

VERBESSERUNG VON AKTUELLEN SCHÄTZUNGEN DURCH VORANGEGANGENE STICHPROBEN – EIN METHODENVERGLEICH

Patricia Dörr

↳ **Schlüsselwörter:** wiederholte Stichproben – Kumulation über die Zeit – Rotationsstichproben

ZUSAMMENFASSUNG

In der amtlichen Statistik sind oft regelmäßig zu veröffentlichende Mittelwertschätzer, zum Beispiel von Haushaltseinkommen oder Konsumausgaben, von Bedeutung. Diese Schätzungen basieren zumeist auf zeitlich wiederholten Stichproben. Dabei kann es passieren, dass der Stichprobenumfang der aktuellen Stichprobe nicht ausreicht, um ein gewünschtes Präzisionsniveau der Punktschätzung zu erreichen. Vorgegangene Stichproben desselben Typs können für die aktuelle Schätzung Informationen enthalten, mit denen sich die Effizienz des Mittelwertschätzers verbessern lässt. Diese Arbeit illustriert und erweitert gängige Methoden im genannten Szenario. Vor- und Nachteile der Methoden werden diskutiert und abschließend anhand einer quasi-design-basierten Simulationsstudie bewertet, welche eine realitätsnah erzeugte synthetische Population nutzt.

↳ **Keywords:** repeated sample surveys – cumulation over time – sample rotation

ABSTRACT

Periodically published mean estimates, for instance of household income or consumption expenditure, often play an important role in official statistics. These estimates are mostly based on surveys that are conducted repeatedly. It may happen however that the size of the current sample is not sufficient to achieve the desired level of precision in the point estimate. Previously conducted sample surveys of the same type may contain information which can be used to improve the efficiency of the current mean estimate. This article illustrates and expands common methods under the scenario outlined above. The advantages and disadvantages of the methods are discussed and finally assessed by means of a quasi-design-based simulation study employing a close-to-reality synthetic population.



Patricia Dörr

ist Doktorandin am Lehrstuhl für Wirtschafts- und Sozialstatistik an der Universität Trier. Im Jahr 2016 hat sie ihr Masterstudium in "Survey Statistics" beendet, ein Jahr später ein Masterstudium in Economics. Ihr Forschungsschwerpunkt liegt in der Varianzschätzung, insbesondere von Business-Indikatoren. Ihre Masterarbeit zum Thema "Comparison of Methods for Combining Surveys over Time" wurde mit dem Gerhard-Fürst-Preis 2017 in der Kategorie Master-/Bachelorarbeiten ausgezeichnet. Die Masterarbeit wurde im Rahmen des Projekts „Forschung und Entwicklung von Innovationen für Stichprobendesigns der amtlichen Statistik (RIFOSS)“ vom Statistischen Bundesamt gefördert.

1

Einleitung

Für die amtliche Statistik sind Schätzungen von Mittelwerten in regelmäßigen Zeitabständen, vor allem aus Haushaltsbefragungen, von hoher Relevanz. Unter anderem bilden sie die zugrunde liegende Information für wirtschafts- und sozialpolitische Maßnahmen. Zu nennen sind hier beispielsweise die durchschnittlichen Ausgabenposten von Haushalten oder das Haushaltseinkommen aus den Laufenden Wirtschaftsrechnungen oder der Einkommens- und Verbrauchsstichprobe.

Wenn der Stichprobenumfang der zugrunde liegenden Befragung für ein gewünschtes Präzisionsniveau nicht ausreichend ist, stellt sich die Frage, ob und wie die aktuelle Stichprobe mit Informationen aus vorangegangenen Stichproben desselben Typs kombiniert werden kann, um die Genauigkeit der Punktschätzung zu erhöhen.

2

Überblick über Kumulationsverfahren

Die vorzustellenden Schätzer werden im Folgenden eingeteilt nach ihrer Zugehörigkeit zum design-basierten beziehungsweise modell-assistierten und dem modell-basierten Ansatz. Im jeweiligen Abschnitt werden weiterhin die Abkürzungen eingeführt, die in der Simulationsstudie verwendet werden. Design-basierte/modell-assistierte Schätzer betrachten die Daten unter Stichproben-Randomisierung, während modell-basierte Schätzer implizit oder explizit einen Daten generierenden Prozess (zumeist ein Regressionsmodell) annehmen. Diesen gilt es zu schätzen, um die Prozessstruktur in der Mittelwertschätzung zu nutzen. Ist der wahre Daten generierende Prozess allerdings anders als der angenommene, so sind die resultierenden Punktschätzungen nicht mehr erwartungstreu. Design-basierte Schätzer hingegen nehmen keinen Daten generierenden Prozess an, sodass kumulierende Schätzer Überlappungen in den Stichproben benötigen, um eine Struktur zwischen den zeitversetzten Stichproben herausfiltern zu können.

Im Bereich der design-basierten/modell-assistierten Schätzer sind vor allem der Beste Lineare Unverzerrte Schätzer (*blue*; Yansaneh, 1992, hier: Seite 106) und der Modifizierte Regressionschätzer (*mr*; Singh/Merkouris, 1995) zur Kombination zeitlich aufeinander folgender Stichproben zu nennen. Auch Hilfsmerkmale in den Stichproben, deren Mittelwerte bekannt sind, lassen sich mithilfe des *blue* schätzen. So ergibt sich auch die Möglichkeit, ein Pendant zum klassischen Generalisierten Regressionschätzer (*greg*) zu schätzen, welcher statt der Horvitz-Thompson-Schätzer (*simple* in der Simulationsstudie) die *blue* verwendet. In der Literatur wird dieser Schätzer als *b1* bezeichnet.

Der *mr* erweitert das assistierende Regressionsmodell um die interessierende Variable zu früheren Zeitpunkten. Folglich ergibt sich ein dynamisches Regressionsmodell, in welchem aufgrund der Endogenität der abhängigen Variablen die Schätzung der Regressionskoeffizienten problematisch ist. Inkonsistente Regressionsparameter wiederum wirken sich negativ auf die Qualität des assistierenden Modells aus und vergrößern somit die mittlere quadratische Abweichung (Mean Squared Errors – MSE) des *mr*. Deshalb wird der *mr* um ein Verfahren – Partial Adjustment – ergänzt (*mr.pa*), welches zum Ziel hat, das Endogenitätsproblem abzumildern. Da Partial Adjustment lediglich für den Einbezug einer zeitversetzten endogenen Variable gedacht ist, wird für die später zu erläuternde Simulationsstudie ebenfalls ein Algorithmus vorgeschlagen, um das Verfahren um weiter zurückliegende Beobachtungen zu erweitern.

Die implizite Annahme des *blue* und des *mr* ist allerdings, dass die Erwartungswerte der Rotationsgruppen identisch sind, das heißt keine strukturellen Unterschiede zwischen Neuzugängen und zum wiederholten Mal erhobenen Einheiten besteht. Ist dies nicht der Fall, weil zum Beispiel wiederholt Befragte aufgrund der Erhebung ihr Ausgabenverhalten verändern oder weniger gewissenhaft antworten, liegt ein Rotationsbias vor. Eine Annahmenverletzung würde somit die (asymptotische) Unverzerrtheit der design-basierten/modell-assistierten Schätzer infrage stellen. Für ausgewählte Schätzer wie den *blue* und den (klassischen) *mr* werden deshalb in Dörr (2016) mögliche Bias-Korrekturen vorgeschlagen, welche lediglich fordern, dass unverzerrte Schätzungen für die betreffende Periode durch die jeweiligen Neuzugänge möglich sind. Diese Annahme ist weniger restriktiv als das Ausbleiben eines Rotationsbias; auch

für einmalige Erhebungen muss angenommen werden, dass durch die erhobenen Einheiten eine unverzerrte Schätzung für Populationsgrößen möglich ist.

Zum Bereich der modell-basierten Schätzer zählt auch die naive Herangehensweise, alle Beobachtungswerte der zeitlich aufeinanderfolgenden Stichproben aufzusummieren. Auch wenn dies nicht explizit genannt ist, so beruht die einfache Kumulation (*cum*) auf der Annahme, dass der Populationsmittelwert über die Zeit konstant geblieben ist. Alternativ können vergangene Stichproben auch nach verschiedenen Verfahren heruntergewichtet werden (*cum.w1*, *cum.w2*), um die mangelnde Aktualität der Daten „zu bestrafen“. Dieses Heruntergewichten veralteter Stichproben lässt sich ebenfalls in ein (Zeitreihen-)Modell überführen, wie in Dörr (2016) gezeigt wird.

Des Weiteren bieten sich aufgrund der zeitlichen Dimension Zeitreihenmodelle an. Allerdings finden Haushaltsbefragungen oft in ein- oder mehrjährigen Abständen statt. Folglich hätten klassische Zeitreihenmodelle, welche lediglich den je Zeitpunkt geschätzten Mittelwert als Datenpunkt in der Zeitreihe betrachten, zu wenig Freiheitsgrade, um ein gutes Modell zu schätzen. Deshalb ist eine Berücksichtigung des im Vergleich zum Längsschnitt doch beachtlichen Umfangs der Querschnittdaten, das heißt der Anzahl aller erhobenen Einheiten, hilfreich, wenn Zeitreihenmodelle angewandt werden sollen. Diese Methode wird in Lind (2005) erläutert und beruht auf der Idee, jede Erhebungseinheit als eine zufällige Ausprägung eines gemeinsamen Modells zu betrachten.

Als letzte Möglichkeit zur Stichprobenkumulation seien Verfahren aus der Small-Area-Schätzung (Small Area Estimation – SAE) genannt. Hier wird üblicherweise ein zugrunde liegendes Regressionsmodell um räumliche Zufallseffekte erweitert. Anstatt der räumlichen Dimension bilden die Zufallseffekte bei der Stichprobenkumulation jedoch eine zeitliche Dimension ab. Weil im Gegensatz zur klassischen SAE eine zeitliche Dimension betrachtet wird, können diese Zufallseffekte möglicherweise auch in Zusammenhang zueinander stehen, beispielsweise über einen autoregressiven Prozess: Der Zufallseffekt zum Zeitpunkt T ist eine (lineare) Funktion der Effekte von $T-1$, $T-2$, ... und eines neuen Zufallsterms. Der klassische Regressionsschätzer wird dann um Prädiktionen dieser Zufallseffekte erweitert; in die

Schätzung des Regressionsmodells und die Prädiktion der Zufallseffekte fließen die vorangegangenen Stichproben ein. In der Simulationsstudie wird der Schätzer, welcher die Zufallseffekte als unabhängig voneinander annimmt, als *sae.ind* bezeichnet, bei einem unterstellten AR(1)-Prozess als *sae.ar1*.

Darüber hinaus wird in Dörr (2016) eine weitere Methode zur Stichprobenkumulation entwickelt: Die empirische Verteilungsfunktion zu verschiedenen Zeitpunkten wird als Folge von – fehlerbehafteten – Abbildern der Verteilungsfunktion in der Population betrachtet. In diesem Rahmen können, bei hinreichender Ähnlichkeit, vorangegangene „Bilder“ zur Stabilisierung des aktuellsten Bildes verwendet werden. Nicht überlappende Teile der Stichprobe stellen in dieser Terminologie zusätzliche Bildpunkte dar; mehrfach befragte Einheiten hingegen zeigen eine Veränderung der Ausprägung ein und desselben „Pixels“. Das heißt, das Perzentil, an dessen Stelle sich die Einheit in der empirischen Verteilung befindet, kann als Punkt im eindimensionalen Raum betrachtet werden. Um in der Terminologie zu bleiben, könnte der tatsächliche Beobachtungswert mit einem Farbwert identifiziert werden. Nach etwaiger Anpassung der „Bilder“ aufeinander (für die der überlappende Teil der Stichproben genutzt werden kann), können die zusätzlichen Bildpunkte im aktuellsten Bild genutzt werden. Mit anderen Worten: Die Information der aktuellsten Stichprobe wird um (eventuell angepasste) Beobachtungen erhöht, wodurch ein Effizienzgewinn möglich ist. Verschiedene Funktionen/Transformationen und Ähnlichkeitsmaße zur Anpassung der älteren Stichproben werden in Dörr (2016) diskutiert. Durch eine vorangestellte Simulationsstudie wurde aus den Kombinationsmöglichkeiten von Ähnlichkeitsmaß und Transformation eine ausgewählt, die Eingang in die hier vorgestellte Simulationsstudie findet und mit *image* bezeichnet wird. Ähnlich des *b1*, welcher die *blue*-Schätzer in das assistierende Modell einsetzt, können auch so geschätzte Hilfsmerkmale in ein assistierendes Regressionsmodell eingesetzt werden (*image.greg*).

Wie andere modell-basierte Ansätze kann diese Methode sowohl mit rotierenden als auch nicht rotierenden Stichproben umgehen. Zudem sind nicht lineare Anpassungsmethoden für die vergangenen empirischen Verteilungsfunktionen umsetzbar, was bei Zeitreihenmodellen und linearen Regressionsmodellen mit Zeiteffekten nur unter Mehraufwand der Fall ist.

Ein Vorteil dieser Methode ist, dass die geschätzten Quantile, welche für die Anpassung der empirischen Verteilungsfunktionen verwendet werden, robuster gegen Ausreißer sind als Mittelwertschätzer. Diese Vorteile, die nicht linearen Transformationen und die Ausreißerstabilität, sind gerade bei Einkommens- und Ausgabenvariablen mit oftmals schiefen Verteilungen ein Argument für diese neue Methode. Weiteres Forschungspotenzial liegt daher in der theoretischen Ausarbeitung der Methode und der Schätzung des Mean Squared Errors (MSE), auf welche in der Arbeit nicht eingegangen wird.

3

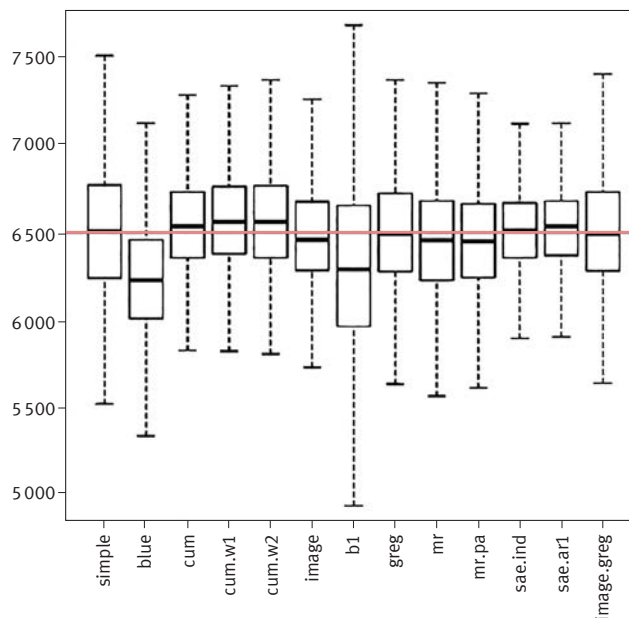
Simulationsstudie

Die Effizienz der vorgestellten Schätzer wird in der abschließenden Simulationsstudie untersucht. Dazu wurde eine synthetisch erzeugte, realitätsnahe Grundgesamtheit genutzt, welche für vier aufeinanderfolgende Zeitpunkte vorliegt und Variablen zu Konsum und Einkommen enthält. Somit sind hier die Populationswerte bekannt. In realen Anwendungen ist dies nicht der Fall. Interessierende Größe sind die mittleren Haushaltskonsumausgaben. Aus der synthetischen Population werden unter verschiedenen Rotationsdesigns wiederholt Stichproben gezogen. In der [Grafik 1](#) ist beispielhaft das Ergebnis für einen Überlappungsgrad von 75% dargestellt. In allen untersuchten Designs wurde eine Stichprobenrotation angewendet, um einen Vergleich zwischen design-basierten und modell-basierten Schätzern zu ermöglichen.

Die ersten sechs Schätzer in der Grafik nutzen keine weiteren Hilfsmerkmale für die Schätzung der mittleren Konsumausgaben. Aufgrund der impliziten Modellannahmen bei der Kumulation, welche in der synthetischen Grundgesamtheit nicht gelten, ist die einfache Stichprobenkumulation leicht verzerrt; allerdings ist ein Effizienzgewinn im Vergleich zur Nicht-Kumulation zu beobachten. Die neu vorgestellte, an die Bildbearbeitung angelehnte Methode schafft es, diese Verzerrung zu korrigieren, schlägt dabei jedoch in die andere Richtung aus. Nichtsdestotrotz bewegt sie sich auf demselben Niveau wie die einfache Stichprobenkumulation. Die überraschende Verzerrung des *blue* kann auf numerische Instabilitäten zurückgeführt werden, die aus der Invertierung einer geschätzten Kovarianzmatrix folgen.

Grafik 1

Boxplot zur Verteilung der Mittelwertschätzer zum aktuellsten Zeitpunkt bei einem 75%-Rotationsdesign



2018 - 01 - 0120


Die anderen Schätzer nutzen neben den älteren Beobachtungen der interessierenden Variable weitere Hilfsmerkmale, deren Populationsmittelwerte als bekannt vorausgesetzt werden. Hier wird deshalb als Vergleich nicht der HT-Schätzer (*simple*) herangezogen, sondern der Generalisierte Regressionsschätzer (*greg*).

Da der *blue* sich bereits als instabil erweist, folgt als Konsequenz auch die Verzerrung des *b1*. Während der klassische Modifizierte Regressionsschätzer ähnliche Ergebnisse zum *greg* liefert, ist eine leichte Verbesserung durch den partial-adjustment Algorithmus (*mr.pa*) zu beobachten. Das weist darauf hin, dass eine schwache Endogenitätsproblematik in der dynamischen Regression vorliegt. Das Einsetzen der mittels der neuen Methodik geschätzten Hilfsmerkmale in das assistierende Regressionsmodell liefert ähnliche Ergebnisse. Das wiederum deutet darauf hin, dass das Verfahren sich nicht nur eignet, die mittleren Konsumausgaben zu schätzen, sondern dass es auch die Mittelwerte der Hilfsmerkmale hinreichend gut schätzt.

Überzeugend sind auch die modell-basierten Schätzer auf Grundlage der SAE-Methodik, *sae.ind* und *sae.ar1*. Deren Performanz ist allerdings mit Vorsicht zu betrachten, da die verwendete synthetische Population

auf Regressionsmodellen beruht. Um hier jedoch einen Zirkelschluss zu vermeiden und die Robustheit der SAE-angelehnten Schätzer zu untermauern, sei auf Folgendes hingewiesen: Das gewählte Regressionsmodell entspricht nicht dem die Population erzeugenden Modell und die Methodik wurde an einem frei zugänglichen Datensatz angewandt.

Um die Robustheit der Ergebnisse zu überprüfen, wurde dieselbe Simulationsstudie auch mit einem realen, öffentlich zugänglichen Mikrodatensatz durchgeführt: dem italienischen Survey on Household Income and Wealth (SHIW) der Jahre 2006, 2008, 2010 und 2012 mit dem verfügbaren Haushaltsnettoeinkommen als interessierende Variable. Die Einheiten, die zu allen vier Zeitpunkten im SHIW vorhanden waren, wurden als Grundgesamtheit genutzt, aus der wiederum die Stichproben gezogen wurden. Auch der restliche Aufbau der zweiten Simulationsstudie entspricht dem hier beschriebenen Vorgehen und die Ergebnisse sind konsistent zu den vorgestellten Resultaten, was wiederum für die Vorteile der SAE-Schätzer spricht.

analytische und experimentelle Betrachtung lohnend scheint. In diesem Sinne sollte auch näher auf eine Schätzung des MSE eingegangen werden. 

4

Fazit und Ausblick

Die Arbeit beleuchtet sowohl theoretisch als auch im Rahmen einer Simulationsstudie Vor- und Nachteile von Schätzern, welche zeitlich vorangegangene Stichproben für aktuelle Mittelwertschätzungen nutzen. Bei geeigneter Wahl aus den vorgestellten Möglichkeiten ist es durchaus möglich, die Präzision der Punktschätzung zu erhöhen im Vergleich zu Schätzern, welche nur die aktuelle Stichprobe nutzen. Mit einbezogen werden sollte hier auch, ob eine etwaige Verzerrung zugunsten einer Varianzreduktion tolerierbar ist. Ein weiterer Aspekt ist ein mögliches Rotationsdesign der Stichprobe. Da manche Schätzer Rotation erfordern, ist auch hierdurch die Methodenauswahl beeinflusst. Eine Schätzung des MSE wird teilweise in Dörr (2016) beleuchtet; eine Simulationsstudie zur experimentellen Untersuchung der MSE-Schätzer steht jedoch noch aus.

Weiterhin haben sich die vorgeschlagenen Erweiterungen beziehungsweise Neuentwicklungen im Rahmen der Arbeit als durchaus wettbewerbsfähig zu den klassischen Schätzern erwiesen, sodass hier eine tiefere

LITERATURVERZEICHNIS

Demnati, Abdellatif/Rao, J. N. K. *Linearization variance estimators for survey data* Technical report. SSC Annual Meeting 2002. Proceedings of the Survey Method Section.

Dörr, Patricia. *Comparison of Methods for Combining Surveys over Time*. Unveröffentlichte Masterarbeit. 2016.

Lind, Jo Thori. *Repeated surveys and the Kalman filter*. In: The Econometrics Journal. Jahrgang 8. Ausgabe 3/2005, Seite 418 ff.

Rao, J. N. K. *Small area estimation*. Hoboken/New Jersey 2005.

Singh, A. C./Merkouris, P. *Composite estimation by modified regression for repeated surveys*. Proceedings of the Survey Research Methods Section. American Statistical Association. 1995. Seite 420 ff.

Yansaneh, Ibrahim Sorie. *Least Squares Estimation for Repeated Surveys*. Ph. D. thesis. Iowa State University 1992.

Herausgeber

Statistisches Bundesamt (Destatis), Wiesbaden

Schriftleitung

Dr. Sabine Bechtold

Redaktionsleitung: Juliane Gude

Redaktion: Ellen Römer

Ihr Kontakt zu uns

www.destatis.de/kontakt

Erscheinungsfolge

zweimonatlich, erschienen im April 2018

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: 1010200-18002-1

ISSN 0043-6143

ISBN 978-3-8246-1069-3

Download (PDF)

Artikelnummer: 1010200-18002-4, 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 (Destatis), 2018

Vervielfältigung und Verbreitung, auch auszugsweise, mit Quellenangabe gestattet.