

Der Einsatz von Small-Area-Verfahren – erste Erfahrungen mit Area-Level-Schätzungen

Unter dem Begriff Small-Area-Verfahren versteht man gemeinhin Methoden, die darauf abzielen, aus stichprobenbasierten Erhebungen belastbare statistische Ergebnisse in tiefer fachlicher oder räumlicher Gliederung abzuleiten. Eines dieser Verfahren, das Area-Level-Modell, wird im Rahmen einer Monte-Carlo-Simulation zur Schätzung der Einzelhandelsumsätze auf Ebene der Landkreise und kreisfreien Städte eingesetzt. Die Schätzergebnisse werden den Ergebnissen der klassischen Schätzmethode, der direkten Schätzung, gegenübergestellt. Die Simulation zeigt, dass Area-Level-Schätzungen sehr häufig, aber nicht immer, genauere Schätzergebnisse liefern als die direkte Schätzung. Da beim Area-Level-Ansatz auch der Varianz der direkten Schätzer eine bedeutsame Rolle zukommt, werden zudem unterschiedliche Ansätze zur Schätzung der Varianz des direkten Schätzers untersucht und miteinander verglichen. Von Dr. Peter Gottfried

In jüngerer Zeit häufen sich im Hessischen Statistischen Landesamt (HSL) die Nutzeranfragen nach regional und fachlich sehr tief gegliederten Daten. Die Gründe für diese Entwicklung sind vielfältig. Regional- und wirtschaftspolitische Maßnahmen werden vermehrt Evaluationen unterzogen, welche sehr detailliertes Datenmaterial erfordern. Auch wirtschaftliche Interessenverbände sind zunehmend bestrebt, ihre wirtschaftliche und gesellschaftliche Bedeutung für Land und Regionen in „harten Zahlen“ darzustellen. Dies sind nur zwei mögliche Motive für ein anhaltend hohes bzw. stetig steigendes Nutzerinteresse an tief gegliederten Daten. Diese Entwicklung stellt das HSL allerdings vor eine neue Herausforderung, da derart detaillierte statistische Angaben, sofern sie auf Stichproben-Erhebungen basieren, mit der herkömmlichen Methodik nicht in der gewohnten Zuverlässigkeit der Daten bereitgestellt werden können.

Small-Area-Verfahren bieten hier die Möglichkeit statistisch belastbarere Ergebnisse zu gewinnen. In der Wissenschaft sind die Verfahren mittlerweile fest etabliert und werden stetig weiterentwickelt. Sie fanden, international betrachtet, auch Einzug in die of-

fizielle Statistik¹⁾, werden aber bislang in der deutschen amtlichen Statistik eher zurückhaltend eingesetzt. Zu den wenigen Ausnahmen zählt die Arbeit von Dieterle (2011)²⁾, die zur Schätzung des Schweinebestandes auf Kreisebene neben dem klassischen Vorgehen auch einen synthetischen und einen zusammengesetzten Schätzer einsetzte und die Ergebnisse einander gegenüberstellte. In ähnlicher Weise befasst sich der vorliegende Beitrag mit Area-Level-Schätzungen, einem weiteren Small-Area-Verfahren³⁾. Als Testanwendung sollen die Umsätze des Einzelhandels auf Ebene der Landkreise und kreisfreien Städte geschätzt werden. Im Kern steht dabei ebenfalls die Frage, inwieweit sich mit Area-Level-Schätzungen belastbare kleinräumige Ergebnisse erzielen lassen. Zur Untersuchung der Frage wird auf Simulationstechnik zurückgegriffen. Da stichprobenbasierte Ergebnisse zufälligen Schwankungen unterliegen, wird das Vorgehen – Stichprobenziehen nebst Auswertung derselben – mehrfach wiederholt, um so ein umfassenderes Bild von der Leistungsfähigkeit des Verfahrens (im vorliegenden Anwendungsfall) zu erhalten. Der Beitrag ist somit eher methodischer Natur und richtet sich vornehmlich an Leser, die

1) Siehe beispielsweise Gonzales, M. E., Placek, P.J. und Scott, C. (1996): Synthetic Estimation of Followback Surveys at the National Center for Health Statistics. In Schaible, W. L. (Herausgeber): Indirect Estimators in U.S. Federal Programs.

2) Dieterle, M. (2011): „Die Schätzung regionaler Daten mithilfe von Small Area-Schätzmethoden“, Statistisches Bundesamt (Herausgeber): Wirtschaft und Statistik, Dezember 2011. Seite 1212-1218.

3) Eine kurze Übersicht zu Small-Area-Verfahren findet sich ebenfalls in Dieterle (2011).

vielleicht ihrerseits erwägen, detaillierte statistische Ergebnisse bereitzustellen und deshalb an Erfahrungsberichten mit einschlägigen Schätzverfahren interessiert sind, oder einfach Interesse dafür haben, was mit dem betrachteten Verfahren erreicht werden kann. Aufgrund dieser Intention sind im Textverlauf an einigen Stellen auch kurze Anmerkungen dazu eingefügt, wie das für die Berechnungen verwendete statistische Softwarepaket SAS eingesetzt wurde, um die jeweiligen Ergebnisse zu erzielen.



© allvision - Fotolia.com

Im nächsten Abschnitt wird das Testszenario etwas eingehender erläutert. Danach folgen die Ergebnisse der direkten Schätzung nebst einer eingehenderen Betrachtung der Varianzschätzung als Basis für die Beurteilung der Zuverlässigkeit der zuvor gewonnenen Schätzergebnisse. In Abschnitt 4 wird das Area-Level-Verfahren vorgestellt. Dem folgen die Präsentation der Schätzergebnisse sowie eine kurze Zusammenfassung und ein Fazit.

Das Testszenario

Ein naheliegender Ansatzpunkt für das Vorhaben, die Umsätze der hessischen Einzelhandelsunternehmen auf Kreisebene herunter zu brechen, wäre die Jahreserhebung im Handel. Aus dieser Stichprobe ließe sich eine direkte Schätzung der gesuchten Umsätze als gewichtete

Summe der Umsätze aller der WZ-Abteilung 47 zugeordneten hessischen Unternehmen vornehmen, mit den in der Stichprobe angegebenen Hochrechnungsfaktoren als Gewichte. Dieses nicht nur in der amtlichen Statistik sehr weit verbreitete Vorgehen ist, nebenbei bemerkt, mit der eingangs erwähnten herkömmlichen Methodik gemeint. Die formellere Bezeichnung des soeben beschriebenen Schätzverfahrens lautet „Horvitz-Thomson-Schätzer“ oder „direkter Schätzer“. Des Weiteren böte es sich an, zur Durchführung der Area-Level-Schätzung zusätzliche Informationen aus dem Statistischen Unternehmensregister (URS) heranzuziehen. Schließlich stellt letzteres die Auswahlgrundlage für die Jahreserhebung im Handel dar. Ein Vergleich der beiden Schätzergebnisse wäre allerdings von einer Reihe von Faktoren beeinträchtigt, die nicht zwangsläufig die Schätzgüte der beiden Verfahren reflektieren⁴⁾. Ferner wäre einer Wiederholung der Schätzungen mit einer anderen Zufallsauswahl von Einheiten nicht möglich und vor allem hätte man keine Anhaltspunkte dafür, wie stark die jeweiligen Ergebnisse vom tatsächlichen Wert abweichen.

Aus diesen Gründen wurde ein etwas anderer Weg beschritten: Als Ausgangspunkt diente die Registerkopie des URS 2014 für Hessen. Alle darin enthaltenen Unternehmen, die der WZ-Abteilung 47 zugeordnet sind und einen steuerbaren Umsatz größer 0 aufweisen, bilden die Grundgesamtheit oder Auswahlgrundlage für den Test. Sie umfasst 26 955 Einzelhandelsunternehmen. Für diese Grundgesamtheit lassen sich die nach den 26 hessischen Landkreisen und kreisfreien Städten zusammengefassten Umsätze unmittelbar errechnen. Diese Umsätze repräsentieren die tatsächlichen Werte, die es im Folgenden über eine Stichprobe aus der Grundgesamtheit zu schätzen gilt.

Für die Stichprobe wurde die Grundgesamtheit nach den Merkmalen *Branche* und *Umsatzgrößenklasse* geschichtet. Die Brancheneinteilung erfolgte nach WZ-Gruppen (Dreisteller), wobei die Gruppe 47.7 (Einzelhandel mit sonstigen Gütern) zusätzlich in Apotheken und Nicht-Apotheken untergliedert wurde, sodass insgesamt 10 Branchenschichten entstanden. Jede davon wurde in 5 Umsatzgrößenklassen unterteilt. Die

4) Z. B. die Frage bezüglich der Konsistenz beider Quellen: Unternehmensregister auf der einen und Jahreserhebung im Handel auf der anderen Seite.

Box 1: Stichprobenfehler, Schichtung, relativer Standardfehler und Mean Squared Error (MSE)

Statistische Ergebnisse, die auf der Auswertung von Stichproben basieren, sind im Grundsatz mit Fehlern behaftete Schätzungen des tatsächlichen Werts. Schließt man systematische Fehler aus, die über die Auswahl der Einheiten in die Stichprobe oder durch die Verwendung eines bestimmten Schätzverfahrens entstehen können, so sind die Schätzwerte immer noch mit zufallsbedingten Fehlern behaftet. Je nachdem, welche Einheiten zufällig in die Stichprobe gezogen werden, fällt der Schätzwert von Stichprobe zu Stichprobe unterschiedlich aus. Diese zufallsbedingten Fehler werden auch als **Stichprobenfehler** bezeichnet.

Bei einfachen Zufallsstichproben hängt die Höhe der zufälligen Schwankungen in den Schätzergebnissen im Wesentlichen von zwei Faktoren ab: der Heterogenität der Einzelwerte in der Grundgesamtheit und dem Stichprobenumfang. Je unterschiedlicher die Einzelwerte in der Grundgesamtheit sind, desto höher fällt tendenziell auch der Stichprobenfehler aus. Je höher dagegen der Stichprobenumfang gewählt wird, desto geringer schwanken in der Regel die Schätzergebnisse. Eine Erhöhung des Stichprobenumfangs ist aber kostspielig, da z. B. mehr Einheiten zu befragen sind, die Ergebnisse zu plausibilisieren sind etc. Daher sucht man günstigere Alternativen, um den Stichprobenfehler möglichst klein ausfallen zu lassen. Eine davon ist die Verwendung von **geschichteten Stichproben**. Dabei wird die Grundgesamtheit in homogenere Teilmengen unterteilt und aus jeder dieser Teilmengen eine Teilstichprobe gezogen. Nach welchen Kriterien die Teilmengen gebildet werden, hängt von den jeweiligen Untersuchungszielen ab. Die bundesweite Jahrerhebung im Handel ist beispielsweise nach Bundesländern, Branchen (Wirtschaftszweige (WZ)-Gruppen) und Umsatzgrößenklassen geschichtet. Stabilere Ergebnisse im Vergleich zur einfachen Zufallsstichprobe ließen sich bereits dadurch erzielen, dass aus allen Schichten derselbe Anteil von Einheiten in die Stichprobe gezogen wird, d. h. für alle Teilstichproben derselbe **Auswahlsatz** verwendet wird. Durch Variation der Auswahlsätze, Totalerfassung der Einheiten in der höchsten Umsatzgrößenklasse und kontinuierlich abnehmenden Auswahlsätzen für die darunterliegenden Umsatzschichten, ist ein noch stabileres Schätzergebnis (bei insgesamt gleichem Stichprobenumfang) erzielbar. Der unterschiedlichen Besetzungszahl der einzelnen Schichten wird bei der Auswertung der Stichprobe durch entsprechende **Gewichtung** der Einzelwerte Rechnung getragen. In der Regel entspricht das jeweilige Gewicht dem Kehrwert des Auswahlsatzes.

Als Maß für den Stichprobenfehler dient die **Fehlervarianz** der Schätzwerte $\hat{\theta}$. Sie beschreibt die erwartete quadratische Abweichung der Schätzwerte vom Erwartungswert der Schätzung $E(\hat{\theta})$:

$$V(\hat{\theta}) = E\left(\left(\hat{\theta} - E(\hat{\theta})\right)^2\right).$$

Da die Fehlervarianz sehr hohe Werte annehmen kann und aufgrund ihrer quadratischen Dimension sperrig zu interpretieren ist, wird an ihrer Stelle häufig der **Standardfehler** als Maßgröße herangezogen. Er errechnet sich einfach als Wurzel aus der Fehlervarianz und beschreibt damit die mittlere Abweichung der einzelnen Schätzwerte von ihrem Erwartungswert. Noch etwas griffiger und in der amtlichen Statistik am häufigsten verwendet ist der **relative Standardfehler**, bei dem der Standardfehler ins Verhältnis zum Schätzwert gesetzt ist:

$$RS(\hat{\theta}) = \frac{\sqrt{V(\hat{\theta})}}{\hat{\theta}}.$$

Er stellt ein dimensionsloses und einfach zu interpretierendes Maß dar. Ein relativer Standardfehler von beispielsweise 3 % gibt an, dass die einzelnen Schätzwerte im Durchschnitt um 3 % (des mittleren Schätzwerts) vom mittleren Schätzwert abweichen. Bei **erwartungstreuen** Schätzungen entspricht der Erwartungswert des Schätzers dem tatsächlichen Wert. Dann ließe sich – auf das Beispiel bezogen – sagen, dass die Einzelschätzungen durchschnittlich um 3 % vom tatsächlichen Wert abweichen.

Dem Qualitätsbericht der Jahrerhebung im Handel zufolge lag der relative Standardfehler für das Merkmal Umsatz im Handel insgesamt bei 0,4 %⁵⁾. Je tiefer die Ergebnisse gegliedert sind, desto höher fällt jedoch der relative Standardfehler trotz Schichtung aus. In der amtlichen Statistik ist es gängige Praxis, Ergebnisse in Klammern zu setzen, wenn der relative Standardfehler zwischen 10 % und 20 % liegt und auf den Ergebnisausweis zu verzichten, wenn der relative Standardfehler 20 % übersteigt⁶⁾.

Bei nicht erwartungstreuen Schätzfunktionen, wie etwa der hier betrachteten Area-Level-Schätzung, wird üblicherweise ein umfassenderes Maß – der **MSE** – verwendet. Er kennzeichnet die mittlere quadratische Abweichung der Schätzwerte vom tatsächlichen Wert:

$$MSE(\hat{\theta}) = E((\hat{\theta} - \theta)^2) = (B(\hat{\theta}))^2 + V(\hat{\theta}).$$

Er lässt sich in die Verzerrung $B(\hat{\theta})$, die Abweichung des mittleren Schätzwerts vom tatsächlichen Wert und die Fehlervarianz zerlegen. Das Maß umfasst somit neben der Streuung der einzelnen Schätzwerte ebenso die Treffsicherheit bzw. Genauigkeit des Schätzers. Bei erwartungstreuen Schätzern reduziert sich der MSE auf die Fehlervarianz. Analog zu den vorangehenden Maßen ist auch hier die Verwendung des **Root Mean Squared Error (RMSE)** – die Wurzel aus dem MSE – gebräuchlicher. Als relative Größe ergibt sich der **Relative Root Mean Squared Error (RRMSE)** nach Division des RMSE durch den tatsächlichen Wert:

$$RRMSE(\hat{\theta}) = \frac{\sqrt{MSE(\hat{\theta})}}{\theta}.$$

Einteilung erfolgte branchenspezifisch, wobei die Klassengrenzen aus der Jahrerhebung im Handel 2015 übernommen wurden. Gleiches gilt für die Stichproben-Auswahlsätze in den einzelnen Schichten⁷⁾. Insgesamt ist die Vorgehensweise damit sehr eng an das Vorgehen bei der Jahrerhebung angelehnt, ist aber kontrollierbarer und transparenter, da u. a. der Einfluss des Konzepts der Stichprobenrotation und die Problematik rund um echte und unechte Antwortausfälle ausgeblendet sind. Das Szenario entspricht gewissermaßen der Vorstellung, dass die im Unternehmensregister ausgewiesenen Größen alle zeitnah und valide sind und die Ergebnisse der Jahrerhebung ebenfalls unmittelbar ohne zeitliche Verzögerung vorliegen. Ferner kann die Stichprobenziehung und -auswertung beliebig oft wiederholt werden, wodurch ein umfassenderes Bild von der Qualität der Schätzverfahren entsteht.

Direkte Schätzung

Um die Güte der klassischen Vorgehensweise in diesem Szenario beurteilen zu können, wurde zunächst eine direkte Schätzung der Einzelhandelsumsätze auf Ebene der Landkreise und kreisfreien Städte vorgenommen. Tabelle 1

zeigt die Ergebnisse. Als Vergleichsgröße sind in der ersten Tabellenspalte die „tatsächlichen Werte“ aufgelistet. Sie ergeben sich durch Aufsummieren der Umsatzwerte nach Landkreisen und kreisfreien Städten über alle Einheiten der Grundgesamtheit hinweg. Die Spalten 2 und 3 stellen die Schätzergebnisse aus einer einzelnen Stichprobenziehung dar. Für die Ziehung der Stichprobe wurde die SAS-Prozedur SURVEYSELECT verwendet, für die Auswertung die Prozedur SURVEYMEANS. Die Schätzwerte errechnen sich als gewichtete Summe der Umsätze der Unternehmen (des/der jeweils betrachteten Landkreises bzw. kreisfreien Stadt) in der Stichprobe, wobei die aus dem Stichprobendesign gewonnenen Hochrechnungsfaktoren⁸⁾ als Gewichte dienen. Nähere Hinweise zur Rolle der Gewichte wie auch zur Berechnung der im weiteren Verlauf des Beitrags herangezogenen Maße zur Beurteilung der Güte der Schätzung finden sich in der Box 1 „Stichprobenfehler, Schichtung, relativer Standardfehler und Mean Squared Error (MSE)“.

Die in Spalte 3 aufgelisteten relativen Standardfehler werden ermittelt, indem man die geschätzte Standardabweichung zu den Umsatz-Schätzwerten ins Verhältnis setzt. Die pro-

5) Statistisches Bundesamt: Qualitätsbericht, Jahrerhebung im Handel sowie in der Instandhaltung und Reparatur von Kfz 2014, Wiesbaden 2016.

6) Siehe beispielsweise: Statistisches Bundesamt: Qualitätsbericht, Einkommens- und Verbrauchsstichprobe EVS 2013, Wiesbaden 2016, Seite 8.

7) Die hierzu notwendigen Informationen wurden dankenswerterweise vom Statistischen Bundesamt zur Verfügung gestellt.

8) Die Hochrechnungsfaktoren sind im vorliegenden Fall einfach der Kehrwert der jeweiligen Stichproben-Auswahlsätze.

Tabelle 1: Ergebnisse der direkten Schätzung der Einzelhandelsumsätze

Kreisfreie Stadt (St.) Landkreis	Tatsächlicher Umsatz	Einzelschätzung		10000 Schätzungen	
		Schätzwert	Geschätzter relativer Standard- fehler	Mittlerer Schätzwert	Relativer Standard- fehler
	1000 €	1000 €	%	1000 €	%
	1	2	3	4	5
Darmstadt, Wissenschaftsst.	605 790	529 384	18,1	605 112	8,6
Frankfurt am Main, St.	2 952 605	2 936 499	15,5	2 952 720	3,5
Offenbach am Main, St.	296 135	426 115	17,5	296 509	14,3
Wiesbaden, Landeshauptst.	936 084	1 043 615	14,9	937 009	7,2
Bergstraße	838 543	826 490	16,7	838 883	7,8
Darmstadt-Dieburg	608 199	633 462	12,7	608 434	10,5
Groß-Gerau	1 163 169	1 175 765	41,8	1 162 882	5,6
Hochtaunuskreis	642 328	702 669	14,6	642 083	10,6
Main-Kinzig-Kreis	1 974 214	2 090 032	25,4	1 973 569	4,3
Main-Taunus-Kreis	774 718	706 918	20,2	774 151	7,7
Odenwaldkreis	275 050	327 274	24,2	275 194	14,5
Offenbach	921 122	711 237	21,7	921 981	8,0
Rheingau-Taunus-Kreis	322 264	283 851	15,6	322 878	15,6
Wetteraukreis	1 769 164	1 803 192	33,0	1 769 441	4,3
Gießen	1 120 724	1 012 148	20,6	1 121 076	6,8
Lahn-Dill-Kreis	762 324	825 895	15,2	761 299	9,6
Limbürg-Weilburg	459 798	395 998	16,2	459 312	12,4
Marburg-Biedenkopf	729 466	770 354	13,5	729 461	9,8
Vogelsbergkreis	310 176	314 263	24,0	310 355	15,3
Kassel, documenta-St.	725 974	734 871	16,6	726 451	9,2
Fulda	2 085 578	2 122 960	47,0	2 085 811	3,4
Hersfeld-Rotenburg	411 452	497 515	23,2	410 576	14,0
Kassel	630 557	773 808	18,8	630 951	11,4
Schwalm-Eder-Kreis	1 689 450	1 600 101	45,3	1 688 645	3,9
Waldeck-Frankenberg	495 914	428 535	15,6	496 192	12,2
Werra-Meißner-Kreis	286 819	324 544	19,7	286 089	15,5

zentualen Angaben beziehen sich somit auf die geschätzten Umsatzwerte. Selbst wenn man berücksichtigt, dass Kreiszugehörigkeit kein Schichtungsmerkmal darstellt und die Standardfehler für ungeschichtete Auswertungsvariablen tendenziell größer ausfallen als für geschichtete, erscheinen die aufgelisteten Werte überraschend hoch. Legt man die in Box 1 angesprochene gängige Praxis bezüglich des Ausweises von unsicheren Werten auf die Tabellenwerte der Einzelschätzung an, so würde man 9 von 26 Schätzwerten durch einen Schrägstrich ersetzen, da der relative Standardfehler 20 % übersteigt und die übrigen Werte, deren Standardfehler zwischen 10 % und 20 % liegt, in Klammern setzen. Ferner überrascht, dass diejenigen Schätzwerte, deren relative Standardfehler mit über 40 % am

höchsten eingeschätzt werden, doch relativ nahe an den tatsächlichen Werten liegen, mitunter um weniger als 2 % von diesen abweichen.

Die beiden letzten Tabellenspalten zeigen die Ergebnisse der Monte-Carlo-Simulation. Simuliert wurden 10 000 Stichprobenziehungen mit anschließender direkter Schätzung der Umsätze. Der in Spalte 4 ausgewiesene mittlere Schätzwert ist das arithmetische Mittel aus allen Einzelschätzungen. Durch Vergleich der Mittelwerte mit den tatsächlichen Werten erkennt man unschwer die der direkten Schätzung inhärente Eigenschaft der Erwartungstreue. Die mittleren Schätzwerte weichen um weniger als 0,2 % vom Vorgabewert ab. Damit ist auch die aus den Einzelschätzungen errechnete Standardabweichung

chung eine relativ gute Approximation des tatsächlichen Standardfehlers. Nach Division der Standardfehler durch den jeweiligen mittleren Schätzwert ergeben sich wiederum die in Spalte 5 ausgewiesenen relativen Standardfehler. Offensichtlich fallen diese deutlich kleiner aus als die entsprechenden Schätzungen in Spalte 3. Nur in 15 von 26 Fällen liegt der relative Standardfehler tatsächlich über 10 % und in allen Fällen unter 20 %.

Dieses Ergebnis legt zwei Überlegungen nahe. Zum einen befindet man sich typischerweise in der Situation, die durch die beiden Spalten der Einzelschätzung reflektiert wird: man beurteilt die Güte der ermittelten Schätzwerte mit Blick auf die relativen Standardfehler. Dabei sollte man aber nicht aus den Augen verlieren, dass die Standardfehler ihrerseits nur Schätzwerte darstellen, die aus der einen verfügbaren Stichprobe abgeleitet sind. Zum anderen existieren in der Literatur eine Reihe von alternativen Schätzern für die Varianz des direkten Summenschätzers, die auf Basis unterschiedlicher vereinfachender Annahmen abgeleitet wurden. Da die Varianzschätzungen für den direkten Schätzer auch unmittelbar in die Small-Area-Schätzung eingehen und ihnen somit eine doppelte Bedeutung zukommt, erscheint es angesichts der Resultate von Tabelle 1 ratsam zu prüfen, ob im vorliegenden Fall – bei welchem Aggregate zu schätzen sind, die nicht zu der Schichtkonzeption passen – nicht geeignetere Schätzer existieren. Dies wird im nächsten Abschnitt untersucht.

Alternative Varianzschätzer

Getestet wurden 5 verschiedene Alternativen. Eine formale Darstellung der einzelnen Schätzugleichungen befindet sich im Anhang. Die ersten beiden Alternativen gehen unmittelbar auf die allgemeine Darstellung⁹⁾ der Varianz des direkten Summenschätzers unter Verwendung von Inklusionswahrscheinlichkeiten¹⁰⁾ zurück. Variante I berücksichtigt dabei, dass die gemeinsame Inklusionswahrscheinlichkeit davon abhängt, ob die beiden betrachteten Einheiten zur selben Schicht gehören oder nicht, Variante II unterstellt hingegen durchweg Unabhängigkeit der Inklusionswahrscheinlichkeiten.

Da eine Schätzung der Einzelhandelsumsätze nach Landkreisen und kreisfreien Städten beim Stichprobendesign durch entsprechende Schichtung nicht einbezogen wurde, wird die Zahl der Stichprobeneinheiten je Verwaltungsbezirk von Stichprobe zu Stichprobe unterschiedlich ausfallen. Schätzvariante III trägt dem Umstand Rechnung, dass der Stichprobenumfang je Landkreis bzw. kreisfreier Stadt eine Zufallsvariable darstellt. Bei den Varianten IV und V handelt es sich um die in der Prozedur SURVEYMEANS implementierten Varianzschätzer. Sie kommen standardmäßig zum Einsatz, wenn kein explizites Schätzverfahren über die Option VARMETHOD aufgerufen wird. Bei Variante III wurden der Prozedur die Besetzungszahlen der Schichten mitgegeben, bei Variante IV nicht. Varianzschätzungen auf der Basis von Bootstrap-Methoden, die ebenfalls in SURVEYMEANS implementiert sind, werden in Hinblick auf die längere Rechenzeit, die diese Verfahren benötigen, nicht getestet.

Als Beurteilungsmaßstab wird der RRMSE verwendet. Je kleiner der RRMSE ausfällt, desto effizienter und damit zuverlässiger ist die Schätzung. Dabei dient die durch Simulation ermittelte Varianz des Schätzers als Näherungslösung für den tatsächlichen Wert. Die RRMSE's der einzelnen Schätzvarianten sind in Tabelle 2 aufgeführt. Augenscheinlich entsprechen sich die Varianten I und IV sowie die Varianten III und V. Dies gilt nicht nur in Bezug auf den RRMSE, die Varianten führen trotz unterschiedlicher Ansätze auch zu identischen Schätzwerten.

Hinsichtlich der Schätzgüte produzieren die Varianten I und IV noch die verlässlichsten Schätzungen, wobei auch hier die Schätzwerte für die Varianz bei einigen Landkreisen, wie etwa dem Odenwaldkreis und dem Landkreis Hersfeld-Rotenburg, um durchschnittlich mehr als 40 % von den tatsächlichen Werten abweichen können. Recht ähnliche Ergebnisse werden mit Variante II erzielt, wobei die Annahme der Unabhängigkeit zwischen den Inklusionswahrscheinlichkeiten (insbesondere auch dann, wenn zwei Einheiten aus derselben Schicht vorliegen) zu tendenziell leicht höheren Schätzwerten führt. Bei Verwendung der explizit auf ungeplante Domains ausgerichteten Variante III wie auch bei der Varianz-

9) Die formale Darstellung findet sich beispielsweise in Münnich, R. (2008): Varianzschätzung in komplexen Erhebungen. Austrian Journal of Statistics, Vol. 37, Seite 321.

10) Die Inklusionswahrscheinlichkeit kennzeichnet die Wahrscheinlichkeit dafür, dass die entsprechenden Einheiten in die Stichprobe gezogen werden.

schätzung über SURVEYMEANS ohne Angabe der Schichtenbesetzung tritt hingegen relativ häufig der Fall auf, dass die Varianz sehr stark überschätzt wird. Die in Tabelle 1 ausgewiesenen Standardfehler wurden über die Variante V bestimmt und sind hierfür ein Beispiel. Wird SAS SURVEYMEANS zur direkten Schätzung von Summen oder Mittelwerten verwendet, dann sollten, soweit möglich, der Prozedur auch Angaben zur Schichtgröße in der Grundgesamtheit und die Schichtungsmerkmale übergeben werden.

Das Area-Level-Modell

Area-Level-Schätzer stellen eine Möglichkeit dar, gegenüber einer direkten Schätzung insbesondere dann verlässlichere Schätzwerte zu liefern, wenn die Areas in der Stichprobe nur sehr schwach (oder gar nicht) vertreten sind. Die Stabilisierung wird durch die Einbeziehung von Hilfsinformationen bzw. Hilfsvariablen erzielt. Das grundlegende Area-Level-Modell unterstellt dabei einen linearen Zusammenhang zwischen den tatsächlichen Umsatzmittelwerten \bar{u}_d und den Mittelwerten der p Hilfsvariablen $Z_d = (z_{d1}, \dots, z_{dp})^T$,

$$\bar{u}_d = Z_d^T \beta + v_d,$$

wobei der lineare Zusammenhang durch einen kreissspezifischen Random Effekt v_d überlagert wird¹¹⁾. Die v_d sind hierbei identisch und unabhängig verteilt mit Erwartungswert 0 und Varianz σ_v^2 .

Weiterhin würde eine direkte Schätzung die Umsatzmittelwerte erwartungstreu schätzen

$$\hat{u}_d = \bar{u}_d + e_d,$$

wobei die Stichprobenfehler e_d unabhängig mit Erwartungswert 0 und Varianz D_d verteilt sind. Beides zusammengenommen ergibt das Modell

$$\hat{u}_d = Z_d^T \beta + v_d + e_d.$$

Aus diesem Zusammenhang lassen sich die Random Effekte abschätzen und damit auch die Area-Level-Schätzer bestimmen. Letztere lassen sich mit

$$\hat{u}_d^{FH} = \hat{\gamma}_d \hat{u}_d + (1 - \hat{\gamma}_d) Z_d^T \hat{\beta} \quad \text{und} \quad \hat{\gamma}_d = (\hat{\sigma}_v^2 + \hat{D}_d)^{-1} \hat{\sigma}_v^2$$

als Linearkombination der direkten Schätzer und dem Regressionsteil darstellen. Dabei werden die \hat{u}_d und \hat{D}_d üblicherweise von der direkten Schätzung übernommen, die übrigen Größen im

Tabelle 2: Effizienz der getesteten Varianzschätzer – bewertet über Relative Root Mean Squared Errors (RRMSE)

Kreisfreie Stadt (St.) Landkreis	Schätzvarianten				
	I	II	III	IV	V
Darmstadt, Wissenschaftsst.	40,3	41,8	312,1	40,3	312,1
Frankfurt am Main, St.	19,2	24,9	1 924,3	19,2	1 924,3
Offenbach am Main, St.	45,3	46,4	113,9	45,3	113,9
Wiesbaden, Landeshauptst.	28,6	31,2	360,7	28,6	360,7
Bergstraße	28,0	30,2	338,8	28,0	338,8
Darmstadt-Dieburg	30,9	33,1	81,7	30,9	81,7
Groß-Gerau	31,7	33,8	5 504,0	31,7	5 504,0
Hochtaunuskreis	30,9	32,8	85,1	30,9	85,1
Main-Kinzig-Kreis	22,0	25,0	3 753,9	22,0	3 753,9
Main-Taunus-Kreis	31,3	33,0	502,0	31,3	502,0
Odenwaldkreis	48,2	49,1	275,5	48,2	275,5
Offenbach	26,1	29,0	409,5	26,1	409,5
Rheingau-Taunus-Kreis	36,4	38,2	50,8	36,4	50,8
Wetteraukreis	28,4	30,7	6 587,4	28,4	6 587,4
Gießen	26,1	27,9	702,2	26,1	702,2
Lahn-Dill-Kreis	24,7	27,0	179,8	24,7	179,8
Limburg-Weilburg	35,0	36,4	67,8	35,0	67,8
Marburg-Biedenkopf	29,0	31,4	108,3	29,0	108,3
Vogelsbergkreis	42,7	43,8	158,9	42,7	158,9
Kassel, documenta-St.	32,1	34,0	182,7	32,1	182,7
Fulda	29,5	31,3	20 400,9	29,5	20 400,9
Hersfeld-Rotenburg	44,4	45,7	210,4	44,4	210,4
Kassel	29,0	30,6	229,9	29,0	229,9
Schwalm-Eder-Kreis	34,6	36,1	13 819,2	34,6	13 819,2
Waldeck-Frankenberg	32,1	33,6	72,3	32,1	72,3
Werra-Meißner-Kreis	43,0	44,1	103,4	43,0	103,4

I Näherungslösung für Standardformel mit Korrektur bei gleicher Schicht
 II Näherungslösung für Standardformel ohne Korrektur
 III Näherungsformel für zufällige Domains
 IV SAS SURVEYMEANS mit Angabe der Besetzungszahlen der Schichten
 V SAS SURVEYMEANS ohne Angabe der Besetzungszahlen

Modell geschätzt. Dieses Modell geht auf Fay und Herriot (1979)¹²⁾ zurück, weshalb die Schätzer üblicherweise als FH-Schätzer \hat{u}_d^{FH} bezeichnet werden.

Anmerkungen zum Vorgehen bei der Schätzung

Für die nachfolgenden Schätzungen wurde nur eine Hilfsvariable verwendet. Wie bereits die kurze Modellbeschreibung nahe legt, schneidet der Fay-Herriot-Schätzer wohl umso besser ab, je enger der Zusammenhang zwischen der Hilfsvariablen und der Schätzvariablen ist, d. h. je höher sie miteinander korreliert sind. Wollte man die

11) Eine detailliertere Darstellung des grundlegenden Area-Level-Modells findet sich beispielsweise in Rao, J. N. K. und Molina, I. (2015): Small Area Estimation, 2nd Edition, Seite 76ff.

12) Fay, R. E. und Herriot, R. A. (1979): Estimates of Income for Small Places: An Application of James-Stein Procedures to Census Data. Journal of the American Statistical Association, Bd. 74, S. 269-277.



© rosifan19 - Fotolia.com

Einzelhandelsumsätze in den Verwaltungsbezirken mit diesem Schätzverfahren aus der Jahrerhebung schätzen, so böten sich die in der Registerkopie des URS hinterlegten steuerbaren Umsätze, entsprechend zusammengefasst, als Hilfsvariable an. Bereits auf der Ebene der Unternehmen liegt eine hohe Korrelation zwischen den jeweiligen Umsatzwerten vor. In dem hier zu Grunde gelegten Testszenario entfällt diese Möglichkeit. Deshalb wird aus den im URS ausgewiesenen steuerbaren Umsätzen eine künstliche Hilfsgröße erzeugt¹³⁾. Dieses Vorgehen hat den Vorteil, dass über die Konstruktion der Hilfsvariablen auch die Korrelation zur Schätzgröße „gesteuert“ werden kann. Der Korrelationskoeffizient zwischen den Kreismittelwerten der konstruierten Hilfsvariablen und der Zielvariablen beträgt rund 0,96. Ein ähnlich hoher Wert ist für die einschlägigen Größen aus dem URS und der Jahrerhebung zu verzeichnen.

Wie bereits zuvor wurden die Schätzungen unter Verwendung von SAS durchgeführt. Allerdings stellt SAS keine spezielle Prozedur für Area-Level-Modelle bereit. Für die Varianzschätzungen wurde die Prozedur MIXED eingesetzt, die Berechnung der Schätzwerte erfolgte im Anschluss daran in der IML (Interactive Matrix Language)-Umgebung von SAS gewissermaßen manuell. Auf eine explizite Beschreibung des Vorgehens wird verzichtet und auf den Beitrag von Mukhopadhyay und McDowell (2011)¹⁴⁾ im SAS Global Forum verwiesen. Er beschreibt sehr

detailliert anhand von Beispielen die Umsetzung einer Reihe von Small-Area-Modellen in SAS Programmcode, unter anderem auch das hier verwendete Fay-Herriot-Modell.

Schätzergebnisse

Die eingangs aufgeworfene und eher allgemein gehaltene Frage hinsichtlich der Belastbarkeit von Area-Level-Schätzungen gegenüber dem klassischen Verfahren wird hier durch zwei mehr pragmatisch orientierte Teilfragen konkretisiert: Wie gut werden die vorgegebenen tatsächlichen Werte geschätzt? Wie stark variieren die Schätzwerte von Stichprobe zu Stichprobe?

In Tabelle 3 sind die Ergebnisse zweier Einzelschätzungen – einer direkten Mittelwertschätzung und einer Area-Level-Schätzung – einander gegenübergestellt. Die geschätzten Umsatz-Mittelwerte wurden anschließend über die Anzahl der Unternehmen in den einzelnen Landkreisen und kreisfreien Städten in Gesamtbeträge umgerechnet. Vergleicht man die jeweiligen Schätzwerte mit dem tatsächlichen Umsatz, so zeigt sich, dass die FH-Schätzwerte bei 19 von 26 Landkreisen und kreisfreien Städten oder in rund 73 % aller Fälle näher am Vorgabewert liegen als die direkten Schätzwerte. Ausgehend von den in diesen Fällen größeren Abweichungen der direkten Schätzwerte reduziert die FH-Schätzung diese Abweichungen um durchschnittlich 50 %. Für 7 Landkreise bzw. kreisfreie Städte (23 % der Fälle) fällt die FH-Schätzung dagegen schlechter aus. Die Verschlechterung macht dabei 70 % der jetzt größeren Abweichung aus. Würde man Verbesserungen und Verschlechterungen der Schätzergebnisse miteinander verrechnen, um zu einer (zugegebenermaßen) kruden Gesamtschätzung zu gelangen, so fielen die Abweichungen der FH-Schätzwerte von den entsprechenden Vorgabewerten in dieser Stichprobe um durchschnittlich 20 % geringer aus als die Abweichungen der direkten Schätzwerte von den Vorgabewerten. Da diese durchschnittliche Verbesserung rein rechnerischer Natur ist und sich nicht gleichmäßig über die Schätzwerte erstreckt, sei die Einschätzung der Sinnhaftigkeit dieser Zusammenfassung der eigenen Beurteilung überlassen.

13) Rund 15 % der Umsatzwerte werden zufällig ausgewählt und durch die Werte einer gleichverteilten Zufallsvariablen ersetzt. Die Variable nimmt dabei Werte zwischen 20 und 600 an. Die nicht ersetzten Umsatzwerte werden um den Faktor 10 gekürzt und mit einer normalverteilten Zufallsvariablen überlagert. Die normalverteilte Zufallsvariable hat den Erwartungswert 0 und die Standardabweichung 100.

14) Mukhopadhyay, P. K. und McDowell, A. (2011): Small Area Estimation for Survey Data Analysis Using SAS Software, SAS Global Forum 2011, Paper 336-2011.

Tabelle 3: Ergebnisse einer direkten Schätzung und einer Area-Level-Schätzung im Vergleich

Kreisfreie Stadt (St.) Landkreis	Tatsächlicher Umsatz	Direkte Schätzung		Area-Level-Schätzung	
		Direkter Schätzwert	Geschätzter relativer Standard- fehler	FH-Schätzwert	Geschätzter RRMSE
	1000 €	1000 €	%	1000 €	%
	1	2	3	4	5
Darmstadt, Wissenschaftsst.	605 790	558 165	15,6	595 755	10,8
Frankfurt am Main, St.	2 952 605	2 924 090	7,9	2 909 474	7,1
Offenbach am Main, St.	296 135	498 477	23,1	373 230	21,2
Wiesbaden, Landeshauptst.	936 084	891 238	13,1	942 293	10,9
Bergstraße	838 543	801 097	13,2	820 773	10,9
Darmstadt-Dieburg	608 199	696 312	16,7	681 739	16,0
Groß-Gerau	1 163 169	1 060 720	15,3	1 164 562	9,5
Hochtaunuskreis	642 328	627 707	13,2	613 814	11,6
Main-Kinzig-Kreis	1 974 214	2 217 350	11,7	2 123 433	9,7
Main-Taunus-Kreis	774 718	774 199	15,3	757 028	11,3
Odenwaldkreis	275 050	166 569	19,8	215 920	10,8
Offenbach	921 122	695 411	10,1	725 102	7,4
Rheingau-Taunus-Kreis	322 264	422 758	20,6	391 253	20,8
Wetteraukreis	1 769 164	1 809 175	13,6	1 550 378	8,6
Gießen	1 120 724	1 060 180	13,8	1 097 227	9,9
Lahn-Dill-Kreis	762 324	793 258	16,6	780 463	12,6
Limburg-Weilburg	459 798	628 208	19,8	509 598	17,8
Marburg-Biedenkopf	729 466	728 825	13,2	711 992	11,0
Vogelsbergkreis	310 176	274 502	21,1	280 152	13,7
Kassel, documenta-St.	725 974	644 924	13,6	664 220	10,1
Fulda	2 085 578	2 171 946	15,5	2 151 320	10,3
Hersfeld-Rotenburg	411 452	418 768	18,8	377 803	12,0
Kassel	630 557	841 135	18,9	737 402	15,0
Schwalm-Eder-Kreis	1 689 450	1 726 630	18,6	1 322 889	7,9
Waldeck-Frankenberg	495 914	577 334	16,0	506 122	13,8
Werra-Meißner-Kreis	286 819	380 000	21,9	327 590	17,1

Für eine erste Einschätzung der Stabilität der Schätzwerte sind in den Spalten 3 und 5 von Tabelle 3 die relativen Standardfehler der direkten Schätzwerte und die RRMSE der FH-Schätzwerte aufgelistet. Beim Vergleich der beiden Größen ist allerdings zu beachten, dass der RRMSE nicht nur die Streuung der FH-Schätzwerte sondern auch deren Verzerrung umfasst¹⁵⁾. Da die direkte Schätzung erwartungstreu ist, kann der relative Standardfehler auch als RRMSE interpretiert werden. Insofern sind die Werte der Spalten 3 und 5 durchaus vergleichbar, reflektieren aber eher die Effizienz der Schätzer insgesamt. Bedenkt man, dass die Verzerrung quadratisch in den MSE eingeht, kann die relative Streuung der FH-Schätzwerte nicht größer ausfallen als der

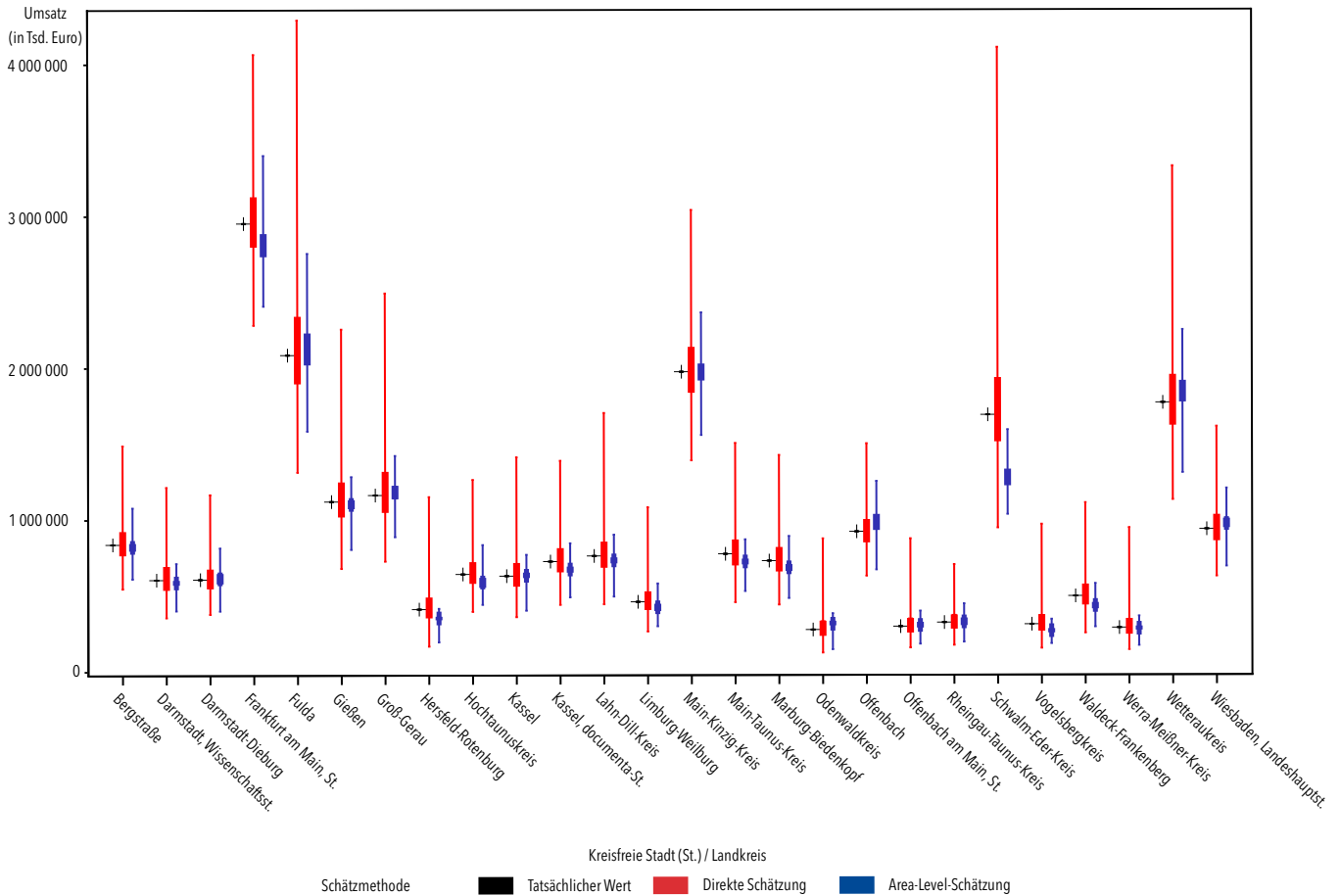
RRMSE. Da die RRMSE-Schätzwerte bereits kleiner sind als die entsprechenden Schätzwerte für den relativen Standardfehler der direkten Schätzung, lässt sich unmittelbar folgern, dass auch die Streuung des FH-Schätzers deutlich kleiner ist als diejenige des direkten Schätzers.

Einen umfassenderen Eindruck bezüglich des Abschneidens der beiden Schätzverfahren vermitteln die in Abbildung 1 dargestellten Ergebnisse der Simulationsrechnung. Simuliert wurden wiederum jeweils 10 000 Einzelschätzungen. Die Verteilungen der Schätzwerte sind in der Abbildung 1 in Form von Box-Plots dargestellt¹⁶⁾. Zusätzlich zu den Verteilungen sind in der Figur auch die tatsächlichen Werte in Form von Kreuzen gekennzeichnet. Bereits auf den ersten Blick

15) Siehe Box 1 „Stichprobenfehler, Schichtung, relativer Standardfehler und Mean Squared Error (MSE)“.

16) Bei jedem der 52 dargestellten Plots kennzeichnet die dickere Linie die mittleren 50 % der jeweiligen Schätzwerte. Die an das Rechteck oben und unten anschließenden dünnen Linien umfassen die höchsten und die niedrigsten 25 % der Schätzwerte.

Abbildung 1: Direkte Schätzung und Area-Level-Schätzung im Vergleich – Ergebnisse der Simulationsrechnung



lässt sich erkennen, dass die blau dargestellten FH-Schätzwerte deutlich weniger streuen als die direkten Schätzwerte. Aus den Simulationsergebnissen errechnet, liegen die relativen Standardfehler der FH-Schätzungen je nach betrachtetem Landkreis bzw. betrachteter kreisfreier Stadt zwischen 3,3 % und 9,5 %. Area-Level-Schätzungen führen somit in der Tat zu stabileren Ergebnissen.

Im Idealfall verteilen sich dabei die geringer streuenden FH-Schätzwerte mehr oder minder symmetrisch um den tatsächlichen Wert. Diese Situation ist in Abbildung 1 beispielsweise bei den Schätzungen für den Landkreis Bergstraße, für Darmstadt-Dieburg oder für den Main-Kinzig-Kreis zu beobachten. Es treten jedoch auch Fälle auf, bei denen die Stabilisierung ober- oder unterhalb des tatsächlichen Wertes erfolgt. In den allermeisten Fällen halten sich diese „Verschiebungen“ jedoch in Grenzen, sodass die FH-Schätzung den tatsächlichen Wert sehr wahrscheinlich nicht genau trifft, aber mit hoher Wahrscheinlichkeit näher am tatsächlichen Wert liegt als die direkte Schätzung. Andererseits zeigen die Ergebnisse für den Landkreis Hersfeld-Rotenburg und vor allem diejenigen für den Schwalm-Eder-Kreis, dass die FH-Schätzungen hier den tatsächlichen Wert allesamt unterschätzen. In Einzelfällen kann dies bei direkten Schätzungen ebenfalls auftreten, die Wahrscheinlichkeit dafür ist aber wesentlich geringer.

Über alle getätigten Schätzungen hinweg betrachtet, liegt der FH-Schätzwert in 69 % der Fälle näher am Vorgabewert als der direkte Schätzwert. Dabei fallen bei der FH-Schätzung die Abweichungen des Schätzwerts zum tatsächlichen Wert um durchschnittlich 59 % geringer aus als bei der direkten Schätzung. In 31 % der Fälle stellt sich das umgekehrte Resultat ein und die FH-Schätzung trägt durchschnittlich 51 % zur größeren Abweichung bei. Verrechnet man wiederum positive wie negative Veränderungen so führt die FH-Schätzung gegenüber der direkten Schätzung zu durchschnittlich 25 % geringeren Abweichungen.

Das insgesamt gute Abschneiden der Area-Level-Schätzung wird nicht zuletzt von der Güte der Hilfsvariablen, genauer gesagt, der hohen Korrelation zwischen Hilfs- und Zielvariablen getragen. In einem weiteren Simulationslauf wurde deshalb der Einfluss der Korrelation auf die Schätzergebnisse geprüft. Hierzu wurde die Hilfsvariable so konstruiert, dass der Korrelationskoeffizient von ursprünglich 0,96 auf nunmehr 0,68 absinkt¹⁷⁾. Ohne auf die Ergebnisse im Einzelnen einzugehen, führte die Area-Level-Schätzung unter diesen Voraussetzungen in immerhin 61 % der Fälle zu im Vergleich zur direkten Schätzung näher am tatsächlichen Wert liegenden Schätzwerten. Gegenüber der direkten Schätzung sind die Abweichungen um durchschnittlich 36 % kleiner. Bei 39 % der Schätzungen ergab hingegen die Area-Level-Schätzung die größeren Abweichungen, wobei die Vergrößerung im Schnitt 44 % der (nun größeren) Abweichung ausmachte. Neben der abgesunkenen Präzision der Area-Level-Schätzung steigt auch die Streuung der einzelnen Schätzwerte. Sie liegt aber auch unter diesen Voraussetzungen deutlich unterhalb der Streuung der direkten Schätzwerte.

Zusammenfassung und Fazit

Der Beitrag befasste sich mit der Frage, wie gut eine Area-Level-Schätzung gegenüber dem klassischen Schätzverfahren – einer direkten Schätzung – abschneidet. Insbesondere wenn es darum geht, tief gegliederte Ergebnisse zu schätzen, bei denen das Gliederungsmerkmal nicht bei der Konzeption der Stichprobe berücksichtigt wurde. Als Anwendungsfall sollten die im Statistischen Unternehmensregister für Hessen ausgewiesenen, auf Ebene der Landkreise und kreisfreien Städte aggregierten Einzelhandelsumsätze auf der Basis einer Stichprobe geschätzt werden. Dieser Ansatz hat den Vorteil, dass die „wahren Werte“ vorab bekannt sind und mit den jeweiligen Schätzergebnissen verglichen werden können. Für die Area-Level-Schätzung wurde eine einzige, aus den Daten erzeugte Hilfsvariable verwendet. Stichprobenziehung und Schätzung wurden mehrfach wiederholt, um eine umfassendere Einschätzung der Ergebnisse zu erhalten.

Wenn die Korrelation zwischen der zu schätzenden Größe und der Hilfsvariablen sehr hoch ist, dann führt die Area-Level-Schätzung sowohl zu insgesamt präziseren als auch zu stabileren Schätzungen. Gegenüber der direkten Schätzung verbessern sich die Schätzergebnisse jedoch nicht gleichmäßig. Für jede Stichprobe sind einige Schätzwerte zu verzeichnen, bei denen die direkte Schätzung den Vorgabewert besser trifft als die Area-Level-Schätzung. In einzelnen Fällen unterschätzt der Area-Level-Schätzer bei mehrfacher Wiederholung der Stichprobenziehung konsequent den tatsächlichen Wert. Dieses Ergebnis ist zum Teil dem etwas mechanischen Vorgehen in der Simulation geschuldet. Im konkreten Anwendungsfall würde man auf Basis der (einen) vorliegenden Stichprobe mehr Sorgfalt auf die Passgenauigkeit des Modells legen und gegebenenfalls im Regressionsteil Ausreißer aus der Parameterschätzung ausschließen.

Insgesamt legen die erzielten Resultate die Schlussfolgerung nahe, dass Area-Level-Schätzungen unter den hier vorliegenden Voraussetzungen das geeignetere Schätzverfahren zur Erzielung von kleinräumigen Ergebnissen sind.

Anhang: In der Simulation getestete Schätzgleichungen zur Bestimmung der Varianz des direkten Summenschätzers

Notation:

\hat{V}	Geschätzte Varianz	\hat{U}_d	Geschätzte Umsatzsumme von Landkreis bzw. kreisfreier Stadt d
π_i	Inklusionswahrscheinlichkeit erster Ordnung für Einheit i	s_d	Indexmenge der Stichprobeneinheiten von Landkreis bzw. kreisfreier Stadt d
π_{ij}	Inklusionswahrscheinlichkeit zweiter Ordnung für die Einheiten i und j	y_i	Umsatz von Unternehmen i
$\pi_{l,i}$	Inklusionswahrscheinlichkeit erster Ordnung für Einheit i aus Schicht l	$y_{l,i}$	Umsatz von Unternehmen i in Schicht l
\hat{V}_l	Varianzschätzung für Schicht l	n	Stichprobenumfang
f_l	Auswahlsatz für Schicht l	n_l	Umfang der Teilstichprobe für Schicht l

Schätzgleichungen I und II:

Aus der Varianz des HT-Summenschätzers ergibt sich unmittelbar die erste getestete Schätzgleichung

$$\hat{V}^{HT}(\hat{U}_d) = \sum_{i \in s_d} (1 - \pi_i) \left(\frac{y_i}{\pi_i} \right)^2 + 2 \sum_{i \in s_d} \sum_{\substack{j \in s_d \\ j > i}} \left(1 - \frac{\pi_i \pi_j}{\pi_{ij}} \right) \frac{y_i y_j}{\pi_i \pi_j}$$

17) Die geringere Korrelation wurde durch Ersetzung eines höheren Anteils von Einzelwerten mit den Werten einer gleichverteilten Zufallsvariablen erzielt.

Hierbei kennzeichnen die π_i die Inklusionswahrscheinlichkeiten erster Ordnung – also die Wahrscheinlichkeit, dass Einheit i in die Stichprobe gezogen wird. In einfachen Modellen entspricht diese in der Regel dem Auswahlatz für die Schicht zu der Einheit i gehört. Entsprechend kennzeichnet π_{ij} die Wahrscheinlichkeit, dass die Einheiten i und j gemeinsam in die Stichprobe gezogen werden. Diese Inklusionswahrscheinlichkeiten zweiter Ordnung wurden auf 2 unterschiedliche Arten bestimmt: In Variante I wurde $\pi_{ij} = \frac{n_{Li} \cdot n_{Lj} - 1}{N_{Li} \cdot N_{Lj} - 1}$ gewählt, falls die Einheiten i und j aus derselben Schicht stammen, ansonsten $\pi_{ij} = \pi_i \pi_j$; in Variante B wurde unabhängig von der Schichtzugehörigkeit der Einheiten vereinfacht durchweg $\pi_{ij} = \pi_i \pi_j$ unterstellt. Somit entfällt in Variante II der zweite Term auf der rechten Seite der obigen Schätzgleichung. Variante B ist beispielsweise in der einschlägigen Schätzgleichung des R-packages „sae“¹⁸⁾ umgesetzt.

Schätzgleichung III

Als dritte Alternative wurde die in Lehtonen und Veijanen (2009)¹⁹⁾ angegebene Schätzgleichung für Domain-Analysen getestet, die für den Fall vorgeschlagen wird, dass die betrachteten Domains – hier die Landkreise und kreisfreien Städte – nicht in der Stichprobenkonzeption berücksichtigt wurden. Die Schätzgleichung lautet

$$\hat{V}^{LV}(\hat{U}_d) = \frac{n}{n-1} \sum_{i \in S} \left(I_{d,i} \frac{y_i}{\pi_i} - \frac{\hat{\tau}_d}{n} \right)^2 \quad \text{mit} \quad \hat{\tau}_d = \sum_{i \in S} I_{d,i} \frac{y_i}{\pi_i}.$$

Demnach wird die Varianz des Schätzwerts \hat{U}_d unter Einbeziehung aller Stichprobenwerte bestimmt, wobei die Indikatorvariable $I_{d,i}$ die Kreiszugehörigkeit kennzeichnet. Sie nimmt den Wert 1 an, wenn die Einheit i zu Landkreis bzw. kreisfreier Stadt d gehört und den Wert 0 wenn nicht.

18) R ist ein kostenfrei verfügbares und im wissenschaftlichen Umfeld weit verbreitetes statistisches Softwarepaket, welches für zahlreiche statistische Aufgaben spezielle Prozeduren bzw. Packages – so auch das Package „sae“ (Small-Area-Estimation) – bereitstellt.

19) Lehtonen, R. und Veijanen, A. (2009): Design-based Methods of Estimation for Domains and Small Areas. Sample Surveys: Inference and Analysis, Vol. 29B, S. 227.

Schätzgleichungen IV und V:

Die in der SAS Prozedur SURVEYMEANS implementierte Schätzgleichung setzt die Varianzschätzung aus den Teilvarianzen in den einzelnen Schichten zusammen:

$$\hat{V}^{SAS}(\hat{U}_d) = \sum_{l=1}^L \hat{V}_l(\hat{U}_d),$$

wobei der Index $l = 1, \dots, L$ die Schichten kennzeichnet. Erfolgt die Varianzschätzung für Domains bzw. Landkreise und kreisfreie Städte, dann kennzeichnet auch hier eine Indikatorvariable $I_{l,i}$ die Domain- bzw. Kreiszugehörigkeit. Die Variable nimmt den Wert 1 an, wenn die betrachtete Einheit i von Schicht l zu Domain d gehört, ansonsten den Wert 0. Die Teilvarianz errechnet sich über

$$\hat{V}_l(\hat{U}_d) = \frac{n_l(1-f_l)}{n_l-1} \sum_{i=1}^{n_l} \left(I_{l,i} \frac{y_{l,i}}{\pi_{l,i}} - \bar{y}_l \right)^2 \quad \text{mit} \quad \bar{y}_l = \frac{1}{n_l} \sum_{i=1}^{n_l} I_{l,i} \frac{y_{l,i}}{\pi_{l,i}}.$$

Wird in SURVEYMEANS eine Varianzschätzung gemäß Taylor-Approximation für eine Domain-Analyse gewählt, greift die Prozedur auf diese Schätzformel zurück. Bei Angabe der Besetzungszahlen der Schichten in der Grundgesamtheit (Verwendung der Option TOTAL = in Verbindung mit dem Statement STRATA in der Prozedur) werden die Schätzgleichungen direkt umgesetzt. Ohne die Angabe geht die Prozedur davon aus, dass es nur eine Schicht gibt. Man bemerkt die ähnliche Struktur zwischen der obigen Schätzgleichung und Schätzgleichung III. In der Tat führen beide zum selben Ergebnis, wenn in Gleichung IV/V sowohl auf den Korrekturfaktor $(1-f_l)$ als auch auf eine schichtweise Berechnung der Teilvarianzen verzichtet wird.

Peter Gottfried; Tel: 0611 3802-868;
E-Mail: peter.gottfried@statistik.hessen.de