen. Ziel dieses Validierungsschrittes war es, eventuell aufgetretene Fehler bei der Zuordnung der Einheiten zu identifizieren und nachträglich zu korrigieren.

Ursachen, die durch das Erhebungsdesign bedingt sind, können zum Beispiel unterschiedliche Grundgesamtheiten der Statistiken oder unterschiedliche Konzepte der Datengewinnung sein. Da die im MDL-Projekt verwendeten Erhebungen Totalerhebungen oder Stichproben mit jeweils unterschiedlichen Stichprobenumfängen und Abschneidegrenzen sind, konnten nicht alle Einheiten datenquellenübergreifend verknüpft werden. Ferner bilden die Erhebungen zum Teil unterschiedliche Grundgesamtheiten ab: Die Grundgesamtheit der Strukturstatistiken sind alle aktiven Unternehmen in Deutschland innerhalb der Wirtschaftsbereiche B bis N (ohne K) und S95 der Klassifikation der Wirtschaftszweige, Ausgabe 2008 (WZ 2008). Die Grundgesamtheit der Außenhandelsstatistik bezieht sich nicht auf eine Wirtschaftszweigabgrenzung, sondern umfasst alle Unternehmen, die Außenhandel mit Waren betreiben. Vor diesem Hintergrund sind auch die relativ niedrigen Verknüpfungsquoten der Unternehmensstrukturstatistiken mit der Außenhandelsstatistik zu sehen (siehe Tabelle 2). Nicht verknüpfbare Einheiten aufgrund unterschiedlicher Erhebungsdesigns sind eine Konsequenz der verfügbaren Datenlage und können nicht nachträglich korrigiert werden.

Grundsätzlich wäre es für die Einzeldatenverknüpfung wünschenswert, Stichprobenerhebungen aufeinander abzustimmen und das Erhebungsdesign zu vereinheitlichen. Dadurch würden sich die Verknüpfungsquoten der beteiligten Stichprobenerhebungen erhöhen mit dem Vorteil, dass bei der Hochrechnung der verknüpften Daten aus Stichproben die Unsicherheit der Ergebnisse reduziert wird. Die Koordinierung der in das MDL-Projekt einfließenden Stichprobenerhebungen kann jedoch nachträglich nicht beeinflusst werden.

Neben den beschriebenen Faktoren hinsichtlich der Erhebungsmethodik können auch Fehler bei der Verknüpfung über die Identifikationsnummer auftreten. Diese Fehler können prinzipiell nachträglich korrigiert

werden. Beispielsweise kann eine Einheit in verschiedenen Datenerhebungen unterschiedliche Identifikationsnummern aufweisen. Dies kann bei den Einzeldaten des Unternehmensregisters und aus den Unternehmensstrukturstatistiken auftreten, wenn einem Unternehmen durch den Umzug in ein anderes Bundesland eine neue Identifikationsnummer im Unternehmensregister zugewiesen wird, dies zum Zeitpunkt der Ziehung der Stichprobe für die Unternehmensstrukturstatistik(en) jedoch noch nicht bekannt war. Solche Diskrepanzen zwischen den Zeitständen der Datensätze sind unter anderem dafür verantwortlich, dass die Verknüpfungsquoten von URS und Unternehmensstrukturstatistiken in allen Berichtsjahren nicht exakt 100% erreichen (siehe Tabelle 1).

Außerdem ist es möglich, dass eine Einheit innerhalb einer Datenerhebung unterschiedliche Identifikationsnummern in verschiedenen Berichtsjahren aufweist. In den verknüpften Einzeldaten ist die betroffene Identifikationsnummer dann nicht durchgehend im Längsschnittdatensatz vorhanden. Für das MDL-Projekt wurde dieser Aspekt eingehend untersucht. Zumeist konnten Stichprobenrotationen, Wechsel des Wirtschaftszweiges oder des Bundeslandes als Ursachen identifiziert werden.

Umverteilung von Organschaftsmeldungen im Außenhandel

Im Rahmen des MDL-Projekts trat bei der Verknüpfung der Strukturstatistiken mit den Außenhandelsdaten das Problem auf, dass in der Außenhandelsstatistik teilweise statistische Meldungen von steuerlichen Organschaften vorliegen. Steuerliche Organschaften sind ökonomisch miteinander verbundene rechtliche Einheiten, die gemeinsam steuerrechtlich als eine Einheit behandelt werden und beispielsweise gemeinsam ihre Umsatzsteuer entrichten (Wagner, 2004). Bei Einheiten, die zu einer steuerrechtlichen Organschaft gehören, ist der Organträger einer Organschaft stellvertretend für alle Mitglieder der Organschaft meldepflichtig für die Außenhandelsumsätze. Dies hat zur Folge, dass sich die unter der Identifikationsnummer des Organträgers aufgeführten Außenhandelsumsätze auf die gesamte Organschaft beziehen. Im Rahmen dieses Validierungsschrittes zeigte sich dies daran, dass im MDL-Datensatz die Exporte einer Einheit ein Vielfaches der in den Unternehmensstrukturstatistiken gemeldeten Umsätze betragen.

Um die oben beschriebene Problematik zu beheben, wurden die von den Organträgern im Außenhandel gemeldeten Exporte und Importe innerhalb der Organschaft umverteilt. Dazu wurden über das Unternehmensregister die zu den Organträgern gehörenden Organgesellschaften ermittelt und deren Identifikationsnummern zu den Außenhandelsdaten hinzugefügt. Anschließend wurden die vom Organträger gemeldeten Außenhandelsumsätze (Exporte oder Importe) innerhalb des Organkreises umverteilt. Als Schlüsselgröße für die Verteilung wurden die umverteilten Umsätze⁶ aus dem Unternehmensregister verwendet. Die Umverteilung der Außenhandelswerte wird durch folgende Formel beschrieben:

$$(1) A_{ip} = \frac{\alpha_i}{\sum_{i=1}^k \alpha_i} \beta_p^{org} A^{org},$$

wobei A_{ip} der geschätzte Außenhandelsumsatz von Organkreismitglied i mit Partnerland p ist. A^{org} ist der gesamte Außenhandelsumsatz eines Organträgers. Der geschätzte Umsatz des Organkreismitglieds i ist α_i , dieser wird dividiert durch die Summe der Umsätze aller k Mitglieder des Organkreises: $\sum_{i=1}^k \alpha_i$. Daraus ergibt sich also der Anteil des Organkreismitglieds i an den Umsätzen des jeweiligen Organkreises. β_p^{org} ist der Anteil des Partnerlands p an den gesamten Außenhandelsumsätzen des Organträgers.

Ausgenommen von der Umverteilung der Exporte wurden Organkreismitglieder, deren primäre wirtschaftliche Tätigkeit nicht im Verarbeitenden Gewerbe oder im Binnenhandel lag, da nur solche Wirtschaftszweige miteinbezogen werden sollten, in denen Güterexporte in nennenswertem Umfang zu erwarten sind. Durch die Umverteilung der Außenhandelsumsätze innerhalb der Organschaften wurden die Konsistenz des MDL-Datensatzes verbessert und die Plausibilität der Daten erhöht.

Vorgehen bei demografischen Ereignissen

Die Unternehmenslandschaft in einer Volkswirtschaft ist nicht statisch, sondern unterliegt einem beständigen Wandel durch Neugründungen und Schließungen von Unternehmen sowie durch andere sogenannte demografische Ereignisse (zum Beispiel Fusionen, Übernah-

⁶ Eine Umverteilung der von Organträgern gemeldeten Umsätze innerhalb des Organkreises wird regelmäßig bei der Erstellung der Bundeskopien des Unternehmensregisters durchgeführt.

men, Ausgliederungen, Konzernumstrukturierungen, Gründungen).⁷ Diese demografischen Ereignisse können innerhalb des Berichtszeitraums auftreten und sich im MDL-Datensatz auf unterschiedliche Weise niederschlagen. Zum Beispiel kann sich eine Unternehmensübernahme dadurch zeigen, dass die Identifikationsnummer des übernommenen Unternehmens aus dem Datensatz verschwindet, wohingegen in den Merkmalen des übernehmenden Unternehmens auffallende Abweichungen, zum Beispiel bei den tätigen Personen und Umsätzen, zu verzeichnen sind. Solche Konstellationen können bei Längsschnittanalysen von Wachstumsraten problematisch sein, da unternehmensinternes Wachstum nicht von „künstlichem“ Wachstum, zum Beispiel durch Fusionen und Übernahmen, unterschieden werden kann. Um den MDL-Datensatz für Längsschnittanalysen konsistent zu machen, wurde ein eigenes Merkmal für dieses Phänomen erstellt.

Da dieses Merkmal für den Validierungsschritt bedeutsam war, wird es im Folgenden genauer vorgestellt. Zunächst wurden für alle Einheiten der Strukturstatistiken vier Größenklassen, basierend auf ihren Angaben zu tätigen Personen und Umsätzen, erstellt. Innerhalb dieser Größenklassen wurde dann die durchschnittliche Veränderung der Merkmale Umsatz und Anzahl der Lohn- und Gehaltsempfänger in Vollzeitseinheiten (VZE) in jeweils zwei aufeinanderfolgenden Jahren berechnet. Diese wird durch den Term $\overline{\Delta x_{gt}}$ gekennzeichnet. Zudem wurde die Standardabweichung (σ_{gt}) des jeweiligen Merkmals innerhalb der Größenklasse im selben Zeitraum berechnet. Auf Basis dieser Größen konnten Einheiten aus den Unternehmensstrukturstatistiken selektiert werden, deren Angaben beim Merkmal Umsatz oder Lohn- und Gehaltsempfänger (VZE) in zwei aufeinanderfolgenden Jahren um mehr als drei Standardabweichungen der durchschnittlichen Veränderung innerhalb ihrer Größenklasse abweichen. Dieser Zusammenhang wird durch folgende Formel ausgedrückt:

$$(2) |\Delta x_{it}| > |\overline{\Delta x_{gt}}| + 3\sigma_{gt},$$

wobei Δx_{it} die Veränderung des Merkmalswertes bei Einheit i im Zeitraum t darstellt. Dieses Kriterium sollte Einheiten selektieren, die ungewöhnlich hohe Schwankungen bei den Merkmalen Umsatz und Lohn- und Gehaltsempfänger (VZE) aufweisen.

⁷ Für eine genaue Abgrenzung demografischer Ereignisse siehe Rink und andere, 2013, hier: Seite 425 f.

In einem weiteren Schritt wurde überprüft, ob im selben Zeitraum Veränderungen bei weiteren Merkmalen aus dem Unternehmensregister (zum Beispiel Unternehmensbezeichnung, Gruppenzugehörigkeit, Art der Einheit oder Wirtschaftszweig) aufgetreten waren. Dabei zeigte sich, dass bei allen zuvor selektierten Einheiten Auffälligkeiten bei dem URS-Merkmal „Art der Einheit“ zu erkennen waren. Dieses Merkmal enthält Informationen darüber, ob es sich bei der Einheit um ein Unternehmen handelt oder um einen Betrieb. Bei einem Großteil der selektierten Einheiten änderte sich im betroffenen Zeitraum der Status bei diesem Merkmal von „Unternehmen“ zu „Betrieb“. Es ist davon auszugehen, dass bei diesen Einheiten ein demografisches Ereignis stattgefunden hat. Die betroffenen Einheiten wurde im MDL-Datensatz entsprechend gekennzeichnet.

3.2 Ansätze zur Hochrechnung

Bei der Verknüpfung von Einzeldaten aus Stichproben kann das Problem auftreten, dass errechnete Kennzahlen auf Basis der verknüpften Daten nicht mit entsprechenden Eckwerten bereits existierender Veröffentlichungen aus den einzelnen Erhebungen übereinstimmen. Um die Konsistenz der Ergebnisse zu gewährleisten, muss deshalb eine geeignete Hochrechnungsmethode gewählt werden. Die fehlende Konsistenz der Merkmalswerte war im MDL-Projekt insbesondere im Hinblick auf die Eckwerte von wichtigen Merkmalen der Strukturstatistiken und Inward-FATS zu beachten. Im Folgenden werden deshalb Lösungsansätze vorgestellt, die die Konsistenz bei der Hochrechnung verknüpfter Einzeldaten aus mehreren Quellen gewährleisten sollen.

Kalibrierte Hochrechnung

Die kalibrierte Hochrechnung ist eine Form der gebundenen Hochrechnung. Die Vorgehensweise ähnelt der freien Hochrechnung, unterscheidet sich jedoch dadurch, dass Korrekturfaktoren zur Kalibrierung der Hochrechnungsfaktoren berechnet werden. Ziel der Kalibrierung ist es, die Hochrechnungsfaktoren so anzupassen, dass ein vorgegebener Merkmalseckwert erreicht wird. Die Berechnung der Korrekturfaktoren erfolgt anhand folgender Formel:

$$(3) \quad KF_j = \frac{Y_j}{\hat{Y}_j},$$

wobei $\hat{Y}_j = \sum_{i \in j} \hat{y}_i$ die Summe des hochgerechneten Merkmals y der Einheiten i in WZ-Abschnitt j darstellt und $Y_j = \sum_{i \in j} y_i$ die Summe des Merkmals y der Einheiten i in WZ-Abschnitt j in der Originalstatistik ausdrückt. Der Korrekturfaktor wird mit dem ursprünglichen Hochrechnungsfaktor multipliziert, um den kalibrierten Hochrechnungsfaktor zu erhalten.

Für das MDL-Projekt wurde zur Hochrechnung der verknüpften Einzeldaten eine kalibrierte Hochrechnung mit Korrekturfaktoren durchgeführt. Ziel war, Konsistenz mit den wichtigsten Merkmalseckwerten der Inward-FATS-Statistik herzustellen unter der Nebenbedingung, dass die Abweichungen zu den (wichtigen) Eckwerten der Unternehmensstrukturstatistiken (gesamt) minimiert werden. Diese Eckwerte wurden als Referenz ausgewählt, da in den geplanten Projektveröffentlichungen insbesondere Merkmale von auslandskontrollierten sowie kleinen und mittleren Unternehmen im Fokus stehen sollten. Es wurde deutlich, dass diese Methode zwar unkompliziert in der Durchführung ist, sie aber hinsichtlich der Anwendungsmöglichkeiten begrenzt ist. Bei einer geringen Anzahl an Merkmalen kann die Konsistenz mit bereits veröffentlichten Ergebnissen hergestellt werden. Allerdings ist diese Methode bei einer größeren Anzahl zu treffender Eckwerte nicht mehr praktikabel. Aufgrund dieses Nachteils ist eine verbesserte Hochrechnungsmethodik für künftige Arbeiten erstrebenswert.

General Regression Estimator

Als Alternative zur kalibrierten Hochrechnung mit Korrekturfaktoren kommt eine regressionsbasierte Kalibrierung der Hochrechnungsfaktoren mit dem General Regression Estimator (GREG) infrage (Deville/Särndal, 1992). Mit GREG können Hochrechnungsfaktoren mithilfe von Hilfsvariablen kalibriert werden, um vorgegebene Eckwerte zu erreichen. Der Vorteil dieser Methode ist, dass die Eckwerte mehrerer Statistiken gleichzeitig erreicht werden können. Auch ermöglicht GREG die Evaluation der Schätzgenauigkeit in Form des relativen Standardfehlers. Diese Methode der kalibrierten Hochrechnung wird zum Beispiel auch für den Mikrozensus eingesetzt (Afen-takis/Bihler, 2005). Für ein Nachfolgeprojekt ist geplant, die Hochrechnung mit GREG durchzuführen.

4

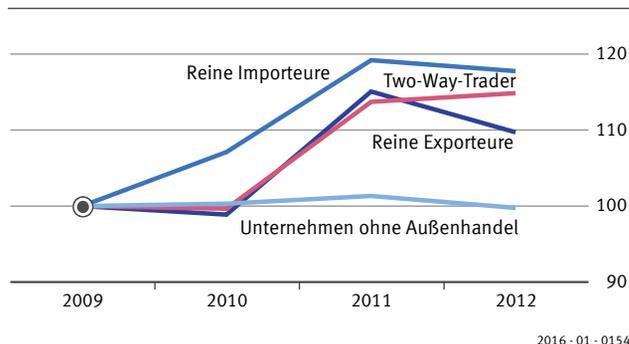
Auswertungsmöglichkeiten

Mithilfe der Einzeldatenverknüpfung verschiedener Unternehmensstatistiken bietet sich die Möglichkeit, statistikübergreifende Fragestellungen zu beantworten, die bisher nur eingeschränkt untersucht werden konnten. Beispielsweise lassen sich aus der Verknüpfung von Einzeldaten der Außenhandelsstatistik mit Daten aus der Unternehmensstrukturstatistik und der Inward-FATS-Statistik neue Erkenntnisse über die Richtung und Struktur grenzüberschreitender Wertschöpfungsketten gewinnen. Zusammenhänge zwischen Kostenstrukturen, Produktivität, Marktaustritten und Änderungen in den Besitz- und Kontrollstrukturen sind weitere Themen für empirische Analysen, die durch Datenverknüpfungen möglich sind.

Im Folgenden wird beispielhaft die Beschäftigungsentwicklung von Außenhandel betreibenden und nicht Außenhandel betreibenden Unternehmen im Zeitraum 2009 bis 2012¹⁸ untersucht. Dafür wurde die Entwicklung der Anzahl der Beschäftigten je Unternehmen für Importeure, Exporteure, Two-Way-Trader¹⁹ sowie für Unternehmen, die keinen Außenhandel betreiben, dargestellt. [↘ Grafik 2](#)

Grafik 2

Entwicklung der Anzahl der Beschäftigten je Unternehmen nach Außenhandelsaktivität
2009 = 100



Die Beschäftigtenzahl von Unternehmen, die nicht am Außenhandel beteiligt sind, blieb zwischen 2009 und

8 Das Berichtsjahr 2008 wurde nicht mit einbezogen, da uns aufgrund der begrenzten Speicherdauer von Einzeldaten in der Außenhandelsstatistik hierfür keine vollständigen Außenhandelsdaten vorliegen.

9 Dabei handelt es sich um Unternehmen, die sowohl importieren als auch exportieren.

2012 nahezu konstant. Unternehmen, die im Außenhandel aktiv sind, verzeichneten demgegenüber eine wesentlich höhere Beschäftigungsdynamik. Insbesondere die reinen Importeure konnten bis 2012 einen Beschäftigungszuwachs um knapp 20% verzeichnen. Etwas geringer fielen die Beschäftigungszuwächse bei den reinen Exporteuren und den Two-Way-Tradern mit 10% beziehungsweise 17% aus. Eine mögliche Erklärung für die besonders starke Beschäftigungsentwicklung in importierenden Unternehmen könnte sein, dass diese von gesunkenen Einfuhrpreisen im Zuge der Weltwirtschaftskrise profitieren konnten.¹⁰ Um jedoch die Zusammenhänge zwischen Handelsbeziehungen mit Unternehmen im Ausland und Beschäftigungszuwächsen aussagekräftig untersuchen zu können, sind detailliertere Analysen notwendig, als sie im Rahmen dieses Artikels vorgestellt werden können.

5

Ausblick

Die Verknüpfung von Einzeldaten aus mehreren Datenquellen wird im Bereich der amtlichen Statistik in Zukunft verstärkt eine Rolle spielen, da auf diese Weise neue Erkenntnisse gewonnen werden können, ohne Unternehmen zusätzlich mit Auskunftspflichten zu belasten.

Mit dem EU-Projekt "Micro-data linking of Structural Business Statistics and other business statistics" 2014/2015 ist es gelungen, das Auswertungspotenzial vorhandener Unternehmensstatistiken zu erhöhen und die Möglichkeiten und Grenzen der Verknüpfung von Mikrodaten aus verschiedenen Datenquellen aufzuzeigen. Dabei wurden Einzeldaten der Berichtsjahre 2008 bis 2012 aus sechs¹¹ verschiedenen Erhebungen miteinander verknüpft und ausgewertet. Die Projektergebnisse werden in drei "Statistics explained"-Artikeln auf der Homepage von Eurostat veröffentlicht. Zum gegenwärtigen

10 Wagner und Weche Gelübcke (2014) finden für importierende Unternehmen im Zeitraum der globalen Rezession höhere Überlebenswahrscheinlichkeiten als für nicht Außenhandel betreibende Unternehmen. Als Grund hierfür sehen sie verbesserte Bedingungen für importierende Unternehmen auf Faktormärkten.

11 Nach der Projektausschreibung war es möglich, bis zu vier weitere Datenquellen, wie die Outward-FATS-Statistik oder den Community Innovation Survey (CIS), in die MDL-Datenbank mit einzubeziehen. Für einige teilnehmende Länder liegen deshalb zusätzlich Daten aus anderen Datenquellen vor.

gen Zeitpunkt ist bereits ein Artikel erschienen, in dem Statistiken für gruppenabhängige und -unabhängige kleine, mittlere und große Unternehmen präsentiert und erläutert werden (Airaksinen und andere, 2015).

In einem Nachfolgeprojekt wird der MDL-Datensatz um das Berichtsjahr 2013 erweitert. Zusätzlich werden die wirtschaftlichen Aktivitäten von schnell wachsenden Unternehmen als weiterer Analyseschwerpunkt im Fokus stehen. Zudem bietet das Nachfolgeprojekt die Möglichkeit, die Methodik weiterzuentwickeln. So soll die Hochrechnung der Einzeldaten künftig mit GREG durchgeführt werden.

Die im Rahmen des MDL-Projekts 2014/2015 verknüpften Daten bieten ein Analysepotenzial, das deutlich über die Möglichkeiten der einzelnen Statistiken hinausgeht. Zum Beispiel lassen sich Zusammenhänge zwischen ökonomischen Kennzahlen, wie Beschäftigungsentwicklung, Produktivität oder Rentabilität, und anderen Unternehmenscharakteristika untersuchen. Durch die Verknüpfung mit den Daten der Unternehmensdemografie besteht die Möglichkeit, Einflussfaktoren auf die Wahrscheinlichkeit des Überlebens im Markt zu bestimmen. Es ist vorgesehen, weitere ausgewählte Analysen aus dem MDL-Projekt zu veröffentlichen. 

LITERATURVERZEICHNIS

Afentakis, Anja/Bihler Wolf. *Das Hochrechnungsverfahren beim unterjährigen Mikrozensus ab 2005*. In: *Wirtschaft und Statistik*. Ausgabe 10/2005, Seite 1039 ff.

Airaksinen, Aarno/Luomaranta, Henri/Alajääskö, Pekka/ Roodhuijzen, Anton. *Dependent and independent SMEs and large enterprises*. *Statistics Explained*. 2015. Webdokument [Zugriff am 2. März 2016]. Verfügbar unter: ec.europa.eu

Allafi, Sabine. *Intrahandelsresultate nach Wirtschaftszweigen*. In: *Wirtschaft und Statistik*. Ausgabe 6/2011, Seite 546 ff.

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

Voshage, Ramona/Drafz, Ute/Habla, Heike/Klumpe, Matthias/Meisdrock, Claudia/Nowak, Kristin/Raab, Stefanie/Richter, Alexander/Rößner, Michael/Schmidtke, Kerstin. *Formal, faktisch oder absolut nachgefragt? Die Auswirkungen der Entgeltumstellung auf die Entwicklung der Nachfrage in den Forschungsdatenzentren der Statistischen Ämter des Bundes und der Länder*. FDZ-Arbeitspapier Nr. 49, Seite 14 ff. Wiesbaden 2015.

Nahm, Matthias/Söllner, René. *Die wirtschaftliche Bedeutung auslandskontrollierter Unternehmen in Deutschland 2012*. In: *Wirtschaft und Statistik*. Ausgabe 12/2014, Seite 754 ff.

Rauland, Jan/Bauer, Oliver. *Statistische Analyse des Einflusses von Informations- und Kommunikationstechnologien auf die Produktivität von Unternehmen*. In: *Wirtschaft und Statistik*. Ausgabe 12/2010, Seite 1066 ff.

Rink, Anke/Seiwert, Ines/Opfermann, Rainer. *Unternehmensdemografie: methodischer Ansatz und Ergebnisse 2005 bis 2010*. In: *Wirtschaft und Statistik*. Ausgabe 6/2013, Seite 422 ff.

Statistisches Bundesamt (Herausgeber). *Qualitätsbericht Außenhandel 2014*. Wiesbaden 2014. [Zugriff am 2. März 2016]. Verfügbar unter: www.destatis.de

Statistisches Bundesamt (Herausgeber). *Qualitätsbericht Nutzung von Informations- und Kommunikationstechnologien (IKT) in Unternehmen 2015*. Wiesbaden 2015. [Zugriff am 2. März 2016]. Verfügbar unter: www.destatis.de

Wagner, Ingo. *Schätzung fehlender Umsatzangaben für Organschaften im Unternehmensregister*. In: *Wirtschaft und Statistik*. Ausgabe 9/2004, Seite 1001 ff.

Wagner, Joachim/Weche Gelübcke, John P. *Risk or Resilience? The Role of Trade Integration and Foreign Ownership for the Survival of German Enterprises during the Crisis 2008-2010*. In: *Jahrbücher für Nationalökonomie und Statistik*. Band 234. Heft 6/2014, Seite 757 ff.

Herausgeber

Statistisches Bundesamt, Wiesbaden

www.destatis.de

Schriftleitung

Dieter Sarreither, Präsident des Statistischen Bundesamtes

Redaktionsleitung: Kerstin Hänsel

Redaktion: Ellen Römer

Ihr Kontakt zu uns

www.destatis.de/kontakt

Erscheinungsfolge

zweimonatlich, erschienen im April 2016

Das Archiv aller Ausgaben ab Januar 2001 finden Sie unter www.destatis.de/publikationen

Print

Einzelpreis: EUR 18,- (zzgl. Versand)

Jahresbezugspreis: EUR 108,- (zzgl. Versand)

Bestellnummer: 1010200-16002-1

ISSN 0043-6143

ISBN 978-3-8246-1044-0

Download (PDF)

Artikelnummer: 1010200-16002-4, 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: Metapaper Smooth, FSC-zertifiziert, klimaneutral, zu 61% aus regenerativen Energien

© Statistisches Bundesamt, Wiesbaden 2016

Vervielfältigung und Verbreitung, auch auszugsweise, mit Quellenangabe gestattet.