

Ronning, Gerd

Working Paper

Vertraulichkeit und Verfügbarkeit von Mikrodaten

IAW Diskussionspapiere, No. 101

Provided in Cooperation with:

Institute for Applied Economic Research (IAW)

Suggested Citation: Ronning, Gerd (2014) : Vertraulichkeit und Verfügbarkeit von Mikrodaten, IAW Diskussionspapiere, No. 101, Institut für Angewandte Wirtschaftsforschung (IAW), Tübingen

This Version is available at:

<https://hdl.handle.net/10419/92992>

Standard-Nutzungsbedingungen:

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

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

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

Terms of use:

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

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

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

Vertraulichkeit und Verfügbarkeit von Mikrodaten

Gerd Ronning

Institut für Angewandte Wirtschaftsforschung e.V.
Ob dem Himmelreich 1 | 72074 Tübingen | Germany
Tel.: +49 7071 98960 | Fax: +49 7071 989699

ISSN: 1617-5654

IAW-Diskussionspapiere

Das Institut für Angewandte Wirtschaftsforschung (IAW) Tübingen ist ein unabhängiges außeruniversitäres Forschungsinstitut, das am 17. Juli 1957 auf Initiative von Professor Dr. Hans Peter gegründet wurde. Es hat die Aufgabe, Forschungsergebnisse aus dem Gebiet der Wirtschafts- und Sozialwissenschaften auf Fragen der Wirtschaft anzuwenden. Die Tätigkeit des Instituts konzentriert sich auf empirische Wirtschaftsforschung und Politikberatung.

Dieses IAW-Diskussionspapier können Sie auch von unserer IAW-Homepage als pdf-Datei herunterladen:

<http://www.iaw.edu/Publikationen/IAW-Diskussionspapiere>

ISSN 1617-5654

Weitere Publikationen des IAW:

- IAW-News (erscheinen 4x jährlich)
- IAW-Forschungsberichte

Möchten Sie regelmäßig eine unserer Publikationen erhalten? Dann wenden Sie sich bitte an uns:

IAW e.V. Tübingen, Ob dem Himmelreich 1, 72074 Tübingen,
Telefon 07071 / 98 96-0
Fax 07071 / 98 96-99
E-Mail: iaw@iaw.edu

Aktuelle Informationen finden Sie auch im Internet unter:

<http://www.iaw.edu>

Der Inhalt der Beiträge in den IAW-Diskussionspapieren liegt in alleiniger Verantwortung der Autorinnen und Autoren und stellt nicht notwendigerweise die Meinung des IAW dar.

Vertraulichkeit und Verfügbarkeit von Mikrodaten

GERD RONNING *

Dezember 2013

Zusammenfassung

Der Vortrag beschreibt die Entwicklungen der letzten zehn Jahre bezüglich des Zugangs zu amtlichen Mikrodaten, insbesondere Unternehmens- und Betriebsdaten. Dabei werde ich vor allem auf Ergebnisse aus drei Projekten eingehen, an denen ich selbst mitgewirkt habe: (a) Anonymisierung von Querschnittsdaten, (b) Anonymisierung von Paneldaten und (c) Datenzugang über "Remote Access". Alle drei Projekte wurden gemeinsam von Mitarbeitern der Statistischen Ämter und des IAB sowie von Wissenschaftlern aus dem Bereich der empirischen Wirtschaftsforschung bearbeitet. Mein Vortrag spiegelt naturgemäß eher die Sicht eines Daten-Nutzers als die eines Datenschützers wider. Als Resümee ist festzuhalten, daß die - datenverändernde - Anonymisierung von Daten für Unternehmen und Betriebe einen Zugang zu diesen Mikrodaten schafft, der nicht ohne Probleme ist. Allerdings sind diese Daten im Zusammenhang mit dem Fernrechnen ("remote access"), das als bester Zugangsweg anzusehen ist, als "Strukturfiles" einsetzbar.

Stichwörter: Faktische Anonymität, Scientific Use Files, stochastische Überlagerung, (Multiple) Imputation, Datenstrukturfiles, Enthüllungsrisiko, Input-, Prozeß und Output-Kontrolle.

*Wirtschafts- und Sozialwissenschaftliche Fakultät, Fachbereich Wirtschaftswissenschaften, Universität Tübingen, D-72074 Tübingen. email: gerd.ronning@uni-tuebingen.de. Vortrag beim Wirtschaftswissenschaftlichen Seminar Ottobeuren, September 2013. Ich danke Jörg Drechsler und Stefan Bender, IAB Nürnberg, Martin Rosemann, ISG Berlin, und Hans Schneeweiß, Universität München, für Kommentare zu einer früheren Fassung dieses Vortrags. Außerdem hat mich Stefan Bender dankenswerterweise über den aktuellen Stand des Fernrechnens im FDZ des IAB informiert.

Inhaltsverzeichnis

1	Einleitung	3
2	Amtliche Mikrodaten als Scientific Use Files	5
2.1	Operationalisierung der 'Faktischen Anonymität'	6
2.2	Anonymisierungsverfahren für stetige Merkmale	6
2.3	Anonymisierung für diskrete, insbesondere nominale Merkmale	9
2.4	Korrekturschätzer	10
2.5	Kann man SUF-basierte empirische Ergebnisse publizieren ? . .	10
3	Fernrechnen (Remote Access)	11
3.1	Input- und Prozeß-Kontrolle	12
3.2	Enthüllungsrisiken statistischer Prozeduren	12
3.3	Faktische Anonymität von Output statt Daten	14
3.4	Automatisiertes Fernrechnen im IAB	14
4	Ein persönliches Resumee	15
5	Literatur	16

1 Einleitung

Mikrodaten für die wissenschaftliche Forschung sind seit langem in den experimentellen Wissenschaften verfügbar, vor allem in der Biologie, aber auch in den Sozialwissenschaften, vor allem in der Psychologie und in der Soziologie. Im Bereich der Wirtschaftswissenschaften waren es vor allem die Marktforschungsinstitute, die Daten für einzelne Unternehmen erhoben. Der seit 1949 bestehende Konjunkturtest des ifo-Instituts dürfte eine der ersten Erhebungen gewesen sein, die Informationen über die Entwicklung einzelner Branchen oder der Wirtschaft eines Landes insgesamt erhoben haben. Allen diesen Daten war gemeinsam, daß sie, soweit sie von Personen, Haushalten oder Unternehmen stammten, auf freiwilliger Basis erhoben wurden. Im Gegensatz dazu wurden von staatlichen Stellen Daten in der Art erhoben, daß ihre Bereitstellung verpflichtend war und ist. Heute spricht man in Deutschland von der "Amtlichen Statistik", deren Arbeit, vor allem die Durchführung von Erhebungen, aber auch die Weitergabe dieser Daten an Dritte, durch gesetzliche Vorschriften geregelt ist.

Natürlich wurde (und wird) auch bei Daten, die auf freiwilliger Basis gewonnen wurden, auf Vertraulichkeit geachtet. Mit der Tatsache, daß eine Preisgabe von Information gesetzlich angeordnet wird und bei Zuwiderhandlung mit Strafen belegt werden kann, bekommt die Vertraulichkeit jedoch eine ganz andere Qualität: Wenn keine Weigerung möglich ist, dann muß auch entsprechend sicher sein, daß die preisgegebenen Daten für einzelne Personen, Haushalte oder Unternehmen nicht an Dritte weitergegeben werden. Deshalb regelt das "Gesetz über die Statistik für Bundeszwecke oder kurz "Bundestatistik-Gesetz" vom Januar 1987 sehr penibel die Weitergabe dieser Daten. Eine Sonderregelung erfahren die Daten, die von der Bundesagentur für Arbeit erhoben werden, durch das Sozialgesetzbuch (SGB).¹

Obwohl die wissenschaftliche empirische Forschung naturgemäß nicht an individuellen Informationen, sondern an Aussagen über Gesamtheiten interessiert ist, kann keineswegs ausgeschlossen werden, daß die Daten mit dem Ziel der Enthüllung von Einzeldaten ausgewertet werden. Deshalb bestand lange Zeit für Forscher keine Möglichkeit, an die Einzel- bzw. Mikrodaten heranzukommen. Allenfalls Tabellen oder Aggregate wurden von der Amtlichen Statistik zur Verfügung gestellt. Die Daten, die ganz offensichtlich umfangreiche Information über die wirtschaftliche und soziale Situation in Deutschland bedeuten, lagen also ungenutzt in den Archiven der Statistischen Ämter. Mit dem jetzt gültigen Statistik-Gesetz von 1987 wurde diese seit langem kritisierte strikte Nicht-Weitergabe aufgelockert: Die Forscher können nun "faktisch anonymisierte" Mikro-Daten von den Statistischen Ämtern erwerben. Diese Datensätze werden auch als "Scientific Use Files"(SUF) bezeichnet. Dabei ist dieser Begriff, der übrigens im Gesetz gar nicht auftaucht, wie folgt - indirekt - definiert:²

¹Siehe beispielsweise Hochfellner u.a. (2012, Abschnitt 2.1).

²Absatz 6 von §16 BStatG hat folgenden Wortlaut: " (6) Für die Durchführung wissenschaftlicher Vorhaben dürfen vom Statistischen Bundesamt und den statistischen Ämtern der Länder Einzelangaben an Hochschulen oder sonstige Einrichtungen mit der Aufgabe un-

Einzelangaben können weitergegeben werden, wenn sie "...nur mit einem unverhältnismäßig großen Aufwand an Zeit, Kosten und Arbeitskraft zugeordnet werden können".

Im Jahr 1999 wurde vom Wissenschaftsministerium eine "Kommission zur Verbesserung der informationellen Infrastruktur zwischen Wissenschaft und Statistik" (KVI) eingerichtet, die 2001 ihr umfangreiches Gutachten vorlegte³ und u.a. folgende Empfehlungen aussprach:

- die Einsetzung des Rates für Wirtschafts- und Sozialdaten (RatSWD): Der Rat wurde 2004 vom Bundesministerium für Bildung und Forschung mit der Zielsetzung eingerichtet, die Forschungsdateninfrastruktur für die empirische Forschung nachhaltig zu verbessern und somit zu ihrer internationalen Wettbewerbsfähigkeit beizutragen.
- die Einrichtung von Forschungsdatenzentren (FDZ'en): Den Anfang machten das Statistische Bundesamt und die Statistischen Landesämter, die Rentenversicherung und die Bundesagentur für Arbeit/IAB. Insgesamt 25 Datenzentren wurden bis Anfang 2013 vom RatSWD akkreditiert. Gemeinsam ermöglichen sie der Wissenschaft einen klar definierten Zugang zu den Mikodaten der verschiedenen Institutionen.
- die Finanzierung von faktisch anonymisierten Mikrodaten (Scientific Use Files) als Zugang zu den Mikrodaten der Amtlichen Statistik. Über die Forschungsdatenzentren sind heute auch andere Formen des Zugangs zu ausgesuchten Datenbeständen der amtlichen Statistik möglich. Die verschiedenen Optionen sind wie folgt:
 1. Nutzung der Mikrodaten vor Ort (On-Site-Nutzung): Arbeitsplätze für Gastwissenschaftler, Kontrollierte Datenfernverarbeitung
 2. Nutzung der Daten außerhalb des Datenhalters (Off-Site-Nutzung): Scientific-Use-Files, Public-Use-Files, CAMPUS-Files

Demnach sind inzwischen Scientific-Use-Files nur noch **eine (!)** mögliche Art, Mikrodaten zu nutzen.

abhängiger wissenschaftlicher Forschung übermittelt werden, wenn die Einzelangaben nur mit einem unverhältnismäßig großen Aufwand an Zeit, Kosten und Arbeitskraft zugeordnet werden können und die Empfänger Amtsträger, für den öffentlichen Dienst besonders Verpflichtete oder Verpflichtete nach Absatz 7 sind." Und dieser Absatz 7 lautet wie folgt: "(7) Personen, die Einzelangaben nach Absatz 6 erhalten sollen, sind vor der Übermittlung zur Geheimhaltung zu verpflichten, soweit sie nicht Amtsträger oder für den öffentlichen Dienst besonders Verpflichtete sind. "

³Kommission zur Verbesserung der informationellen Infrastruktur zwischen Wissenschaft und Statistik (Hrsg.): Wege zu einer besseren informationellen Infrastruktur Gutachten der vom Bundesministerium für Bildung und Forschung eingesetzten Kommission zur Verbesserung der informationellen Infrastruktur zwischen Wissenschaft und Statistik. Nomos Verlagsgesellschaft, Baden-Baden 2001

2 Amtliche Mikrodaten als Scientific Use Files

Initiativen zur Erzeugung von Scientific Use Files (SUF's) gab und gibt es immer noch in großer Zahl; allerdings war die Kommunikation zwischen den einzelnen Projekten sowie die Dokumentation der Ergebnisse - für mich - eher unbefriedigend, beispielsweise bezüglich einer Initiative von Eurostat, weil kaum Informationen weitergegeben wurden.⁴ Eine allgemein gültige Definition für SUF's ist nicht verfügbar. Üblicherweise versteht man darunter einen Satz von Mikro-Daten, der derart anonymisiert ist, daß einerseits die Vertraulichkeit gewahrt und andererseits der statistische Gehalt der Daten nicht allzusehr gestört bzw. zerstört ist. In Deutschland setzt man, wie bereits in Abschnitt 1 erwähnt, diese Files auch mit "faktisch anonymisierten" Datenfiles gleich, im Gegensatz zu Public Use Files (PUF's) , die im deutschen Sprachraum auch mit "absolut anonymisierten" Datensätzen gleichgesetzt werden. Letztere sind in weit stärkerem Maße anonymisiert⁵ und werden beispielsweise in Lehrveranstaltungen als Beispiels-Datensätze verwendet.

Im Jahr 2003 begann ein vom Wissenschaftsministerium finanziertes Projekt mit dem Titel "Faktische Anonymisierung wirtschaftsstatistischer Einzeldaten", für das sich projektintern das Kürzel FAWE eingebürgert hat. Dabei ging es zunächst um die Bereitstellung von Querschnittsdaten. In einem Anschlußprojekt mit dem Titel "Wirtschaftsstatistische Paneldaten und faktische Anonymisierung " (2005 - 2008, Akkürzung FAWE-PANEL) wurde dann eine Erweiterung der Untersuchung auf Paneldaten vorgenommen. In beiden Projekten ging es, wie der Projekttitel schon andeutet, um Mikrodaten für Unternehmen und Betriebe.⁶ Vor allem im ersten Projekten ging es um folgende Fragen:

- Wie kann man faktische Anonymität in der Praxis definieren ??
- Welche Verfahren der Anonymisierung sind geeignet und welche (besonders) ungeeignet ??
- Gibt es Verfahren, mit denen man mögliche Verzerrungen der Ergebnisse korrigieren kann ?

⁴Ich verzichte in meinem Vortrag auf jegliche Aufzählung von derartigen Projekten. Vor allem im internationalen Kontext würde dies für einen eigenen Vortrag (mehr als) ausreichen.

⁵Eine genaue Abgrenzung liefert das Statistik-Gesetz.

⁶Die folgenden Ausführungen gehen nicht auf die Anonymisierung von Mikrodaten für Personen und Haushalte ein. Die im Folgenden betrachteten Verfahren bezeichnet man als "datenverändernd" im Gegensatz zu einfachen, informationsreduzierenden Verfahren, die sich bei Personen- und Haushaltsdaten bewährt haben.

2.1 Operationalisierung der 'Faktischen Anonymität'

Beim Begriff der - faktischen - Anonymität geht es um die Frage, in welchem Umfang und mit welchem Aufwand ein sogenannter Daten-Angreifer Informationen für einzelne Personen, Haushalte oder Unternehmen aus einem Datensatz herausfiltern kann. Natürlich verfügt der Datensatz über keine identifizierende Information wie Name oder Adresse. Gleichwohl kann der Angreifer mit seinem sogenannten "Zusatzwissen" versuchen, vertrauliche Information aus dem Datensatz zu bekommen. Wenn beispielsweise Umsatz und Beschäftigung im Datensatz - in anonymisierter, also verfremdeter Form - enthalten sind, die wahre Beschäftigung aber auch als Zusatzwissen bzw. als "Überschneidungsmerkmal" zur Verfügung steht, dann kann damit versucht werden, den Umsatz für ein bestimmtes Unternehmen abzuschätzen. Wohlgedenkt, der gefundene Wert ist nicht der wahre Umsatz, wohl aber ein Wert, der diesem "nahe kommt". Je niedriger die Wahrscheinlichkeit ist, daß eine solche Enthüllung erfolgt, desto geringer ist das sogenannte "Enthüllungs-Risiko" (disclosure risk) und umso größer ist das Ausmaß an Anonymität. Wenn nun außerdem die Wahrscheinlichkeit, daß der anonymisierte Wert genügend stark von dem wahren Wert differiert, hoch ist, dann soll dies ebenfalls als eine Verstärkung der Anonymität angesehen werden. Dies führte dann zu einer im Projekt FAWE entwickelten (indirekten) Definition der (faktischen) Anonymität:⁷

Faktische Anonymität ist gegeben, wenn die Wahrscheinlichkeit für die Enthüllung eines Einzelwertes genügend gering und die Wahrscheinlichkeit für eine genügend große Abweichung vom wahren Wert, vorausgesetzt eine Identifikation sprich Enthüllung hat stattgefunden, genügend groß ist. Wenn beispielsweise der Umsatz eines Unternehmens aus dem SUF mittels Zusatzwissen identifiziert wird, der gefundene Wert des Umsatzes aber um mehr als 10 Prozent vom wahren Wert abweicht, dann ist die Anonymität gewahrt!

Natürlich ist diese Definition, soweit sie eine "Abweichung vom wahren Wert" betrachtet, streng genommen nur für stetige Merkmale verwendbar. Vor allem bei qualitativen sprich nominalen Merkmalen ist dieses zweistufige Konzept nicht einsetzbar. Der Unterabschnitt 2.3 beschreibt deshalb, welche speziellen Verfahren bei solchen Merkmalen zur Wahrung der Anonymität eingesetzt werden sollten.

2.2 Anonymisierungsverfahren für stetige Merkmale

Bei stetigen sprich metrischen Merkmale stehen verschiedene Anonymisierungsverfahren zur Verfügung, von denen die wichtigsten hier kurz genannt werden sollen:

⁷Für Details und genau Quellenangaben siehe Handbuch der Anonymisierung wirtschaftsstatistischer Mikrodaten, 2005, Kapitel 12.

- Mikroaggregation: Es wird für jede Untersuchungseinheit nur der Mittelwert einer Teilmenge von Untersuchungseinheiten angegeben. Dies Verfahren ist eng mit der Cluster-Analyse verwandt.
- Stochastische Überlagerung: Dabei wird der Originalwert durch eine Zufallsvariable - additiv oder multiplikativ - überlagert. Dies Verfahren entspricht in der Ökonometrie einem Model mit Meßfehlern ("Fehler-in-den-Variablen-Modell"). Allerdings wird dort vorwiegend mit einer additiven Fehlerüberlagerung gearbeitet, während sich für die Anonymisierung insbesondere von Unternehmensdaten die multiplikative Variante weit besser eignet. Allerdings ist dieses Verfahren nur bei stetigen Merkmalen einsetzbar.
- Multiple Imputation: Dies Verfahren, das auf einen Vorschlag von Donald Rubin zurückgeht, erzeugt einen Datensatz (oder auch mehrere), der die anonymisierten Daten enthält. Dabei betrachtet Rubin alle Werte als fehlend und ersetzt bzw. imputiert sie durch geschätzte Werte. Im Verfahren zur Multiplen Imputation werden dabei mehrere Werte für jeden fehlenden Wert erzeugt. Im Gegensatz zur Überlagerung kann das Imputations-Verfahren auch bei diskreten Merkmalen verwendet werden.

Im FAWE-Projekt wurde das Verfahren der Mikroaggregation als zu grob bewertet und deshalb nur am Rande betrachtet. Für das Verfahren der Imputation ergaben sich erst in letzter Zeit vorzeigbare Ergebnisse. Beispielsweise liegen SUF-Datensätze im FDZ des IAB vor, die mit dem Verfahren der Multiplen Imputation anonymisiert wurden.⁸ Als besonders attraktiv wurde die stochastische Überlagerung angesehen, bei der der jeweilige Originalwert (des stetigen Merkmals) mit einem Zufallsfehler überlagert wird. Wenn wir mit x den Originalwert, mit u den Zufallsfehler und mit x^a den resultierenden anonymisierten Wert bezeichnen, dann lassen sich folgende zwei Varianten der stochastischen Überlagerung unterscheiden:

$$x^a = x + u, \quad \text{wobei } E[u] = 0 \quad (\text{additive Überlagerung})$$

$$x^a = x \cdot u, \quad \text{wobei } E[u] = 1 \quad (\text{multiplikative Überlagerung})$$

Dabei ist in der Anonymisierung die **multiplikative** Variante eindeutig vorzuziehen: In der additiven Variante werden kleine und große Werte mit demselben Zuschlag, beispielsweise 10 Euro, überlagert, während in der multiplikativen Variante die Umsätze mit einem prozentualen Zu- bzw. Abschlag überlagert werden: Sowohl kleine als auch große Umsätze werden um denselben Prozentsatz von beispielsweise 10 % verändert.

Schließlich wurde für die multiplikative Überlagerung noch ein spezieller Ansatz gewählt, in dem der Zufallsfehler nicht der üblichen - eingipfligen - Normalverteilung, sondern einer zweigipfligen Mischungsverteilung folgt. Dieser Ansatz wurde zuerst von Gina Marie Roque in ihrer Dissertation von 2000

⁸Siehe Drechsler u.a. (2008) sowie vor allem Drechsler (2012).

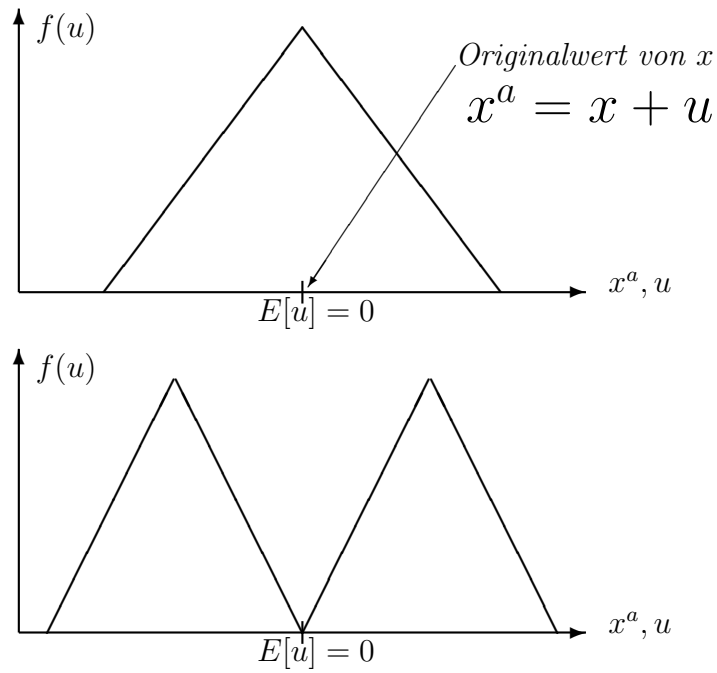


Abbildung 2/1: **Additive** Überlagerung (unimodal und Mischungsverteilung)

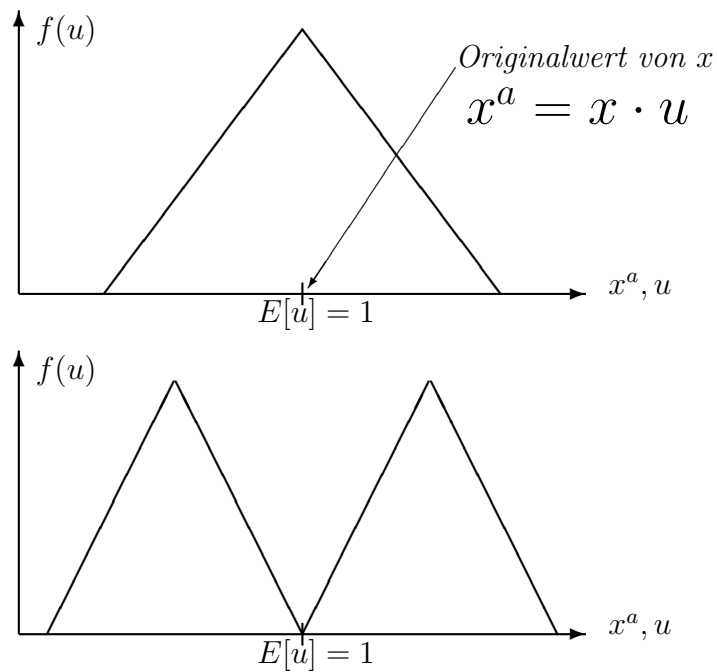


Abbildung 2/2: **Multiplikative** Überlagerung (unimodal und Mischungsverteilung)

vorgeschlagen, dann aber von Jörg Höhne, Statistisches Landesamt Berlin und Brandenburg, weiterentwickelt. Dieses Fehler-Modell läßt sich formal wie folgt schreiben:

$$u = \begin{cases} \delta D + \varepsilon & \text{(additive Überlagerung)} \\ 1 + \delta D + \varepsilon & \text{(multiplikative Überlagerung)} \end{cases}$$

mit $E[\varepsilon] = 0$, wobei für die diskrete Zufallsvariable D gilt:

$$D = \begin{cases} +1 & \text{mit Wahrscheinlichkeit } \alpha \\ -1 & \text{mit Wahrscheinlichkeit } 1 - \alpha \end{cases} .$$

Üblicherweise wird hier $\alpha = 0,5$ gewählt. In diesem Fall ergibt sich $E[D] = 0$ und damit $E[u] = 0$ im additiven und $E[u] = 1$ im multiplikativen Fall. Im letzteren Fall läßt sich der Parameter δ als prozentualer Zuschlag interpretieren. Die resultierende Verteilungen - für additive und multiplikative Variante - zeigen die Abbildungen 2/1 und 2/2: Im unteren Teil ist jeweils die Überlagerung mittels Mischungsverteilung skizziert, wobei in der Praxis statt der Dreiecksverteilung üblicherweise die Normalverteilung verwendet wird.

2.3 Anonymisierung für diskrete, insbesondere nominale Merkmale

Ein Großteil der Merkmale in der Amtlichen Statistik ist nicht stetig: Es gibt Merkmale wie Beruf, Tätigkeitsstatus, Region, Branchen oder auch die Information, ob jemand zur Umsatzsteuer optiert oder nicht. Nicht nur in den genannten FAWE-Projekten, sondern auch in vielen anderen Untersuchungen ist deutlich geworden, daß von den nominalen Merkmalen ein besonders großes Enthüllungs-Risiko ausgeht. Wenn man ein Unternehmen in einer bestimmten Branche und einer bestimmten Region vorgibt und noch ein bißchen Zusatzinformation hinzufügt, dann ist die Wahrscheinlichkeit, das gesuchte Unternehmen im Datensatz auch zu finden, sehr groß.

Es gibt eine Anonymisierungs-Methode für nominal-skalierte Merkmale, die man formal auch als stochastische Überlagerung ansehen kann. Sie wird als Post-Randomisierungs-Methode oder kurz PRAM bezeichnet. Allerdings ist sie für Merkmale mit vielen Ausprägungen und bei gleichzeitiger Anwendung auf **mehrere** nominale Merkmale derart komplex, daß sie sich in der Praxis der Anonymisierung nicht durchgesetzt hat. Statt dessen wird bei der Anonymisierung dieser Merkmale darauf geachtet, daß die Fallzahl in einer bestimmten "Zelle" sprich Merkmalskombination verschiedener nominaler Merkmale genügend groß ist. Falls nicht, werden verschiedene Kategorien sprich Ausprägungen zusammengelegt. Im äußersten Fall wird auf die Angabe der Ausprägung für die "kritischen" Unternehmen bzw. Haushalte ganz verzichtet. Dabei greift man in Deutschland in zunehmendem Maße auf das Verfahren SAFE zurück, an dem Jörg Höhne, STALA Berlin, maßgeblich mitgearbeitet hat und das auch in der jüngsten Volkszählung bei der Daten-Weitergabe eingesetzt wird.

In allerjüngster Zeit wird auch die Multiple Imputation, die bereits bei der Anonymisierung stetiger Merkmale erwähnt wurde (siehe Abschnitt 2.2), zur Anonymisierung von nominalen und ordinalen Merkmalen eingesetzt. Dabei wird für die sogenannte "Imputationsgleichung" ein Modell verwendet, das nominale und ordinale Variable als abhängige Variable spezifiziert, also insbesondere Logit- und Probitmodelle. Und für ganzzahlige Merkmale kann man auf die Zähldatenmodelle zurückgreifen, also insbesondere auf die Poisson-Regression.

2.4 Korrektorschätzer

Eine Besonderheit der beiden FAWE-Projekte waren Untersuchungen dazu, ob man die durch die Anonymisierung erzeugte Verzerrung korrigieren könnte. Diese Untersuchungen betrafen vorwiegend die stochastische Überlagerung, aber auch die Anonymisierung mit PRAM. Statistisch-ökonomisch wird mit der stochastischen Überlagerung ein Fehler in den Variablen-Modell erzeugt, das zu verzerrten Parameter-Schätzungen im Regressionsmodell führt, wenn die Regressor-Variablen mit einem Fehler behaftet sind. Milton Friedman hat als erster darauf hingewiesen. Allerdings kann man bei anonymisierten Daten davon ausgehen, daß die Fehler-Varianz **bekannt** ist, weil sie bei der Beschreibung der Anonymisierungsmethode mitgeteilt wird.⁹

Zu der Frage, wie man Korrektorschätzer im linearen Panel-Modell einsetzen kann, haben Hans Schneeweiß und ich in jüngster Zeit einige Ergebnisse publiziert.¹⁰ Ferner gibt es Ergebnisse, in denen die Simulation-Extrapolation-Methode (SIMEX) zur Schätzung verwendet wurde.¹¹ Schließlich wurde für das Probit-Modell gezeigt, daß eine konsistente Schätzung auch dann möglich ist, wenn die binäre abhängige Variable mittels PRAM anonymisiert wurde. Dabei wurde wieder vorausgesetzt, daß die Parameter der Anonymisierungsmethode, in diesem Fall also PRAM, bekannt sind.¹²

2.5 Kann man SUF-basierte empirische Ergebnisse publizieren ?

Die soeben angesprochene Korrektorschätzung mag man als akademisches Gehirntraining belächeln. Sicher ist, daß ein allgemeiner Einsatz bei anonymisierten Daten nicht in Frage kommt, weil die meisten Nutzer mit der Arbeitsweise der Korrektorschätzer überhaupt nicht vertraut sind. Dann bleibt die Frage, ob es einem Autor gelingt, einen Artikel zu publizieren, in dem die empirischen Ergebnisse aus einem Scientific Use File gewonnen wurden. Und insbesondere stellt sich die Frage bei den hier betrachteten datenverändernden Verfahren, die vor allem auf Unternehmensdaten angewendet werden. Eine ehrliche Debatte

⁹Darüber herrscht allerdings zwischen den Statistischen Ämtern und sonstigen Datenlieferanten keineswegs Einigkeit!

¹⁰Siehe Schneeweiß und Ronning(2010) und Ronning und Schneeweiß (2011).

¹¹Siehe Ronning und Rosemann (2008).

¹²Siehe Ronning (2005) sowie Ronning, Rosemann und Strotmann (2005).

zu dieser Frage hat meines Wissens nie stattgefunden. Der Grund dafür könnte sein, daß überwiegend von vornherein diese Möglichkeit als vollkommen utopisch angesehen wird. Auch in den FAWE-Projekten wurde diese Frage wenn überhaupt nur kurz in Debatten angeschnitten.

Wenn ein SUF nur schwach anonymisiert ist, sind die verzerrenden Auswirkungen gering, andererseits ist das Enthüllungsrisiko groß. Deshalb werden die Lieferanten der Daten, vor allem die Statistischen Ämter, darauf bestehen, daß die Anonymisierung nicht zu schwach gewählt wird. Dies gilt auch für die in den FAWE-Projekten erzeugten SUF's. Dann aber ist die Verzerrung der Ergebnisse so bedeutsam, daß - nach meiner ganz persönlichen Meinung - nur eine Publikation mit Originaldaten akzeptabel ist.

Im Projekt *infinite* wurde bei der Diskussion über die Erzeugung von Datenstruktur-Files auch die Frage erörtert, wie man die Annoymität dieser Datensätze steuern könne. Für die stochastische Überlagerung ist dies mittels der Varianz des Zufallsfehlers, im Höhne-Verfahren zusätzlich über die Variation des Zuschlag-Parameters δ , relativ einfach möglich. Auch für die Mikroaggregation kann durch Variation der Gruppen-Größe die Annoymität gesteuert werden. Dagegen ist im Fall der Multiplen Imputation, die auf eine Modell-Struktur zwischen den einzelnen Variablen zurückgreift, diese Frage nicht einfach zu beantworten. Ganz allgemein gilt in diesem Verfahren, daß die Anonymisierung umso geringer ist, je besser der Fit der dabei verwendeten Imputationsgleichungen ist.

Ergebnis der beiden Projekte zur Erzeugung von Scientific Use Files ist also, daß die ursprüngliche Idee, diese Daten für die empirische Forschung zu verwenden, nicht oder nur schlecht realisierbar ist. Allerdings hat sich bei Untersuchung des alternativen Zugriffs auf Mikrodaten über das Fernrechnen, über den jetzt zu reden sein wird, herausgestellt, daß die SUF's bzw. die dabei verwendeten Methoden eine wichtige Rolle bei diesem Zugangsweg spielen.

3 Fernrechnen (Remote Access)

Schon seit langem ist klar, daß das Fernrechnen, bei dem der Nutzer einen Programm-Code an das Rechenzentrum des jeweiligen Datenlieferanten sendet und ihm die resultierenden Schätzergebnisse praktisch ohne Verzögerung zugesandt werden, der aus Nutzer-Sicht optimale Zugangsweg zu Mikrodaten ist.¹³ Natürlich müssen dafür die technischen Voraussetzungen geschaffen werden. Ein größeres Hindernis schien jedoch lange Zeit der immense Prüfaufwand zu sein, der dabei entsteht. Denn bei Zusendung eines Programm-Codes ist ja keineswegs ausgeschlossen, daß die Vertraulichkeit einzelner Daten verletzt wird:

- Jemand fordert mit einem LIST-Befehl alle Werte eines oder mehrerer

¹³Seit langem gibt es die Möglichkeit, als Gastwissenschaftler in einem Statistischen Amt mit den Mikrodaten zu rechnen. Diese Variante ist aber zeit- und kostenaufwendig!

Merkmale an.

- Es wird das arithmetische Mittel für alle N Beobachtungen im Datensatz, dann aber auch das arithmetische Mittel für $N - 1$ Beobachtungen angefordert. Durch Differenzenbildung erhält man einen Einzelwert.
- Wenn man die geschätzten Werte der abhängigen Variablen und zusätzlich die Residuen anfordert, dann kann man durch Addition die Beobachtungswerte der abhängigen Variablen bestimmen.
- In einem Box-Plot werden üblicherweise die Extremwerte ausgegeben, die aber als Einzelwerte nicht offengelegt werden können.

Offensichtlich sind viel Fachwissen und ein großer Zeitaufwand notwendig, um solche mißbräuchlichen Anfragen ("queries") zu unterbinden! Im Folgenden soll etwas genauer geschildert werden, wie die FDZ's, an die diese Anfragen gerichtet werden, versuchen, den Aufwand und das Risiko bei Kontrolle des Outputs, der aus der Anfrage resultiert, zu minimieren.

3.1 Input- und Prozeß-Kontrolle

Eine erste Einschränkung eines möglichen Mißbrauchs ist die Tatsache, daß nur mit einer Auswahl von statistischen Softwarepaketen gerechnet werden kann, vor allem mit STATA, SPSS oder auch SAS. Eigene in FORTRAN oder GAUSS geschriebene Programme haben also keine Chance! Zudem muß der Programm-Code bestimmte Formvorschriften erfüllen, insbesondere eine ausreichende Kommentierung der einzelnen Befehle enthalten. Teilweise werden auch bestimmte Befehle, z.B. LIST, nicht akzeptiert. Im Projekt infinitE wurde diese "Inputkontrolle" in der Art systematisiert, daß eine Liste von nicht zulässigen Befehlen erarbeitet wurde. In der anschließenden "Prozeßkontrolle" werden Ergebnisse, die die Anonymität verletzen könnten, auf ihr Enthüllungsrisiko überprüft. Beispielsweise werden vom FDZ des IAB Auswertungen, die auf einer zu geringen Fallzahl ($n \leq 20$) basieren, nicht im Output dokumentiert.¹⁴ Deshalb wird gefordert, die Fallzahl für eine bestimmte statistische Kennzahl oder Analyse jeweils mit auszugeben.

Besonders argwöhnisch sind die Kontrolleure bei der Verwendung von Dummy-Variablen. Daß dies seine Berechtigung hat, wird im folgenden Unterabschnitt demonstriert.

3.2 Enthüllungsrisiken statistischer Prozeduren

In der Arbeit von Gomatam et al(2005) werden einige grundsätzliche Aspekte des Enthüllungsrisikos beim Fernrechnen aufgeführt. Hier soll das nur an einigen Beispielen, an denen ich selbst mitgearbeitet habe, demonstriert werden:

¹⁴Siehe Hochfellner u.a. (2012).

Stratgische Dummy-Variable Wenn man aus dem Zusatzwissen eine Dummy-Variable konstruiert, die dafür sorgt, daß nur ein einziges Unternehmen für diese Variable den Wert 1 hat, dann ergibt sich aus einer linearen Regression, in der diese Dummy-Variable als erklärende Variable berücksichtigt wird, für das betrachtete Unternehmen i^*

$$\hat{y}_{i^*} = y_{i^*} \quad .$$

Das heißt, es wird der Wert der abhängigen Variablen für dieses Unternehmen offenbart. Verwendet man nacheinander verschiedene Merkmale als abhängige Variable, so wird für alle diese Merkmale Information des Unternehmens preisgegeben.¹⁵

Gesättigte Modelle Eng verbunden mit dem zuvor erwähnten Risiko ist die Verwendung von sogenannten gesättigten Modellen (saturated models), in denen auf der rechten Seite nur binäre Regressoren (Dummies samt all ihren Interaktionen) spezifiziert sind. Für statistisch Versierte ist im Fall der (linearen) Varianzanalyse bekannt, daß die geschätzten Werte mit den beobachteten Werten der (stetigen) abhängigen Variablen identisch sind, d.h. es gilt

$$\hat{y}_i = y_i \quad \text{für alle } i \quad .$$

Weniger bekannt ist, daß dies auch für die wichtigsten nicht-linearen mikro-ökonomischen Modelle wie Logit und Probit sowie die Poissonregression und das NEGBIN-Modell gilt, worauf ich in einer noch unveröffentlichten Arbeit hingewiesen habe.¹⁶

Factor Scores Die meisten Arbeiten zu Enthüllungsrisiken beschäftigen sich mit linearen Modellen. Weitgehend unbeachtet blieb bisher die Multivariate Analyse, vor allem die Faktoren-Analyse. In einer Arbeit gemeinsam mit Jörg Drechsler und Philipp Bleninger habe ich gezeigt, daß man in einer Faktorenanalyse den Merkmalsvektor für ein bestimmtes Merkmal x^0 aus den Factor Scores bestimmen kann, wenn man dafür sorgt, daß alle anderen Merkmale, die in die Faktorenanalyse eingehen, mit dem Merkmal x^0 (fast) unkorreliert sind. Und zwar gilt dies für alle Varianten, nach denen die Factor Scores berechnet werden können (Kleinst-Quadrate, Bartlett und Thomson). In der Arbeit dazu haben wir an einem empirischen Beispiel (IAB-Betriebspanel) gezeigt, daß die Bedingung der geringen Korrelation dadurch erreicht werden kann, daß man aus einem großen Datensatz diejenigen Merkmale aussucht, deren Korrelation mit x^0 genügend gering ist.¹⁷

¹⁵Siehe Bleninger, Drechsler und Ronning (2011).

¹⁶Siehe Ronning (2011).

¹⁷Siehe Drechsler, Ronning und Bleninger (2013/2014).

3.3 Faktische Anonymität von Output statt Daten

Wenn die Nutzer des Fernrechnens die Mikrodaten gar nicht zu Gesicht bekommen¹⁸, dann stellt sich die Frage, wie die Ergebnisse in den an den Nutzer gelieferten Output-Files aus Sicht des Statistik-Gesetzes zu bewerten sind. In dem dritten Projekt, an dem ich beteiligt war, mit dem Titel "Eine informationelle Infrastruktur für das E-Science Age"¹⁹ oder kurz **infinite** hat sich der Begriff "Faktische Anonymität von Ergebnissen" herausgebildet, der auch in anderen Projekten verwendet wird. Allerdings ist unklar, wie man beispielsweise festlegen kann, ob ein bestimmter Regressions-Koeffizient diese Anonymitäts-Forderung erfüllt. Nach meiner Ansicht fehlt dazu bisher eine konkrete Vorschrift bzw. Definition.

Wie bereits erwähnt, werden beim Fernrechnen Datenstruktur-Files, also SUF's, als Probiermaterial eingesetzt, um eine sinnvolle Spezifikation des zu schätzenden Modells zu bestimmen. Im Projekt infinite wurde die Meinung vertreten, daß Ergebnisse, z.B. Regressionskoeffizienten, die aus diesen Daten bestimmt werden, ebenfalls SUF bzw. faktisch anonym sind. Aber das löst nicht das Problem, daß für die endgültige Analyse die Originaldaten verwendet werden und damit offen bleibt, wie die gefundenen Schätzergebnisse bezüglich Anonymität zu bewerten sind.

3.4 Automatisiertes Fernrechnen im IAB

Als besonders fortschrittlich betrachte ich das Angebot des IAB zum Fernrechnen.²⁰ Nicht nur ist die Dokumentation auf der Internet-Seite zu den Anforderungen für den Zugriff auf die Mikrodaten des IAB vorbildlich. Meines Wissens stellt das IAB als erstes FDZ auch auf eine automatisierte Version²¹ des Fernrechnens um, wobei ausschließlich STATA als Software-Paket genutzt werden kann:

- Grundgedanke des Ansatzes ist, daß ein Großteil des Outputs, der an den Nutzer geliefert wird, nicht zur Veröffentlichung bestimmt ist. Dieser Teil des Outputs wird als "intern" bezeichnet.
- Der interne Output wird dem Nutzer in einem Editor unmittelbar nach Analyseende zur Verfügung gestellt. Der Output durchläuft eine automatische Outputkontrolle und wird nach dem STATA-Lauf in das Verzeichnis des Nutzers kopiert. Der interne Output darf nicht von diesem Editor runtergeladen werden und darf nur projektintern verwendet werden. Zu den Schutzmechanismen gehört - neben Wasserzeichen und der

¹⁸Auch eine Darstellung auf dem Bildschirm ist wegen der Möglichkeit, diesen zu fotografieren, nicht zulässig.

¹⁹Als Untertitel ist dem Projekttitle hinzugefügt: "Auf dem Weg zu Remote Access. Verbesserung der kontrollierten Datenfernverarbeitung bei wirtschaftsstatistischen Daten durch Datenstrukturfiles und automatisierte Ergebniskontrolle."

²⁰siehe www.fdz.iab.de sowie Hochfellner u.a. (2012).

²¹Die Informationen dazu verdanke ich Stefan Bender. Mail vom 6. September 2013.

Unmöglichkeit, den Output zu drucken - auch eine vertragliche Bindung. Dazu werden alle bestehenden Verträge der Nutzer zur Zeit (September 2013) aktualisiert.²²

- Demgegenüber werden diejenigen Ergebnisse, die für eine Veröffentlichung verwendet werden, als "externer" Output bezeichnet. Der externe Output wird wie bisher kontrolliert und kann anschließend auch veröffentlicht werden.²³

Ohne daß dies explizit erwähnt wird, ist damit klar, daß in diesem Verfahren **immer** die Originaldaten verwendet werden. Die anonymisierten Datensätze (Datenstruktur-Files) spielen hier also überhaupt keine Rolle mehr.

4 Ein persönliches Resumee

Scientific Use Files, insbesondere solche, die mit datenverändernden Verfahren erzeugt wurden, sind für die wissenschaftliche empirische Forschung nur bedingt von Nutzen: Man bekommt durch sie einen "ungefähren" Eindruck von der zu untersuchenden Problematik. Das mag für Vorstudien zu der Frage, ob ein Datensatz von Interesse ist, relevant sein. Wenn es aber ans Publizieren oder Kommunizieren geht, werden vermutlich alle Forscher nur mit Ergebnissen, die auf Originaldaten basieren, an die Öffentlichkeit gehen.

Insofern ist es zu begrüßen, daß das Fernrechnen auch in Deutschland mit riesigen Schritten einem Status entgegenstürmt, bei dem der Forscher, der sich an den leicht zu akzeptierenden Kodex für das Fernrechnen, beispielsweise beim IAB, hält, ein komfortables Forschungsumfeld geliefert bekommt. Dabei ist die zunehmende internationale Vernetzung der deutschen FDZ's ein wichtiger Motor, der auch der internationalen Zusammenarbeit in der empirischen Forschung nützt.

²²Herr Bender ergänzt dies wie folgt: "Der Output hat somit den Stellenwert eines SUF's."

²³Herr Bender ergänzt dies wie folgt: "Der Output hat somit den Stellenwert eines PUF's."

5 Literatur

- Bleninger, P., Drechsler, J., und Ronning, G. (2011). Remote Data Access and the Risk of Disclosure from Linear Regression. *Statistics and Operations Research Transactions (SORT) 35 (2011)*, Special Issue: Privacy in Statistical Databases, September 2011, 7-24.
- Drechsler, J., Dundler, A., Rässler, S. und Zwick, T. (2008). A New Approach for Disclosure Control in the IAB Establishment Panel - Multiple Imputation for a Better Data Access. *ASTA Advances in Statistical Analysis* 92, 2008, 259-276.
- Drechsler, J. (2012). New Data Dissemination Approaches in Old Europe - Synthetic Datasets for a German Establishment Survey. *Journal of Applied Statistics* 39, 243-265.
- Drechsler, J., Ronning, G. und Bleninger, P. (2014). Disclosure risk from factor scores. *Journal of Official Statistics* 30 No. 1, erscheint im März 2014 .
- Gomatam, S., A.F. Karr, J.P. Reiter und A.P. Sanil (2005). Data Dissemination and Disclosure Limitation in a World Without Microdata: A Risk-Utility Framework for Remote Access Analysis Servers. *Statistical Science* 20, 163-177.
- Hochfellner, D., D. Müller, A. Schmucker und E. Roß (2012). Datenschutz am Forschungsdatenzentrum. FDZ Methodenreport 06/2012 DE. Version vom 21.12.2012, Bundesagentur für Arbeit.
- Kommission zur Verbesserung der informationellen Infrastruktur zwischen Wissenschaft und Statistik (Hrsg.): Wege zu einer besseren informationellen Infrastruktur Gutachten der vom Bundesministerium für Bildung und Forschung eingesetzten Kommission zur Verbesserung der informationellen Infrastruktur zwischen Wissenschaft und Statistik. Nomos Verlagsgesellschaft, Baden-Baden 2001
- Ronning, G. (2005) Randomized Response and the Binary Probit Model. *Economics Letters* 86 (2005), 221-228.
- Ronning, G. (2008a). Measuring Research Intensity From Anonymized Data: Does Multiplicative Noise With Factor Structure Save Results Regarding Quotients? *Jahrbücher für Nationalökonomie und Statistik/Journal of Economics and Statistics* 228, (2008), 645-653.
- Ronning, G. (2008b). Linear Models With Measurement Errors Arising From Mixture Distributions. *Acta et Commentationes Universitatis Tartuensis de Mathematica* 12,(2008), 55-64.
- Ronning, G. (2011). Disclosure Risk From Interactions and Saturated Models in Remote Access. IAW Diskussionspapier 72, Juni 2011.

- Ronning, G. et al (2005). Handbuch zur Anonymisierung wirtschaftsstatistischer Mikrodaten. Statistisches Bundesamt, Wiesbaden , Reihe "Statistik und Wissenschaft", Band 4, 2005, (gemeinsam mit Roland Sturm, Jörg Höhne, Rainer Lenz, Martin Rosemann, Michael Scheffler und Daniel Vorgrimler)
- Ronning, G, Rosemann, M, und Strotmann, H. (2005). Post-Randomization Under Test: Estimation of the Probit Model *Jahrbücher für Nationalökonomie und Statistik (Journal of Economics and Statistics)* 225 (2005), 544 - 566.
- Ronning, G., und Rosemann, M. (2008). SIMEX Estimation in Case of Correlated Measurement Errors. *AStA Advances in Statistical Analysis* 92 , 391-404.
- Ronning, G., und Schneeweiß, H. (2011). Panel Regression with Multiplicative Measurement Errors. *Economics Letters* 110 (2011), 136-139.
- Roque. G.M. (2000) , Masking Microdata Files with Mixtures of Multivariate Normal Distributions. Dissertation, June 2000, University of California, Riverside.
- Schneeweiß, H. und Ronning, G. (2010). Multiple Linear Panel Regression with Multiplicative Random Noise. in: Thomas Kneib and Gerhard Tutz (Eds.): *Statistical Modelling and Regression Structures*. Festschrift in Honour of Ludwig Fahrmeir. Physica-Verlag, 2010, 399-417.

IAW-Diskussionspapiere

Die IAW-Diskussionspapiere erscheinen seit September 2001. Die vollständige Liste der IAW-Diskussionspapiere von 2001 bis 2011 (Nr. 1-77) finden Sie auf der IAW-Internetseite www.iaw.edu/publikationene/iaw-diskussionspapiere.

IAW-Diskussionspapiere seit 2011:

- Nr. 78
The Role of Employees for Post-Entry Firm Growth (Februar 2012)
Andreas Koch / Jochen Späth / Harald Strotmann
- Nr. 79
Nationality Matters: The Geographic Origin of Multinationals and the Productivity of their Foreign Affiliates (Februar 2012)
Christian Arndt / Julia Spies
- Nr. 80
All You Need Is Trade: On the In(ter)dependence of Trade and Asset Holdings in Gravity Equations (März 2012)
Jörn Kleinert / Katja Neugebauer
- Nr. 81
The Tradeoff between Redistribution and Effort: Evidence from the Field and from the Lab (März 2012)
Claudia M. Buch / Christoph Engel
- Nr. 82
Factor Shares and Income Inequality – Empirical Evidence from Germany 2002-2008 (Mai 2012)
Martin Adler / Kai Daniel Schmid
- Nr. 83
Network and Selection in International Migration to Spain (Mai 2012)
Nina Neubecker, Marcel Smolka, Anne Steinbacher
- Nr. 84
Do Better Capitalized Banks Lend Less? Long-Run Panel Evidence from Germany (Mai 2012)
Claudia M. Buch / Esteban Prieto
- Nr. 85
From the Stability Pact to ESM – What next? (Juni 2012)
Claudia M. Buch
- Nr. 86
The Connection between Imported Intermediate Inputs and Exports: Evidence from Chinese Firms (Juni 2012)
Ling Feng / Zhiyuan Li / Deborah L. Swenson
- Nr. 87
EMU and the Renaissance of Sovereign Credit Risk Perception (August 2012)
Kai Daniel Schmid / Michael Schmidt
- Nr. 88
The Impact of Random Help on the Dynamics of Indirect Reciprocity (September 2012)
Charlotte Klempt
- Nr. 89
Specific Measures for Older Employees and Late Career Employment (Oktober 2012)
Bernhard Boockmann / Jan Fries / Christian Göbel
- Nr. 90
The Determinants of Service Imports: The Role of Cost Pressure and Financial Constraints (Oktober 2012)
Elena Biewen / Daniela Harsch / Julia Spies
- Nr. 91
Mindestlohnregelungen im Maler- und Lackiererhandwerk: eine Wirkungsanalyse (Oktober 2012)
Bernhard Boockmann / Michael Neumann / Pia Rattenhuber
- Nr. 92
Turning the Switch: An Evaluation of the Minimum Wage in the German Electrical Trade Using Repeated Natural Experiments (Dezember 2012)
Bernhard Boockmann / Raimund Krumm / Michael Neumann / Pia Rattenhuber

IAW-Diskussionspapiere

- Nr. 93 (Januar 2013)
Outsourcing Potentials and International Tradability of Jobs
Evidence from German Micro-Level Data
Tobias Brändle / Andreas Koch
- Nr. 94 (Februar 2013)
Firm Age and the Demand for Marginal Employment in Germany
Jochen Späth
- Nr. 95 (Juli 2013)
Messung von Ausmaß, Intensität und Konzentration des Einkommens- und Vermögensreichtums in Deutschland
Martin Rosemann / Anita Tiefensee
- Nr. 96 (Oktober 2013)
Flexible Collective Bargaining Agreements: Still a Moderating Effect on Works Council Behaviour?
Tobias Brändle
- Nr. 97 (Oktober 2013)
New Firms and New Forms of Work
Andreas Koch / Daniel Pastuh / Jochen Späth
- Nr. 98 (November 2013)
Non-standard Employment, Working Time Arrangements, Establishment Entry and Exit
Jochen Späth
- Nr. 99 (Dezember 2013)
Intraregionale Unterschiede in der Carsharing-Nachfrage – Eine GIS-basierte empirische Analyse
Andreas Braun / Volker Hochschild / Andreas Koch
- Nr. 100 (Dezember 2013)
Changing Forces of Gravity: How the Crisis Affected International Banking
Claudia M. Buch / Katja Neugebauer / Christoph Schröder
- Nr. 101 (Januar 2014)
Vertraulichkeit und Verfügbarkeit von Mikrodaten
Gerd Ronning