

Datenqualität am Beispiel des Deutschen Mikrozensus

Von Dr. Ralf Münnich, Abteilung Statistik, Ökonometrie und Unternehmensforschung am Wirtschaftswissenschaftlichen Seminar der Universität Tübingen und Dr. Rolf Wiegert, Universität Tübingen und IAW Tübingen.

Durch die zunehmende Internationalisierung auf vielen Gebieten, wie zum Beispiel auf denen der Politik, der Wirtschaft, der Migration und des supranationalen staatlichen Handelns im Rahmen der Europäischen Union, werden weitergehende Qualitätsanforderungen an die definitorischen Grundlagen und an das Datenmaterial aller amtlichen Statistiken der Mitgliedsländer gestellt. Diese Forderungen werden vorrangig im Interesse einer Harmonisierung im Europäischen Statistischen System (ESS) erhoben, aber natürlich ebenso unter Berücksichtigung zunehmender Technologisierung der Datengewinnung und -verarbeitung. Einen wesentlichen Einfluss darauf, einen zureichenden Begriff der Datenqualität zu entwickeln, hat die Diskussion in Politik und Administration über Qualitätsanforderungen an Statistiken bei gleichzeitig sinkenden Budgets und weitgehender Entlastung zu befragender Haushalte und Unternehmen.

Der vorliegende Beitrag befasst sich zunächst mit dem Begriff der supranationalen Datenqualität entsprechend den Grundvorstellungen eines Europäischen Qualitätsberichts. Anschließend wird als wichtige Konstituente der Datenqualität im Deutschen Mikrozensus sein Stichprobenfehler untersucht. Speziell dazu werden Ergebnisse einer Simulationsstudie herangezogen, die im Forschungsprojekt DACSEIS (Data Quality in Complex Surveys within the New European Information Society; <http://www.dacseis.de>) erzielt wurden. Das Projekt DACSEIS wird von der Europäischen Kommission gefördert.

Die vorliegende Arbeit bezieht sich auf einen von Herrn Münnich gehaltenen Vortrag anlässlich des 6. Statistischen Kolloquiums am 18. Juli 2002 im Statistischen Landesamt Baden-Württemberg in Stuttgart.

Die Autoren danken an dieser Stelle Herrn Dr. Werner Grünewald, Eurostat, für wertvolle Hinweise sowie den Teilnehmern des 6. Statistischen Kolloquiums für eine anregende Diskussion.

1. Zum Begriff der Datenqualität

1.1 Entwicklung der Diskussion

Mehr oder weniger ausführliche Diskussionen um den Begriff der Genauigkeit und Angemessenheit (Adäquation) sowie Zuverlässigkeit von erhobenen und anderweitig gewonnenen statistischen Daten, und damit inhaltlich um den Begriff der Datenqualität, haben seit der Einrichtung offizieller statistischer Ämter ab der Mitte des 19. Jahrhunderts immer wieder stattgefunden. Neben allgemeinen Grundeigenschaften der Qualität von gewonnenen Daten, die sich in der Dreierheit *Genauigkeit, Vollständigkeit, Aktualität*¹ ausdrücken, spielen jedoch auch eine Reihe sich wandelnder, eher aktuellerer Erfordernisse nach Anpassung der statistischen Konzepte und Prozeduren eine besondere Rolle. Diese liegen sowohl in der technischen Durchführung – man denke etwa an Internet-basierte Erhebungen und computerunterstützte Durchführungen von Interviews – als auch in der fort-



Die Autoren: Dr. Ralf Münnich und Dr. Rolf Wiegert

entwickelten Methodik; hierzu gehören moderne Editing- und Imputationsverfahren oder auch leistungsfähigere Messverfahren zu den Indikatoren von Datenqualität.

Im Rahmen der Jahrestagung der deutschen Statistischen Gesellschaft wurde 1984 die Datenqualität in den Mittelpunkt der Tagungsaktivitäten gestellt. 1992 wurde zum Auftakt der Wissenschaftlichen Kolloquiumsreihe im Statistischen Bundesamt in Wiesbaden die Eröffnungsveranstaltung

der statistischen Datenqualität gewidmet. Wichtige Themen waren damals und sind heute unter anderen:²

- Qualität statistischer Daten – Zur Diskussion um einen Kriterienkatalog;
- Qualitätsansprüche der politischen Entscheidungsträger;
- Qualitätsansprüche der Wirtschaft;
- Qualitätssicherung bei Erhebung und Aufbereitung;
- der Einfluss politischer Entscheidungen auf die Datenqualität am Beispiel des Mikrozensus.

¹ Diese drei Worte wurden wohl in Erinnerung an das kompakte Kriterium zur Abgrenzung von Grundgesamtheiten: „zeitlich, räumlich, sachlich“ auf der Tagung der Deutschen Statistischen Gesellschaft in Augsburg 1984 geprägt. Die Hauptversammlung dieser Tagung der deutschen Statistik hatte sich die Untersuchung des Qualitätsbegriffs statistischer Daten als Thema gewählt.

² Weitergehende Ausführungen können dem Tagungsband entnommen werden: Jürgen Chlumsky, Rolf Wiegert u. a. (Hrsg.), Qualität statistischer Daten, Band 25 der Schriftenreihe Forum der Bundesstatistik, Stuttgart: Metzler-Poeschel, 1993.

Die dissonante und lang anhaltende Diskussion in Deutschland um einen weiteren konventionellen Zensus nach dem letzten von 1987 drängte die Debatte um die Datenqualität ein wenig in den Hintergrund. Zusätzlich schuf die rapide gesteigerte Leistungsfähigkeit der Computer und ihr vehementes Potenzial für die Aufgaben der Statistik für die Beurteilung der Datenqualität ein verändertes Bewusstsein sowohl bei den Produzenten wie den Konsumenten der Statistik.

In den letzten beiden zurückliegenden Jahrzehnten nahm aber die Bedeutung des Begriffs der Datenqualität, insbesondere durch seine Internationalisierung, wieder sehr zu. Es zeigte sich eine dringende Notwendigkeit, die Versorgung der sich konsolidierenden EU mit zuverlässigen, weitgehend harmonisierten Daten sicher zu stellen. Eine *Leadership Expert Group (LEG) on Quality in Statistics* (Statistisches Amt der Europäischen Gemeinschaften, Eurostat) unter Leitung von Lars Lyberg, Statistik Schweden, wurde dafür eigens einberufen. Die Ergebnisse der Arbeiten wurden dann im Rahmen der *International Conference on Data Quality in Official Statistics* in Stockholm im May 2001 vorgestellt.³

Seit einiger Zeit gibt es weitere Aktivitäten zur Berichterstattung und Verbesserung von Datenqualität, speziell bei den großen intra-europäischen Surveys. Neben den Arbeitsgruppen *Working Group on Assessment of Quality in Statistics* und der *Task Force Variance Estimation* (beide Eurostat), beschäftigen sich auch einige Forschungsprojekte zumindest mit einigen Aspekten der Datenqualität. Hierzu gehören unter anderen die Projekte EUREDIT⁴ in Bezug auf Editing- und Imputationsmethoden, EURAREA⁵ in Bezug auf Small-Area-Methoden und die Quality-Gruppe von AMRADS.⁶ Eigens dem Thema Datenqualität widmet sich das von der Universität Tübingen koordinierte Forschungsprojekt DACSEIS.⁷ Selbstverständlich haben diese Projekte zum Ziel, die Methodik zu analysieren und voranzutreiben. Darüber hinaus werden jedoch auch geeignete Empfehlungen für die Praxis erarbeitet und so die notwendigen Grundlagen für eine Harmonisierung und Standardisierung eines internationalen Datenqualitätsbegriffes geschaffen.

Neben den europäischen Bestrebungen einer Harmonisierung und Standardisierung von Komponenten der Datenqualität, insbesondere von Eurostat, sind auch international erhebliche Aktivitäten zu beobachten. Dabei sind wesentliche Komponenten in den meisten Ansätzen vergleichbar. Eine besondere Beachtung finden die Berichte des US Bureau of the Census (<http://www.census.gov/srd/www/byyear.html>), von Statistics Canada (<http://www.statcan.ca>), des Internationalen Währungsfonds (http://dsbb.imf.org/dqrs_intro.htm) und von UN/ECE⁸, sowie die

zwei einschlägigen internationalen Konferenzen in Stockholm und Québec⁹, bei denen zahlreiche Arbeiten aus der ganzen Welt vorgetragen wurden.

1.2 Komponenten der Datenqualität nach Eurostat

Begriffe und Kriterien der Datenqualität werden von Eurostat zu einer internationalen Kompatibilität koordiniert und vorangetrieben. Neben den zentralen Komponenten der Datenqualität wird auch dem statistischen Berichtswesen erhöhte Aufmerksamkeit zugewendet; dem folgend werden die Nationalen Statistischen Ämter angehalten, ihre Berichtssysteme anzupassen und zu adjustieren. Im Sinne der Definition von Eurostat werden sieben isolierbare Komponenten einer modernisierten Datenqualität unterschieden:¹⁰

1. Relevanz statistischer Konzepte
2. Genauigkeit der Schätzergebnisse
3. Aktualität und Pünktlichkeit der Datenbereitstellung
4. Zugänglichkeit und Klarheit der Informationen
5. Vergleichbarkeit von Statistiken
6. Kohärenz
7. Vollständigkeit

Die Organisation der Statistik soll, diesen Punkten folgend, zwei inhärente Forderungen nicht unberücksichtigt lassen: den Aufwand für die Antwortenden sowie die Kosten der Statistikämter gering zu halten. Beide Ziele lassen sich aber kaum simultan optimieren und müssen insofern im Sinne eines Kompromisses behandelt werden. Insbesondere lassen rechtliche Rahmenbedingungen, vorrangig in Deutschland, zusätzliche methodische Erweiterungen aus Gründen des Datenschutzes vielfach nicht zu. Hierzu gehören etwa die Verwendung von Registerdaten, entweder im Sinne einer Datenfusion zur Vermeidung von Mehrfachbefragungen zum gleichen Gegenstand oder im Sinne einer geeigneten Verwendung von Zusatzinformationen zur Verbesserung der Schätzungen oder zur geeigneten Kompensation des Nonresponse.

Die Forschungsarbeiten des oben genannten Forschungsprojekts DACSEIS konzentrieren sich auf die Thematik der Genauigkeit von Schätzergebnissen. Sie beziehen sich auf Stichproben- und Nichtstichprobenfehler, wobei der Einfluss von Nonresponse und dementsprechend geeigneten Methoden zur Vermeidung bzw. Kompensation der Folgen von Nichtbeantwortung, also Gewichts- und Imputationsmethoden, auf die Schätzergebnisse untersucht werden soll.

³ Der Schlussbericht der LEG on Quality in Statistics ist unter <http://amrads.jrc.it/WPs%20pages/Quality/Documents/LEGsummary.pdf> verfügbar.

⁴ EUREDIT: The development and evaluation of new methods for editing and imputation ([http://www.cs.york.ac.uk/euredit/](http://www.cs.york.ac.uk/eureedit/)).

⁵ EURAREA: Enhancing Small Area Estimation Techniques to Meet European Needs (http://www.statistics.gov.uk/methods_quality/eurarea/).

⁶ AMRADS: Accompanying Measure to Research and Development in Official Statistics (<http://amrads.jrc.it>). Im Bereich der Arbeitsgruppe Quality sind zahlreiche weitere Informationen zur Datenqualität in der amtlichen Statistik erhältlich.

⁷ Eine Übersicht über das Projekt DACSEIS kann in Münnich, Ralf; Wiegert, Rolf (2001): The DACSEIS project. DACSEIS research paper series 1 (<http://w210.ub.uni-tuebingen.de/dbt/volltexte/2001/428>), nachgelesen werden.

⁸ Garonna, Paolo; Luige, Tiina (2001): Quality of the European Statistical System: the Statistical Dividend of the European Integration. Stockholm: International Conference on Quality in Official Statistics.

⁹ Die Schlusstagung zur LEG on Quality fand in Stockholm statt (<http://www.q2001.scb.se>). In Québec fand das Symposium von Statistics Canada zum Thema Achieving Data Quality in a Statistical Agency: a Methodological Perspective statt (<http://www.statcan.ca/english/conferences/symposium2001/>). Ebenso erwähnenswert ist das FCSM Statistical Policy Working Paper 31: Measuring and Reporting Sources of Error in Surveys (http://www.fcs.gov/01papers/SPWP31_final.pdf).

¹⁰ Vgl. beispielsweise Grünewald, Werner; Linden, Hakan (2001): Quality Measurement – Eurostat Experiences. In: Statistics Canada (Hrsg.), Proceedings of Statistics Canada Symposium 2001, und die dort zitierte Literatur.

Ziel der Untersuchungen ist auch die Festlegung von möglichen Standards für die Genauigkeit von Schätzungen im Sinne von Empfehlungen zur Anwendung bestmöglich geeigneter Methoden, welche dann selbst eine wesentliche Grundlage zur Vergleichbarkeit von Statistiken und auch für eine angestrebte Harmonisierung der Methodiken in der EU sowie in Europa liefern. Von besonderem Interesse im Rahmen dieses Forschungsprojekts sind Untersuchungen zur Genauigkeit von in der Praxis verwendeten komplexen Stichprobenerhebungen, also Erhebungen, die sowohl vom Umfang der Stichprobeneinheiten als auch von der Anzahl der Erhebungsvariablen eine besondere Größe aufweisen. In Deutschland steht insbesondere der Mikrozensus im Vordergrund.

2. Datenqualität im Deutschen Mikrozensus

Der Mikrozensus ist die größte Haushalts- bzw. Personenstichprobenerhebung in Deutschland, in der 1 % der Bevölkerung befragt wird.¹¹ Hauptgegenstand der Befragung sind Fragen zur Struktur der Bevölkerung und zum Arbeitsmarkt sowie der Erwerbsbeteiligung.

Neben dem Umfang einer solchen Erhebung sind auch einige rechtliche Rahmenbedingungen von Bedeutung, die sich insbesondere im Stichprobendesign und somit auch in der davon abhängigen Datenqualität auswirken.

2.1 Das Auswahlverfahren im Mikrozensus

Entscheidend bei der Beschreibung eines Auswahlverfahrens ist zunächst die Einteilung der Grundgesamtheit in Erhebungsgesamtheiten. Im Mikrozensus erfolgt primär eine regionale Einteilung nach Bundesländern und Regierungsbezirken bis hin zu insgesamt 214 Regionalschichten. Diese weisen eine Mindestbesetzung von etwa 200 000 bis 250 000 Einwohnern auf. Anschließend erfolgt eine Einteilung des Gebäudebestandes in Gebäudegrößenklassen. Es werden kleine, mittlere und große Häuser sowie Anstalten und Neubauten unterschieden. Zur Einteilung in Gebäudegrößenklassen werden Ergebnisse der letzten Volkszählung aus dem Jahre 1987 sowie der Bautätigkeitsstatistik herangezogen. Somit ergeben sich insgesamt 1 070 Schichten.

Zur Realisierung einer 1%-Stichprobe und unter Berücksichtigung der Praktikabilität werden diese 1 070 Schichten in Auswahlbezirke etwa gleicher Größe, zumeist 15 bis 20 Personen, eingeteilt.¹² Anschließend werden je 100 Auswahlbezirke zu einer Zone zusammengefasst aus der schließlich jeweils genau ein Auswahlbezirk zufällig gezogen wird.¹³

¹¹ Auf eine eingehende Darstellung des Mikrozensus wird hier verzichtet. Umfassende Darstellungen können beispielsweise in Mayer, K. (1994): Zum Auswahlplan des Mikrozensus ab 1990. In: Gabler, S.; Hoffmeyer-Zlotnik, J.; Krebs, D. (Hrsg.): Gewichtung in der Umfragepraxis. Opladen: Westdeutscher Verlag, S. 106 - 111, nachgelesen werden.

¹² Die Anzahl der Personen in den Auswahlbezirken weicht je nach Gebäudegrößenklasse zum Teil stärker ab. Dadurch können vereinzelt Cluster-Effekte bei Schätzungen auftreten, die eine Verringerung der Effizienz der Schätzung zur Folge haben.

¹³ In der Literatur wird oft von systematischer Auswahl gesprochen. Im Mikrozensus werden jedoch die 100 Auswahlbezirke innerhalb einer Zone zunächst permutiert bevor die Auswahl eines Auswahlbezirkes in den Zonen vorgenommen wird.

Man erhält damit ein klassisches mehrstufiges Auswahl-Design, das auf den ersten drei Stufen aus den Schichten Regionalschicht, Gebäudegrößenklasse und Zonen besteht und als eine vierte Schicht Klumpen von Auswahlbezirken verwendet. Die ersten drei Stufen können mathematisch als eine Stufe angesehen werden. Man erhält so schließlich eine klassische Klumpenstichprobe, wobei die interessierende Schätzvariable zumeist der Gesamtmerkmalsbetrag einer (Teil-)Population ist.¹⁴ Die Effizienz einer solchen Klumpenstichprobe hängt nun unmittelbar mit der Streuung der Gesamtmerkmalsbeträge in und zwischen den Zonen ab. Hierin spiegelt sich auch das Ergebnis der Mikrozensus-Reform von 1990 wider, in der die durchschnittliche Anzahl von Haushalten in den Klumpen zum Teil erheblich reduziert und somit eine wesentlich effizientere Schätzung ermöglicht wurde.¹⁵

2.2 Ausgewählte Ergebnisse

Die Untersuchung der Stichprobenfehler wird im Folgenden vorrangig betrachtet. Es interessiert also nicht nur das eigentliche Schätzergebnis, sondern auch die Genauigkeit dieser Angabe, die von den Eigenschaften der zugrunde liegenden Schätzfunktion abhängt. Als geeignetes Maß in diesem Zusammenhang wird bei unverzerrten Schätzfunktionen die Varianz bzw. der Standardfehler herangezogen. Zur Vermeidung einer Abhängigkeit vom Mittelwert-Niveau des Untersuchungsmerkmals wird gerne auch der relativierte Standardfehler verwendet, der jedoch nur bei verhältnisskalierten Merkmalen verwendbar ist. Die Überlegungen lassen sich auch bei nur approximativ unverzerrten Schätzern anwenden, weil in den meisten Fällen von einer erheblichen Dominanz der Varianz bzw. Standardabweichung der Schätzfunktion gegenüber einer möglichen Verzerrung ausgegangen werden kann.¹⁶

Eine andere Situation liegt vor, wenn Verfahren der Small-Area-Schätzung angewendet werden; sie führen in der Regel zu stärker verzerrten Ergebnissen.¹⁷ Hier sollte eine Zielfunktion eingesetzt werden, die sowohl die Streuung als auch die Verzerrung einer Schätzfunktion berücksichtigt. Dazu gehören beispielsweise der mittlere quadratische Fehler (Mean Square Error: MSE) die Wurzel aus dem MSE dividiert durch den Schätzwert – eine Art die Verzerrung berücksichtigender Variationskoeffizient – und die Überdeckungsrate von Konfidenzintervallen (Coverage Rate: CR).¹⁸ Allerdings ist dabei zu beachten, dass vielfach keine ge-

¹⁴ Zu Darstellung der Klumpen-Stichprobe vergleiche man etwa: Cochran, William G. (1977): Sampling techniques. New York: Wiley. Lohr, Sharon (1999): Sampling : design and analysis. Pacific Grove: Duxbury Press. Krug, Walter; Nourney, Martin; Schmidt, Jürgen (2001): Wirtschafts- und Sozialstatistik : Gewinnung von Daten. München ; Wien: Oldenbourg.

¹⁵ Zur Mikrozensus-Reform von 1990 vgl. Fußnote 11.

¹⁶ Es wird hier von einfachen Schätzfunktionen zur Ermittlung von Gesamtmerkmalsbeträgen bzw. von Verhältnis- und Regressionsschätzverfahren ausgegangen. Ein nennenswerter Einfluss der Verzerrung einer Schätzfunktion ist hier zumeist nur bei inadäquatem Einsatz, etwa bei sehr geringen Stichprobenumfängen zu beobachten. Vgl. auch Münnich, Ralf (1997): Gebundene Hochrechnung bei Stichprobenerhebungen mit Hilfe von Splines. Göttingen: Vandenhoeck und Ruprecht (Angewandte Statistik und Ökonometrie; 43).

¹⁷ Münnich, Ralf; Schmidt, Katrin (2002): Small Area Estimation in der Bevölkerungsstatistik. Baden-Württemberg in Wort und Zahl, Heft 3/2002, S. 139-145.

¹⁸ Bei der Überdeckungsrate von Konfidenzintervallen (coverage rate of confidence intervals) wird simulativ die relative Häufigkeit der zu ermittelnden Konfidenzintervalle über dem wahren zu schätzenden Wert ermittelt. Diese sollte dem tatsächlichen Konfidenzniveau entsprechen. Gegebenenfalls werden unterschiedliche Konfidenzniveaus betrachtet. Vgl. Skinner, C.; Holt, D.; Smith, T. (1997): Analysis of Complex Surveys. Chichester et al.: John Wiley & Sons. S. 24 ff.

eigneten Schätzungen für die Verzerrung (Bias) der Schätzfunktionen zur Verfügung stehen. Insofern muss bei realen Schätzungen, im Gegensatz zu Simulationen, von einer gewissen Unsicherheit ausgegangen werden, wenn man auch Verzerrung einbeziehende Fehlermaße zur Beurteilung der Güte von Schätzfunktionen heranziehen möchte. Nachfolgende Auswertungen beruhen jedoch auf Simulationsergebnissen, bei denen *wahre* Werte zur Beurteilung der Qualität von Schätzergebnissen verfügbar sind.

Für die Klumpenstichprobe ist das Hochrechnungsverfahren im Mikrozensus zur Ermittlung von Totalwerten ein gewichtetes Hochrechnungsverfahren, das unverzerrt ist.¹⁹ Dementsprechend unproblematisch ist hierfür die Angabe eines Varianzschätzers.²⁰ Allerdings geht als wesentliche Determinante in die Berechnung der Gesamtvarianz der Schätzfunktion die Streuung klumpenspezifischer Gesamtmerkmalsbeträge innerhalb der Zonen ein. Diese muss aufgrund fehlender Informationen über die Streuung der Gesamtmerkmalsbeträge innerhalb der Zonen geschätzt werden. Das ist jedoch problematisch, da aufgrund des Mikrozensus-Auswahlplans jeweils nur genau eine von 100 Auswahl-einheiten in jeder Zone gezogen wird, und aus den Angaben zu dieser einen Einheit keine zonen-spezifische Varianz geschätzt werden kann. Somit entsteht die Frage, wie eine geeignete Zusammenfassung von Zonen erfolgen kann, ohne erhebliche Qualitätseinbußen bei der Schätzung dieser Zonenvarianz zu erhalten. Möchte man allzu kompliziert zu ermittelnde Zusammenfassungen vermeiden, dann liegt es nahe, innerhalb jeder dieser 1 070 Schichten die Zonen zusammenzufassen. Alternativ kann man ebenso eine weitere Zusammenfassung innerhalb der Regionalschichten oder Gebäudegrößenklassen in Betracht ziehen.

In einer Simulation wurde hierzu die Anzahl der Männer in Baden-Württemberg geschätzt (4 807 515).²¹ Man erkennt aus *Tabelle 1*, wie nicht anders zu erwarten, dass die Schätzergebnisse aufgrund ihrer Unabhängigkeit zur geschätzten Klumpenstreuung alle identisch sind und im Rahmen der üblichen Simulations-

¹⁹ Man vergleiche hierzu auch die Ausführungen in Quatember, Andreas (2002): A comparison of the five Labour Force Surveys of the DACSEIS project from a sampling theory point of view. DACSEIS research paper series 3 (<http://w210.ub.uni-tuebingen.de/dbt/volltexte/2002/547>), und der dort angegebenen Literatur.

²⁰ Betrachtet man die gewichteten Totalwerte innerhalb der Klumpen als Untersuchungseinheiten, so kann man die Formeln der geschichteten Zufallsstichprobe anwenden. Vgl. beispielsweise Cochran (1977), S. 92.

²¹ Zur Beschreibung der Simulation zum Mikrozensus: Münnich, Ralf; Schürle, Josef (2003): On the Simulation of Complex Universes in the Case of Applying the German Microcensus. DACSEIS research paper series. In Vorbereitung.

Tabelle 1
Schätzung der Anzahl der Männer in Baden-Württemberg nach dem Auswahlverfahren des Mikrozensus*)

Verfahren zur Ermittlung der Zonenvarianz	Schätzwert	Varianz des Schätzwertes	Erwartungswert der Varianz-schätzung	Varianz der Varianz-schätzung
Aggregation über Grundgesamtheit	4 807 598	8,9162 E9	1,7678 E10	1,594 E15
Aggregation über Gebäudegrößenklassen	4 807 598	8,9162 E9	1,6964 E10	1,498 E15
Verfahren der amtlichen Statistik	4 807 598	8,9162 E9	8,9699 E9	8,631 E14
Wahre Zonenvarianz	4 807 598	8,9162 E9	8,8804 E9	0

*) Vergleich der zugehörigen Varianzschätzungen mit verschiedenen Schätzwerten für die Zonenvarianz und ihrem wahren Wert.

genauigkeit auch unverzerrte Ergebnisse ausweisen. Die tatsächliche Varianz des Schätzers muss demnach ebenso identisch sein. Zur Analyse der Stichprobenfehler müssen geeignete Schätzwerte für diese Varianz ermittelt werden. Als Referenzverfahren dient die Verwendung der wahren Zonenvarianz. Man bemerkt an diesem Punkt, dass das von der amtlichen Statistik verwendete Verfahren, die Zonenvarianz aus den Angaben der übergeordneten Schicht zu ermitteln, geeignete Schätzwerte liefert. Sobald man jedoch über die Gebäudegrößenklassen hinweg aggregiert, ergeben sich erhebliche Überschätzungen. Diese Überschätzungen wirken sich zusätzlich auf die Varianz der Varianzschätzung und damit auf die Qualität der Varianzschätzung aus.²² Betrachtet man jedoch das Verfahren der amtlichen Statistik in seiner Praxis, dann erkennt man unmittelbar, dass sich hier Fragen nach weiteren Varianzschätzverfahren wegen der hervorragenden Ergebnisse zu erübrigen scheinen. Diese Ergebnisse lassen sich tatsächlich auch verallgemeinern, solange die Gesamtzahl einer Population oder einer Teilpopulation geschätzt wird und keine weiteren eventuell vorhanden Informationen verwendet werden beziehungsweise auch dann, wenn nicht einzelne hohe Zonenvarianzen auftreten, wodurch eventuell erhebliche Verzerrungen entstehen können.

Eine geeignete Verbesserung der Schätzung lässt sich im Allgemeinen durch die Verwendung von Hilfsmerkmalen erreichen. Im Mikrozensus können bei der Schätzung der erwerbslosen Bevölkerung die Daten der Arbeitslosenstatistik der Bundesanstalt für Arbeit als Hilfsmerkmal herangezogen werden. Auch wenn eine solche Zusammenführung aus Datenschutzgründen derzeit auf Mikrodatenniveau nicht möglich ist, kann diese im Simulationsmodell problemlos nachgestellt werden. Hierzu eignet sich insbesondere die im Mikrozensus gestellte Zusatzfrage (EF214): *Sind Sie beim Arbeitsamt gemeldet?*

Mit Hilfe dieser Zusatzinformation können zunächst die verschiedenen gebundenen Hochrechnungsverfahren, der Differenzschätzer (Diff), der Verhältnisschätzer (Verh) und Regressions-schätzer (Regr), verwendet werden. Es handelt sich bei der Anwendung um kombinierte Schätzer, deren Merkmalsausprägungen aus der Gesamtzahl der Ausprägungen der interessierenden Variablen innerhalb der Klumpen ermittelt werden. Ebenso können separate Schätzer auf der Basis der fünf Gebäudegrößenklassen (GGK) beziehungsweise der 214 Regionalschichten (RS) verwendet werden. In der *Tabelle 2* sind Schätzungen für die Männer in Baden-Württemberg unter 26 (54 734) und deren Teilpopulation nicht deutsche Langzeitarbeitslose (4 781) aufgeführt.

Aus *Tabelle 2* folgt, dass eine geeignete Verwendung dieser Registerdaten mit Hilfe von gebundenen Hochrechnungsverfahren erhebliche Effizienzgewinne einbringen. Zum Vergleich dazu dient der einfache Totalwertschätzer (Total) gemäß der oben beschriebenen Vorgehensweise in der amtlichen Statistik. Geht man davon aus, dass die Anzahl der Erwerbslosen und die Anzahl der Arbeitslosen in

²² Auf die Bedeutung der Varianz der Varianz wird schon seit langem hingewiesen, vgl. etwa: Strecker, Heinrich (1987): The Efficiency of Sample Models with Equal and with Varying Probability of Selection. In: Beckmann, Martin J.; Wiegert, Rolf (Hrsg.), *Statistische Erhebungen, Methoden und Ergebnisse*, Ausgewählte Schriften von Heinrich Strecker. Göttingen: Vandenhoeck und Ruprecht. S. 70 - 82.

Tabelle 2

Schätzung der Anzahl der Männer unter 26 Jahren und der nicht deutschen langzeitarbeitslosen Männer unter 26 Jahren in Baden-Württemberg nach dem Auswahlverfahren des Mikrozensus^{*)}

Schätzverfahren	Schätzwert	Varianz des Schätzwertes	Erwartungswert der Varianzschätzung	Varianz der Varianzschätzung	Wurzel des MSE durch Schätzwert	Überdeckungshäufigkeit von 90 %- Konfidenzintervallen
					%	
Männer unter 26 Jahren						
Differenzschätzung	54 724	2,13E+06	2,20E+06	4,12E+10	2,67	90,65
Verhältnisschätzung	54 754	2,36E+06	2,44E+06	6,06E+10	2,81	90,52
Separate Verhältnisschätzung (nach Gebäudegrößenklassen)	54 840	2,44E+06	2,46E+06	6,31E+10	2,85	90,05
Separate Verhältnisschätzung (nach Regionalschichten)	55 657	3,21E+06	2,55E+06	8,23E+10	3,62	82,52
Regressionsschätzung	54 720	2,05E+06	2,10E+06	3,41E+10	2,61	90,57
Separate Regressionsschätzung (nach Gebäudegrößenklassen)	54 720	2,05E+06	2,08E+06	3,32E+10	2,62	90,37
Separate Regressionsschätzung (nach Regionalschichten)	54 742	2,05E+06	1,98E+06	2,95E+10	2,61	89,46
Einfache Schätzung des Totalwertes	54 693	6,71E+06	6,66E+06	1,42E+11	4,74	89,76
Nicht deutsche langzeitarbeitslose Männer unter 26 Jahren						
Regressionsschätzung	4 787	2,52E+05	2,43E+05	3,50E+09	10,49	88,82
Separate Regressionsschätzung (nach Gebäudegrößenklassen)	4 767	2,24E+05	2,08E+05	3,15E+09	9,93	87,70
Separate Regressionsschätzung (nach Regionalschichten)	4 012	2,26E+05	1,34E+05	2,43E+09	22,53	36,35
Einfache Schätzung des Totalwertes	4 783	6,38E+05	6,42E+05	1,31E+10	16,70	89,68

*) Vergleich verschiedener Schätzverfahren sowie die jeweils zugehörige Varianz des Schätzwertes.

den Klumpen jeweils nur gering voneinander abweichen, dann kann man daraus schließen, dass die Differenzschätzung und Verhältnisschätzung im Verhältnis zur Regressionsschätzung gute Ergebnisse liefern sollte. Tatsächlich kann man jedoch erkennen, dass die Regressionsschätzung bessere Ergebnisse liefert. Dies gilt ebenso bei geringeren Umfängen an zu schätzenden Personen. Allerdings erkennt man auch, dass die Verwendung der separaten Regressionsschätzung hier die für sie typischen Effizienzverluste aufweisen kann. Eine separate Hochrechnung nach Regionalschichten leidet unter zum Teil sehr geringen Besetzungen in den Klumpen der Stichprobe, die sich auf eine Verzerrung der Schätzung und Varianzschätzung niederschlägt. Dagegen bringt eine separate Hochrechnung bezüglich Gebäudegrößenklassen Vorteile, da die einzelnen Schichten im Allgemeinen noch stark besetzt sind, selbst bei selteneren Merkmalskombinationen, und die Heterogenität zwischen den Gebäudegrößenklassen in der Hochrechnung geeignet ausgenutzt wird.

Wie zuvor erörtert, werden für die Anwendung der gebundenen Hochrechnungsverfahren Angaben benötigt, die derzeit kaum verfügbar sind, nämlich die eindeutige Zuordnung von Register zu Erhebungseinheiten. Diese kann jedoch *künstlich* auf Basis der Angabe der Variablen EF214 modelliert werden, sofern man die Angabe des Interviewten als *korrekt* ansieht. Somit lassen sich auf Schichtniveau die für die Hochrechnung benötigten Steigungskoeffizienten ermitteln. Für die Hochrechnung selbst werden schließlich nur aggregierte Werte benötigt, die verfügbar sind.

Die Angabe der Variablen EF214 kann indes sicher nicht als vollständig valide erachtet werden. Dennoch kann ein möglicher Zusammenhang zwischen deren Angabe im Rahmen des Mikrozensus und der tatsächlichen Anzahl aktuell Arbeitsloser auf aggregiertem Niveau im Sinne einer Modellierung mit Störung untersucht werden. Diese wurde von Münnich und Wiegert²³ durchgeführt mit dem Ergebnis, dass unter *normalen* Bedingungen selbst bei diesen weniger validen Angaben noch erhebliche Effizienzgewinne möglich sind und die Varianzschätzung nicht von einer möglichen schlechten Angabe abhängt. Eine Verwendung dieser Zusatzinformationen ist unter gegebenen Bedingungen in jedem Falle zu empfehlen.

Obwohl in *Tabelle 2* nur angedeutet, ersieht man, dass eine Entscheidung für oder gegen ein Verfahren kaum von der Zielfunktion abhängt, da die Ergebnisse eine starke Dominanz der Varianzkomponente aufweisen. Dieses Resultat basiert jedoch auf nahezu unverzerrten Schätzverfahren. Die Verwendung von Small-Area-Schätzern, insbesondere von synthetischen Verfahren, verlangen hier eine eingehendere Analyse der Thematik auf die an dieser Stelle verzichtet wird. Es sei noch darauf hingewiesen, dass die Schätzverfahren bei sinnvoller Anwendung stets auch

²³ Vgl. Münnich, Ralf; Wiegert, Rolf (2002): An Approach to Improving Data Quality in Surveys by Applying Register Data with Limited Assignment of Units, Proceeding of the International Conference on Improving Surveys, Copenhagen (http://www.icis.dk/ICIS_papers/E_2_3.pdf).

eine Normalapproximation der Schätzverteilung zuließen. Ausnahmen bilden lediglich Extremfälle, in denen ein Schätzverfahren nicht mehr als geeignet erachtet wird, wie beispielsweise die separate Regressionsschätzung nach Regionalschichten in *Tabelle 2*, oder wenn die zu betrachtende Subpopulation in den meisten Fällen Klumpen mit Nullbesetzungen oder extrem geringen Besetzungen aufweist.

3. Zusammenfassung und Ausblick

Die Methodik zur Erfassung der Genauigkeit der Schätzergebnisse im deutschen Mikrozensus in der hier betrachteten Form erweist sich als hinreichend aussagefähig für die Schätzung von Gesamt- oder Teilgesamtmerkmalsbeträgen. Die Analyse weiterer Varianzschätzverfahren, etwa von Verfahren, die auf Resampling-Methoden beruhen, erweist sich als hinfällig, insbesondere bei Betrachtung des Aufwandes angesichts der vielen zu ermittelnden Schätzwerte im Mikrozensus.

Im Kontext erhöhter Qualitätsanforderungen bei gleichzeitig eher sinkenden Budgets muss bezüglich möglicher Verbesserungen der Schätzgenauigkeit und adäquater Messung einer definierten Gesamt-Datenqualität die Methodik geeignet angepasst und verbessert werden. Zur Anwendung einschlägiger Verfahren wird man auf Zusatzmerkmale, bevorzugt zuverlässige Registerdaten, kaum verzichten können – und zwar in Bezug auf eine mögliche grundlegende Verbesserung der Datenqualität und eine effizientere Ausnutzung von Informationen zur Eindämmung von zusätzlichen Kosten. Hier ist aber der Gesetzgeber gefordert, geeignete gesetzliche Rahmenbedingungen zu schaffen.

Die Ausführungen von Münnich und Wiegert in: *An Approach to Improving Data Quality*²⁴ weisen bereits auf mögliche Effizienzgewinne unter abgeschwächten Voraussetzungen hin. Bei vorsichtiger Anwendung von geeigneten Registerdaten ließen sich solche Effizienzgewinne noch steigern. Allerdings ist in der Praxis eine Einzelbetrachtung solcher Fragestellungen unumgänglich.

Wesentlich problematischer aus Sicht des Datenschutzes ist der sehr innovative Ansatz der Datenfusion und – eventuell sogar eine Stufe weiter zurückgehend – der Fusion von Fragen an potenzielle Respondenten, wodurch sich der Aufwand der Befragung zwar erhöht, insgesamt sich aber erhebliche Kostenreduktionen bei gleichzeitigen Verbesserungen der Datenqualität erreichen ließen.

Bei allen Vorschlägen sind jedoch im Sinne der Eurostat-Empfehlungen einer Berichterstattung der Datenqualität beziehungsweise von Qualität von Surveys stets Aufwand und Kosten der zu befragenden Haushalte und Unternehmen zu beachten und möglichst gering zu halten. Gerade deshalb hat es derzeit einen besonderen Charme, jedwede sonst wie verfügbare Information heranzuziehen, um Schätzergebnisse zu verbessern und so ihre Qualität zu steigern. Dabei sollte man sich jedoch einer geeigneten effizienten Methodik bedienen. Besonders gute Ergebnisse können die Länder erwarten, die über hoch-qualitative Registerdaten verfügen oder die auf Ergebnisse einer aktuellen Volkszählung zurückgreifen können.

Bleibt anzumerken, dass die eingangs erwähnten Projekte nicht alleine eine Grundlage für diese Überlegungen schaffen. Vielmehr sollen diese Bestrebungen in einem europäischen Kontext erweitert und vertieft werden, um im Rahmen des 6. Rahmenforschungsprogramms der Europäischen Kommission von einem europäischen Netzwerk gemeinsam genutzt zu werden (<http://www.publicstatistics.net>).

²⁴ Vgl. hierzu auch Fußnote 23.