

Dipl.-Ökonom Christopher Nölting

Ermittlung von Kreisergebnissen in der Agrarstatistik

Anwendung klassischer Verfahren der gebundenen Hochrechnung

In Deutschland wurden in der Vergangenheit die zeitlichen Abstände zwischen Vollerhebungen in der amtlichen Agrarstatistik zunehmend größer. Die bislang letzte agrarstatistische Vollerhebung war die Haupterhebung der Landwirtschaftszählung 2010. Im Jahr 2016 wird es eine Strukturerhebung der landwirtschaftlichen Betriebe geben, wobei die Merkmale zur Bodennutzung und zum Viehbestand allgemein erhoben werden.

Bei diesen Statistiken handelt es sich um äußerst umfangreiche und aufwendige Erhebungen, da alle landwirtschaftlichen Betriebe, die eine bestimmte Größe überschreiten, befragt werden. Die Ergebnisse dieser Statistiken zeichnen ein umfassendes und aktuelles Bild der Landwirtschaft zu diesem Zeitpunkt. Darüber hinaus ist eine tief gegliederte regionale Darstellung der Ergebnisse zum Beispiel auf Kreis- und Gemeindeebene möglich.

Die mit der Durchführung von Vollerhebungen verbundenen hohen Kosten und Belastungen der Auskunftspflichtigen sowie die immer knapperen finanziellen Ressourcen der statistischen Ämter bewirken die zunehmenden zeitlichen Abstände zwischen agrarstatistischen Vollerhebungen. Aktuelle regionale Daten aus der Landwirtschaft sind allerdings nach wie vor gefragt. Das Johann Heinrich von Thünen-Institut berechnet zum Beispiel Treibhausgasemissionen aus der Landwirtschaft.¹ Für diese Berechnungen werden teilweise aktuelle regionale Daten zum Viehbestand benötigt. Auch in politischen Diskussionen sind regionale

Emissionen in der Landwirtschaft im Bereich Klima- und Umweltschutz Thema. Hier können aktuelle regionale Daten eine Diskussionsgrundlage bilden. Die zunehmenden zeitlichen Abstände, in denen Vollerhebungen durchgeführt werden, erschweren die Versorgung mit aktuellen regionalen Daten.

Die amtliche Agrarstatistik führt regelmäßig Stichproben-erhebungen durch und befragt dabei nur eine Teilmenge der Erhebungsgesamtheit. Die halbjährlich durchgeführte Viehbestands-erhebung Schweine stellt eine solche Stichproben-erhebung dar. Mithilfe der so erhobenen Daten können verlässliche Zahlen für den Schweinebestand für Deutschland sowie für die einzelnen Bundesländer geschätzt werden. Schätzungen für Regionen, Landkreise oder gar Gemeinden sind aufgrund des geringen Stichprobenumfangs mit einer freien Hochrechnung im Allgemeinen nicht hinreichend präzise möglich.

Dieser Aufsatz untersucht mithilfe der Viehbestands-erhebung Schweine in Niedersachsen und Bayern, inwieweit die klassischen Verfahren der gebundenen Hochrechnung zu einer höheren Präzision der Schätzungen für Kreis-ergebnisse für Schweine und Zuchtsauen beitragen können. Zunächst werden die verwendeten Daten, insbesondere der Auswahlplan der Stichprobe, beschrieben. Anschließend werden die Verfahren, mit denen die Kreisschätzungen berechnet wurden, vorgestellt. Es wird gezeigt, welches Verfahren sich für welche Datenlage eignet. Die Ergebnisse werden kurz dargestellt und anhand des Kriteriums des relativen Standardfehlers miteinander verglichen. Der Aufsatz schließt ab mit einem Ausblick, welche Richtung künftige Untersuchungen nehmen könnten.

¹ Siehe Haanel, H.-D. (Herausgeber): „Berechnung der Emissionen aus der deutschen Landwirtschaft – Nationaler Emissionsbericht (NIR) 2010 für 2008“, Braunschweig 2010, Sonderheft 334, Seite 51 ff.

1 Beschreibung der verwendeten Daten

Die für die Untersuchung verwendeten Daten sind die Auswahlgesamtheiten der Viehbestandserhebung Schweine für Niedersachsen und Bayern der jeweiligen Statistischen Landesämter, die daraus gezogenen Zufallsstichproben der Viehbestandserhebung Schweine zum 3. November 2013 aus den Bundesländern Niedersachsen und Bayern sowie ein Datenauszug zum Stichtag 1. Januar 2013 des Viehbestands Schweine aus der Verwaltungsdatenbank Herkunftssicherungs- und Informationssystem für Tiere (HIT).

Die Daten der Viehbestandserhebung Schweine aus den Bundesländern Niedersachsen sowie Bayern wurden zunächst verwendet, um die Qualität von frei hochgerechneten Landkreisergebnissen für den Viehbestand Schweine und Zuchtsauen einzuschätzen. Darüber hinaus wurden für Bayern und Niedersachsen die zur oben genannten Erhebung gehörenden Auswahlgesamtheiten aus dem Zentralen Betriebsregister der Agrarstatistiken (zeBRA) abgerufen. Die zur Durchführung der gebundenen Hochrechnung benötigten Hilfsinformationen stammen aus der Verwaltungsdatenbank Herkunftssicherungs- und Informationssystem für Tiere. Die gebundenen Hochrechnungen wurden sowohl für Niedersachsen als auch für Bayern zur Schätzung von Totalwerten für den Viehbestand Schweine und Zuchtsauen durchgeführt.

Auswahlgesamtheit zur Viehbestandserhebung Schweine

Die Auswahlgesamtheit zur Viehbestandserhebung Schweine bilden alle im Zentralen Betriebsregister der Agrarstatistiken enthaltenen Betriebe, die dort mit mindestens 50 Schweinen oder mindestens 10 Zuchtsauen registriert sind. Aus dieser Gesamtheit wird die jeweilige Stichprobe gezogen.

Die Betriebe werden im Zentralen Betriebsregister der Agrarstatistiken in der Regel mit den aktuellsten verfügbaren Werten gelistet, also Werten aus vergangenen Stichproben- oder Vollerhebungen. Dies ist Voraussetzung für eine möglichst zuverlässige Schichteinteilung bei geschichteten Stichproben. Gegebenenfalls werden von den Statistischen Ämtern der Länder auch Werte aus Verwaltungsdatenquellen zur Pflege des Registers verwendet. Gibt ein Betrieb zum Stichtag an, dass er nur vorübergehend einen Viehbestand unterhalb der Erfassungsgrenze hat, so bleibt der alte Wert im Register stehen, damit künftig eine zutreffende Schichtzuordnung möglich ist.

Stichprobe Viehbestandserhebung Schweine

Generell handelt es sich bei der Viehbestandserhebung Schweine² um eine nach Bundesländern und Größenklas-

sen geschichtete Stichprobe, die halbjährlich, jeweils zum Stichtag 3. Mai und 3. November, durchgeführt wird. Die Rechtsgrundlagen bilden das Gesetz über Agrarstatistiken³, das Gesetz über die Statistik für Bundeszwecke⁴ sowie die EU-Verordnung Nr. 1165/2008⁵. Im Rahmen der Viehbestandserhebung werden Bestandsgrößen wie zum Beispiel Schweine insgesamt, Zuchtsauen, Mastschweine und Eber abgefragt. Die Schichtungsmerkmale sind zum einen das Merkmal Schweine insgesamt (Y^{SW}) und zum anderen das Merkmal Zuchtsauen (Y^{ZS}). Geschichtet werden Betriebe der Auswahlgesamtheit je Bundesland. Mit N_h wird im Folgenden der Umfang der Schicht h in der Auswahlgesamtheit bezeichnet und mit n_h der Stichprobenumfang der Schicht h . Die Auswahlgesamtheit bilden alle bekannten Gesamtbetriebe N , die mindestens 50 Schweine oder mindestens 10 Zuchtsauen im Bestand haben. In der Praxis handelt es sich dabei um Betriebe, die im Zentralen Betriebsregister der Agrarstatistiken der jeweiligen Bundesländer gespeichert sind. Da es nur zulässig ist, Schichtungsmerkmale im Zentralen Betriebsregister der Agrarstatistiken zu speichern, beschränken sich hier die Merkmale auf Schweine insgesamt sowie Zuchtsauen. Zur Ziehung der Stichprobe sind die Betriebe der Auswahlgesamtheit je Schicht nach einer fünfstelligen Kreiskennziffer aufsteigend sortiert. Der Begriff Kreise wird in diesem Aufsatz verwendet, wenn sowohl Landkreise als auch kreisfreie Städte gemeint sind. Je Schicht werden dann grundsätzlich gleich große aufeinanderfolgende Zonen mit m_h aufeinanderfolgenden Auswahlseinheiten (wobei $m_h \leq N_h$) gebildet. Aus diesen Zonen werden gemäß vorgegebenem Auswahlssatz die Stichprobeneinheiten per Zufallsgenerator bestimmt. Tendenziell ist der Auswahlssatz in den Schichten mit großen Betrieben (zum Beispiel 5 000 Schweine und mehr) höher (teilweise auch 100 %), da hier ein höherer schichtspezifischer Standardfehler einen größeren Effekt auf den landesspezifischen Standardfehler hätte. Die Bildung der Zonen hat in Kombination mit der Anordnung der Betriebe nach Kreisen den gewünschten Effekt, dass die Landkreise annähernd proportional in der Schicht repräsentiert sind. Sind beim Auswahlssatz in der Form x aus m_h die Werte für x und m_h sehr klein, so kommt der Anordnungseffekt besonders stark zur Geltung. Durch das zufällige Auswählen innerhalb von Zonen werden gewisse Nachteile der systematischen Auswahl mit Zufallsstart vermieden.⁶

In dieser Untersuchung wurde die Stichprobe Viehbestand Schweine zum Stichtag 3. November 2013 von Niedersachsen und Bayern verwendet. Laut Auswahlgesamtheit gab es am Stichtag in Niedersachsen 8 605 relevante Betriebe (= M) und in Bayern 7 620 Betriebe (siehe Tabelle 1). Der Stichprobenumfang n beträgt für Niedersachsen 3 641 Betriebe und für Bayern 2 432 Betriebe. In Niedersach-

² Siehe „Erhebung über die Schweinebestände – Qualitätsbericht“ auf der Internetseite des Statistischen Bundesamtes unter www.destatis.de, im Bereich Publikationen > Qualitätsberichte > Land- & Forstwirtschaft > Viehbestand.

³ Gesetz über Agrarstatistiken (Agrarstatistikgesetz – AgrStatG) in der Fassung der Bekanntmachung vom 17. Dezember 2009 (BGBl. I Seite 3886), zuletzt geändert durch Artikel 13 Absatz 5 des Gesetzes vom 12. April 2012 (BGBl. I Seite 579).

⁴ Gesetz über die Statistik für Bundeszwecke (Bundesstatistikgesetz – BStatG) vom 22. Januar 1987 (BGBl. I Seite 462, 565), zuletzt geändert durch Artikel 13 des Gesetzes vom 25. Juli 2013 (BGBl. I Seite 2749).

⁵ Verordnung (EG) Nr. 1165/2008 des Europäischen Parlaments und des Rates vom 19. November 2008 über Viehbestands- und Fleischstatistiken und zur Aufhebung der Richtlinien 93/23/EWG, 93/24/EWG und 93/25/EWG des Rates (Amtsblatt der EU Nr. L 321 Seite 1).

⁶ Siehe Krug, W./Nourney, M./Schmidt, J.: „Wirtschafts- und Sozialstatistik“, 6. Auflage, München 2001, Seite 94 f.

Tabelle 1 Relevante Betriebe und Kreise in Bayern und Niedersachsen

| | Relevante Betriebe ¹ | | Kreise | | |
|-----------------|---------------------------------|-------------------|-----------|--------------------------|-------------------|
| | laut Grundgesamtheit | in der Stichprobe | insgesamt | mit relevanten Betrieben | in der Stichprobe |
| Bayern ... | 7 620 | 2 432 | 96 | 86 | 79 |
| Niedersachsen . | 8 605 | 3 641 | 46 | 45 | 42 |

Stichtag: 3. November 2013.

¹ Mit mindestens 50 Schweinen oder mindestens 10 Zuchtsauen im Bestand.

sen gibt es 38 Landkreise und 8 kreisfreie Städte, sodass es hier insgesamt 46 verschiedene Kreiskennziffern gibt. Laut Auswahlgesamtheit existieren in 45 Kreisen relevante Betriebe. Stichprobenbetriebe gibt es in 42 Kreisen, sodass 3 Kreise nicht in der Stichprobe repräsentiert sind, obwohl es relevante Betriebe gibt. In Bayern gibt es 71 Landkreise und 25 kreisfreie Städte, also 96 verschiedene Kreiskennziffern. Davon existieren in 86 Kreisen laut Auswahlgrundlage relevante Betriebe, von diesen 86 Kreisen sind 79 Kreise mit Stichprobenbetrieben vertreten. Dies entspricht einer Differenz von 7 Kreisen, in denen relevante Betriebe existieren, diese jedoch nicht in der Stichprobe vertreten sind. Für die nicht in der Stichprobe repräsentierten 3 Kreise in Niedersachsen und 7 Kreise in Bayern ist weder eine freie noch eine gebundene Hochrechnung mit den untersuchten Verfahren möglich. Jedoch ist wegen des zuvor beschriebenen Auswahlplans und des Ziehungsprozesses zu erwarten, dass es sich hier um Kreise handelt, in denen es nur wenige relevante Betriebe gibt. Laut Auswahlgesamtheit existieren in den 3 niedersächsischen Kreisen insgesamt 10 relevante Betriebe und in den 7 bayerischen Landkreisen insgesamt 12 relevante Betriebe. In solchen Kreisen ist es äußerst fragwürdig, ob eine Veröffentlichung von Ergebnissen überhaupt zulässig wäre; Totalergebnisse müssen hier in der Regel geheim gehalten werden.

Verwaltungsdaten aus der HIT-Datenbank für Schweine

Die Halter von Schweinen sind nach § 26 der Viehverkehrsordnung dazu verpflichtet, zum Stichtag 1. Januar eines jeden Jahres die Anzahl der im Bestand vorhandenen Zuchtschweine, Mastschweine sowie Ferkel an die zuständige Behörde zu übermitteln. Die Angaben werden dann zentral in der HIT-Datenbank⁷ erfasst. Die Registriernummer (HIT-Nummer), mit der sich der Schweinehalter für die Datenbank registriert, entspricht nicht der Betriebsnummer im Zentralen Betriebsregister der Agrarstatistiken. In der amtlichen Agrarstatistik gilt nach § 91 Absatz 4 in Verbindung mit Absatz 4a Bundesstatistikgesetz das Betriebssitzprinzip. Der Betriebssitz ist in der Regel das Grundstück, auf dem sich das wichtigste Wirtschaftsgebäude befindet. Besteht der Betrieb aus mehreren voneinander entfernt liegenden Betriebsteilen, so sind die Meldungen dem Betriebssitz zuzuordnen. Daraus können im Einzelfall räumliche Ver-

zerrungen für die Viehbestände resultieren. Die Zuordnung der HIT-Nummern zum zugehörigen Betrieb (Betriebssitz) erfolgt in Bayern und Niedersachsen durch die jeweiligen Statistischen Landesämter. Beim verwendeten Datenmaterial konnten für Niedersachsen etwa 70 % der Betriebe die zugehörigen HIT-Nummern und Bestände zugeordnet werden, in Bayern für etwa 87 % der Betriebe (siehe Tabelle 2). Dass HIT-Werte fehlen, resultiert zum Teil daraus, weil ein nicht aktueller Zuordnungsschlüssel verwendet wurde. Ein aktueller Zuordnungsschlüssel lag zum Zeitpunkt der Untersuchung nicht vor. Eine andere mögliche Erklärung wäre, dass Betriebe aus der Grundgesamtheit nicht in der HIT-Datenbank registriert sind. Diese Möglichkeit lässt sich ohne einen vollständigen Zuordnungsschlüssel nicht exakt überprüfen.

Tabelle 2 Betriebe in der Auswahlgrundgesamtheit und mit zugeordnetem HIT-Wert

| | Betriebe in der Auswahlgrundgesamtheit | Betriebe mit zugeordnetem HIT-Wert | |
|-------------------|--|------------------------------------|------|
| | Anzahl | | % |
| Bayern | 7 619 | 6 614 | 86,8 |
| Niedersachsen ... | 8 605 | 6 025 | 70,0 |

2 Freie und gebundene Hochrechnungsverfahren zur Schätzung von Kreisergebnissen

Das Stichprobendesign der Viehbestandserhebung Schweine zielt darauf ab, präzise Schätzungen mittels freier Hochrechnung auf Bundeslandebene zu gewährleisten. Geschichtet wird innerhalb von Deutschland auf NUTS-1-Ebene (Bundesländer) und nach Größenklassen von Schweinebeständen. Eine Ausnahme bildet Nordrhein-Westfalen, welches auf NUTS-2-Ebene (Regionen) schichtet. Der Stichprobenumfang je Schicht ist durch das Stichprobendesign fest vorgegeben. Für Kreise gilt dies beim derzeitigen Stichprobendesign nicht. Einzelne Kreise können in mehreren Schichten vertreten sein. Der Stichprobenumfang auf Kreisebene ist nicht fix. Es resultiert ein zusätzlicher fehlervergrößernder Effekt, da die Stichprobenumfänge für einzelne Kreise dem Zufall überlassen bleiben. Um hinreichend präzise Kreisergebnisse mittels freier Hochrechnung zu erhalten, müssten die Auswahlsätze in den Schichten deutlich erhöht werden, sodass alle Kreise mit einer hinreichend hohen Anzahl an Betrieben in der Stichprobe vertreten sind. Praktisch würde das bedeuten, dass nahezu alle Betriebe befragt werden müssten. Mit dem in § 19 Agrarstatistikgesetz gesetzlich festgelegten maximalen Stichprobenumfang für die Viehbestandserhebungen ist eine solche Befragung nicht möglich.

Im Folgenden wird zunächst die freie Hochrechnung in der Form vorgestellt, in der sie im Rahmen der Viehbestandserhebung Schweine zur Schätzung des Bestands auf Bundeslandebene verwendet wird. Daran schließt sich die Darstellung der freien Hochrechnung für Kreisergebnisse an. Die Verfahren der gebundenen Hochrechnung werden dann ausschließlich für Kreisergebnisse vorgestellt.

⁷ Siehe Kokott, K.: „HI Tier – Info Schweinedatenbank“, im Internet unter www.hi-tier.de/Entwicklung/Konzept/Sonstiges/schweine002.htm (abgerufen am 3. November 2014) zu finden.

Freie Hochrechnung bei geschichteter Stichprobe für Landes- und Kreisergebnisse

Die freie Hochrechnung (Horvitz-Thompson-Schätzer) ist ein allgemeines statistisches Schätzverfahren, das lediglich Informationen aus einer Erhebung zur Schätzung von Total- oder Mittelwerten verwendet. Das Verfahren wird häufig in der amtlichen Statistik angewandt (in der Agrarstatistik beispielsweise bei der Viehbestandserhebung Schweine und der Bodennutzungshaupterhebung), da es sich um ein erwartungstreu (unverzerrtes) Schätzverfahren⁸ handelt, zur Schätzung nur die Auswahlwahrscheinlichkeiten und die Werte der Erhebungsmerkmale der Stichprobeneinheiten benötigt werden und es einfach zu handhaben ist. Die Eingängigkeit des Verfahrens ermöglicht es zudem, den Nutzern den Entstehungsprozess der Ergebnisse zu erklären.

Der Wert des Merkmals Anzahl Schweine (beziehungsweise Zuchtsauen) im Betrieb i der Schicht h ($h = 1, \dots, H_L$) wird mit y_{hi} bezeichnet. Die Menge aller Stichprobenbetriebe der Schicht h wird mit S_h bezeichnet. Der schichtspezifische Hochrechnungsfaktor w_h , der auch im Stichprobenmaterial vorhanden ist, ist definiert als:

$$(1) \quad w_h = \frac{N_h}{n_h}$$

Der Horvitz-Thompson-Schätzer für den Totalwert t_L des Landes L ist dann definiert durch:

$$(2) \quad \hat{t}_L^{HT} = \sum_{h=1}^{H_L} w_h \sum_{i \in S_h} y_{hi}$$

Als Maß für die Präzision der Schätzungen wird der Standardfehler verwendet. Je kleiner der Standardfehler ist, desto präziser ist der zugehörige Schätzwert. Damit die Güte der Schätzungen für unterschiedliche Bundesländer verglichen werden kann, wird der Standardfehler ins Verhältnis zum geschätzten Totalwert gesetzt. Dabei handelt es sich um den relativen Standardfehler (RSF). Der Standardfehler der freien Hochrechnung auf Landesebene wird geschätzt durch:

$$(3) \quad \hat{s}(\hat{t}_L^{HT}) = \sqrt{\sum_{h=1}^{H_L} N_h (w_h - 1) s_{hy_L}^2}$$

und der relative Standardfehler

$$RSF = \frac{\hat{s}(\hat{t}_L^{HT})}{\hat{t}_L^{HT}}$$

wobei

$$(4) \quad s_{hy_L}^2 = \frac{1}{n_h - 1} \sum_{i \in S_h} (y_{hi} - \bar{y}_h)^2$$

die Varianz der Einzelwerte in der Schicht h und

$$\bar{y}_h = \frac{1}{n_h} \sum_{i \in S_h} y_{hi}$$

die Mittelwerte der Einzelwerte in der Schicht h bezeichnet⁹.

Um das Ergebnis der freien Hochrechnung für Kreis d zu erhalten, werden nur die hochgerechneten Merkmalswerte summiert, die zum Kreis d gehören. Daher ergibt sich als Schätzung für den Totalwert des Kreises t_d :

$$(5) \quad \hat{t}_d^{HT} = \sum_{h=1}^{H_L} w_h \sum_{i \in S_h} y_{hid}$$

mit

$$y_{hid} = \begin{cases} y_{hi} & \text{für } i \in d \\ 0 & \text{sonst} \end{cases}$$

Insbesondere bei der Berechnung des Standardfehlers muss berücksichtigt werden, dass nur Merkmalswerte, die zum Kreis d gehören, mit einem Wert von Null verschieden eingehen. Dabei ist zu beachten, dass sich die Berechnung der Varianz der Einzelwerte von Kreis d in Schicht h formal über alle n_h Stichprobenbetriebe erstreckt. Die Formel zur erwartungstreuen Schätzung des Standardfehlers des Totalwertes für den Kreis d lautet¹⁰:

$$(6) \quad s(\hat{t}_d^{HT}) = \sqrt{\sum_{h=1}^{H_L} N_h (w_h - 1) s_{hy_d}^2}$$

wobei

$$(7) \quad s_{hy_d}^2 = \frac{1}{n_h - 1} \sum_{i \in S_h} (y_{hid} - \bar{y}_{hd})^2$$

mit

$$\bar{y}_{hd} = \frac{1}{n_h} \sum_{i \in S_h} y_{hid}$$

Um die Programmierung zu erleichtern, kann Formel (7) umgeformt werden zu:

$$(8) \quad s_{hy_d}^2 = \frac{1}{n_h - 1} \left(\sum_{i \in S_h} y_{hid}^2 - \frac{1}{n_h} \left(\sum_{i \in S_h} y_{hid} \right)^2 \right)$$

Gebundene Hochrechnungsverfahren bei geschichteter Stichprobe für Kreisergebnisse

Anders als bei der freien Hochrechnung werden bei Verfahren der gebundenen Hochrechnung zusätzlich Hilfsinformationen verwendet. Mit x_i wird im Folgenden der Wert des Hilfsmerkmals des i -ten Betriebs bezeichnet. Der Totalwert des Hilfsmerkmals für den Kreis d wird mit X_d bezeichnet. Die Grundidee der gebundenen Hochrechnung ist es, die Hilfsinformationen in der Schätzfunktion zu nutzen, um die Präzision des Schätzergebnisses zu verbessern. Dabei werden im Wesentlichen schichtweise Unterschiede zwischen frei hochgerechneten Totalwerten des Hilfsmerkmals und bekannten Totalwerten des Hilfsmerkmals zur Korrektur von frei hochgerechneten Totalwerten des Erhebungsmerk-

⁸ Siehe Münnich, R./Burgard, J.P./Vogt, M.: „Small Area-Statistik: Methoden und Anwendungen“ in Wirtschafts- und Sozialstatistisches Archiv, Band 6, Hefte 3-4, Jahr 2013, Seite 149 ff., hier: Seite 154.

⁹ Siehe Särndal, C.-E./Swensson, B./Wretman, J.: „Model Assisted Survey Sampling“, 1. Auflage, New York 1992, Seite 103.

¹⁰ Siehe Krug, W./Nourney, M./Schmidt, J. (Fußnote 6), hier: Seite 116 f.

mals verwendet. In dieser Arbeit wurden drei unterschiedliche Verfahren der gebundenen Hochrechnung untersucht. Zunächst wird die Verhältnisschätzung vorgestellt, anschließend die Differenzschätzung und letztlich die Regressionschätzung. Unabdingbare Anwendungsvoraussetzung für die hier vorgestellten Verfahren ist das Vorliegen von Werten eines Hilfsmerkmals (Bezugsmerkmale) für alle Einheiten der Auswahlgesamtheit. Es gilt für alle Verfahren, dass der Nutzen der Hilfsmerkmale von der Korrelation zwischen Erhebungs- und Hilfsmerkmal abhängt.

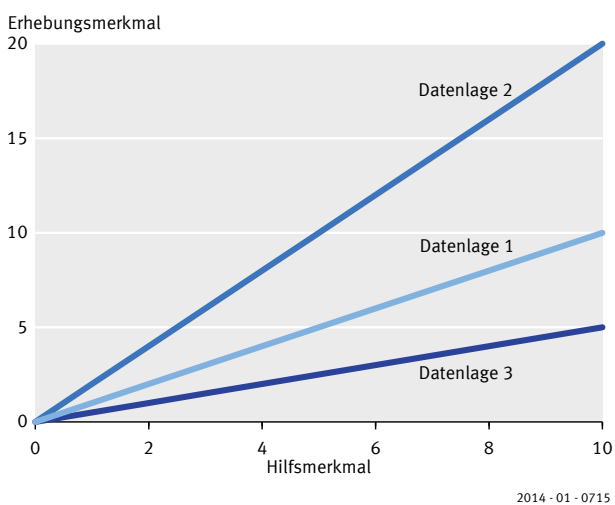
Verhältnisschätzung

Im Rahmen der Verhältnisschätzung wird ein multiplikativer Zusammenhang zwischen Erhebungs- und Bezugsmerkmal unterstellt.

Schaubild 1 veranschaulicht drei Idealfälle, bei denen die Punktwolke des Streudiagramms von x und y jeweils vollständig auf einer Geraden durch den Ursprung liegt. Dabei werden an die Steigung der Geraden keine weiteren Anforderungen gestellt. Bei der Verhältnisschätzung wird ein Korrekturfaktor aus dem Verhältnis des bekannten Kreistotalwertes des Hilfsmerkmals und dem frei hochgerechneten Kreistotalwert des Hilfsmerkmals der Stichprobenbetriebe gebildet. Diese Verhältnisse lassen sich je Kreis und Schicht separat oder über alle Schichten hinweg kombiniert je Kreis berechnen. Daher wird zwischen separater und kombinierter Verhältnisschätzung unterschieden. Da sich die schichtspezifischen Verhältnisswerte von Erhebungs- und Bezugsmerkmalen nicht stark voneinander unterscheiden, wird auf die Darstellung des separaten Schätzverfahren verzichtet. Damit ergibt sich für den Totalwert bei kombinierter Verhältnisschätzung (kVhs) die folgende Schätzfunktion¹¹:

$$(9) \hat{t}_d^{kVhs} = \frac{\sum_{h=1}^{H_L} W_h \sum_{i \in S_h} Y_{hid}}{\underbrace{\sum_{h=1}^{H_L} W_h \sum_{i \in S_h} X_{hid}}_{v_d}} \cdot X_d,$$

Schaubild 1 Darstellung von idealen Datenlagen für eine Verhältnisschätzung (perfekter multiplikativer Zusammenhang)



11 Siehe Krug, W./Nourney, M./Schmidt, J. (Fußnote 6), hier: Seite 184 ff.

wobei

$$x_{hid} = \begin{cases} x_{hi} & \text{für } i \in d \\ 0 & \text{sonst} \end{cases}$$

Der Standardfehler lässt sich näherungsweise abschätzen durch:

$$(10) \hat{s}(\hat{t}_d^{kVhs}) = \sqrt{\sum_{h=1}^{H_L} N_h (w_h - 1) (s_{hx_d}^2 - 2v_d s_{hdx_y} + v_d^2 s_{hy_d}^2)}$$

mit

$$s_{hdx_y} = \frac{1}{n_h - 1} \left(\sum_{i \in S_h} x_{hid} y_{hid} - \frac{1}{n_h} \sum_{i \in S_h} x_{hid} \sum_{i \in S_h} y_{hid} \right)$$

und

$$s_{hx_d}^2 = \frac{1}{n_h - 1} \left(\sum_{i \in S_h} x_{hid}^2 - \frac{1}{n_h} \left(\sum_{i \in S_h} x_{hid} \right)^2 \right)$$

Die Verhältnisschätzung ist im Allgemeinen nicht erwartungstreu. Die Verzerrung (Bias) kann jedoch abgeschätzt werden. Für die Abschätzung der Verzerrung gilt näherungsweise die folgende Formel:

$$(11) B(\hat{t}_d^{kVhs}) = \frac{1}{X_d} \sum_{h=1}^{H_L} N_h \frac{N_h - n_h}{n_h} (v_d s_{hx_d}^2 - s_{hdx_y})$$

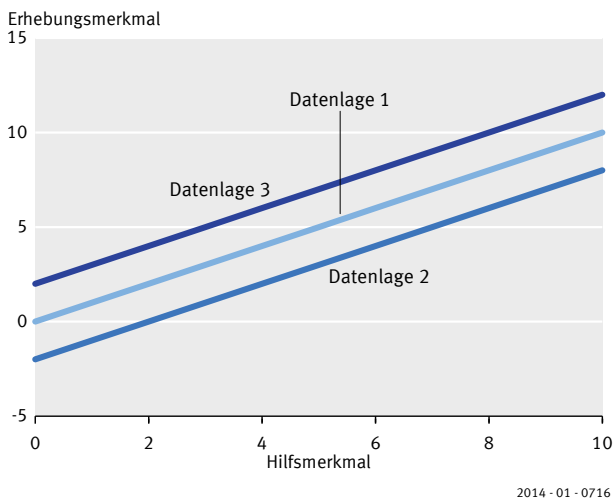
Die Formel zur Abschätzung des Standardfehlers sowie die Formel zur Schätzung der Verzerrung gelten nur näherungsweise. In Kreisen mit wenigen Stichprobenbetrieben sind die üblichen Anwendungsvoraussetzungen für derartige Abschätzungen nicht immer voll erfüllt. Die Ergebnisse der Verhältnisschätzung sind verzerrt, aber die Verzerrung nimmt mit steigendem Stichprobenumfang deutlich ab.

Differenzschätzung

Die Schätzfunktion besteht aus zwei Teilen, zum einen aus dem frei hochgerechneten Teil und zum anderen aus einem Korrekturteil (siehe Formel 12). Dieser entspricht der Differenz zwischen dem bekannten Totalwert X_d und dem frei hochgerechneten Wert. Dies hat zur Folge, dass sinnvolle Bezugsmerkmale spezielleren Anforderungen genügen müssen, als dies bei den beiden anderen Verfahren der Fall ist. Sie müssen von gleicher Art, Dimension und Größenordnung sein. Außerdem wird ausschließlich additiven Zusammenhängen Rechnung getragen. Die Anwendungsvoraussetzung veranschaulicht Schaubild 2.

Zunächst zeigt die Grafik, was unter einem rein additiven Zusammenhang zu verstehen ist. Die eingezeichneten Geraden haben alle die Steigung 1. Die Steigung der Geraden ist anders als bei der Verhältnisschätzung nicht beliebig. Sie kann jedoch nach oben und unten verschoben werden. Eine ideale Datenlage für die Anwendung der Differenzschätzung wäre eine Punktwolke, die sich auf einer Geraden mit Achsenabschnitt nahe dem Ursprung und mit der Steigung 1 konzentriert. Das veranschaulicht noch einmal die besonders hohen Anforderungen an das Hilfsmerkmal als Anwendungsvoraussetzung für die Differenzschätzung.

Schaubild 2 Darstellung von idealen Datenlagen (perfekter additiver Zusammenhang)



2014 - 01 - 0716

Anders als die beiden anderen vorgestellten Ansätze ist der Differenzenschätzer erwartungstreu (unverzerrt). Für die Schätzung des Totalwertes in Kreis d ergibt sich¹²:

$$(12) \hat{t}_d^{diff} = \sum_{h=1}^{H_L} w_h \sum_{i \in S_h} y_{hid} + \underbrace{\left(X_d - \sum_{h=1}^{H_L} w_h \sum_{i \in S_h} x_{hid} \right)}_{\text{Korrekturterm}}$$

Die Formel zur erwartungstreuen Schätzung des Standardfehlers lautet:

$$(13) s(\hat{t}_d^{diff}) = \sqrt{\sum_{h=1}^{H_L} N_h (w_h - 1) (s_{hy_d}^2 - 2s_{hxdy} + s_{hx_d}^2)}$$

Regressionsschätzung

Der Regressionsansatz ist von den hier vorgestellten Ansätzen der umfassendste. Er beruht auf einem linearen Zusammenhang zwischen Hilfs- und Erhebungsmerkmal. Der Ansatz ermöglicht daher sowohl die Berücksichtigung von multiplikativen als auch von additiven Zusammenhängen zwischen Erhebungs- und Bezugsmerkmalen. Der Ansatz kombiniert die Möglichkeiten der Verhältnis- und der Differenzschätzung. Dabei wird das frei hochgerechnete Kreisergebnis mithilfe eines durch einen Regressionskoeffizienten gewichteten Korrekturterms korrigiert (siehe hierzu Formel 14). Das Verfahren stellt eine Verallgemeinerung der Differenzschätzung dar, da die Gewichtung des Korrekturterms von vornherein nicht fest vorgegeben ist. Das Gewicht ist dann besonders hoch, wenn die Kovarianz zwischen Erhebungs- und Bezugsmerkmal besonders hoch ist (siehe Formel 15). Das Schätzergebnis des Regressionsschätzers ist im Allgemeinen verzerrt. Allerdings nimmt die Verzerrung mit steigendem Stichprobenumfang ab. Anders als bei der Verhältnisschätzung lässt sich die Verzerrung in der Regel nicht abschätzen. Auch hier kann – je nachdem, ob der

Regressionskoeffizient je Kreis über alle Schichten hinweg oder je Kreis und Schicht berechnet wird – zwischen kombinierter und separater Regressionsschätzung unterschieden werden. Aufgrund der nahezu einheitlichen Regressionskoeffizienten je Schicht und Kreis wird nur die kombinierte Form dargestellt. Für den Totalwert des Kreises d ergibt sich bei der kombinierten Regressionsschätzung (kReg) die folgende Formel¹³:

$$(14) \hat{t}_d^{kReg} = \sum_{h=1}^{H_L} w_h \sum_{i \in S_h} y_{hid} + \hat{\beta}_d \underbrace{\left(X_d - \sum_{h=1}^{H_L} w_h \sum_{i \in S_h} x_{hid} \right)}_{\text{Korrekturterm}}$$

wobei

$$(15) \hat{\beta}_d = \frac{\sum_{h=1}^{H_L} N_h (w_h - 1) s_{hxdy}}{\sum_{h=1}^{H_L} N_h (w_h - 1) s_{hx_d}^2}$$

Der Standardfehler lässt sich näherungsweise abschätzen durch:

$$(16) s(\hat{t}_d^{kReg}) = \sqrt{(1 - r_{dxy}^2) \sum_{h=1}^{H_L} N_h (w_h - 1) s_{hy_d}^2}$$

wobei

$$(17) r_{dxy} = \frac{\sum_{h=1}^{H_L} N_h (w_h - 1) s_{hxdy}}{\sqrt{\left(\sum_{h=1}^{H_L} N_h (w_h - 1) s_{hx_d}^2 \right) \left(\sum_{h=1}^{H_L} N_h (w_h - 1) s_{hy_d}^2 \right)}}$$

die kreisspezifische empirische Korrelation beschreibt. Formel (16) verdeutlicht, dass der Regressionsschätzer gemessen am Standardfehler immer mindestens die Qualität der freien Hochrechnung erreicht. Bei einer Korrelation von -1 oder 1 sinkt der Standardfehler des Regressionsschätzers auf null. Sobald die empirische Korrelation betragsmäßig größer null ist, ist bezüglich des Standardfehlers der Regressionsschätzer gegenüber der freien Hochrechnung vorteilhaft. Der Standardfehler lässt sich nur näherungsweise abschätzen. Auch hier muss erwähnt werden, dass in kleinen Kreisen mit einer sehr geringen Anzahl an Stichprobenbetrieben die gängigen Anwendungsvoraussetzungen für gute Schätzungen nicht immer voll erfüllt sind.

3 Aufbereitung und Prüfung der Daten

Um die vorgestellten Verfahren anwenden zu können, müssen die Daten gewissen Anforderungen genügen. Die Anwendung der freien Hochrechnung erfordert lediglich die Daten der Viehbestandserhebung Schweine. Für die Anwendung der Verfahren der gebundenen Hochrechnung muss zunächst eine Datei erstellt werden, die einen Datensatz für jeden Betrieb der Auswahlgesamtheit enthält. Außerdem müssen die Datensätze die Merkmalsausprägungen der zu schätzenden Merkmale (hier: Schweine und Zuchtsauen) aus der Auswahlgesamtheit (hier: Zentrales Betriebsregister der Agrarstatistiken) enthalten. Darüber hinaus sollte auch

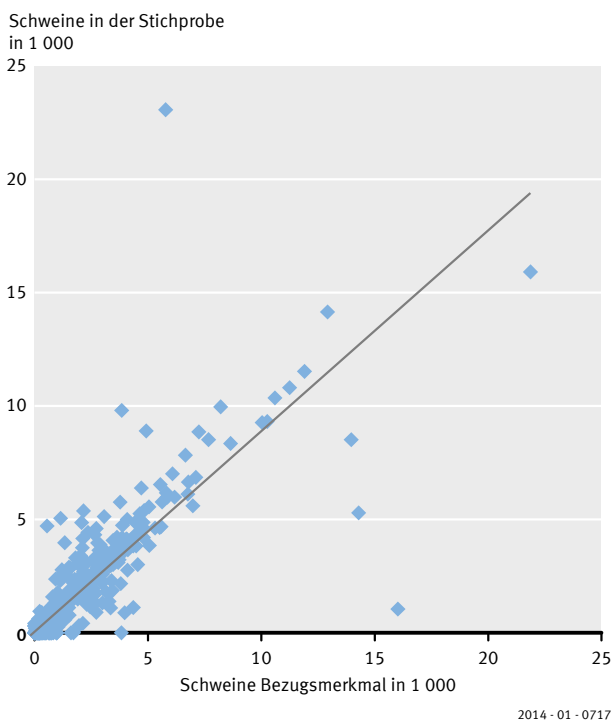
¹² Siehe Krug, W./Nourney, M./Schmidt, J. (Fußnote 6), hier: Seite 194 f.

¹³ Siehe Krug, W./Nourney, M./Schmidt, J. (Fußnote 6), hier: Seite 196 ff.

jeder Betrieb eine Merkmalsausprägung aus der Datenbank HIT als Hilfsmerkmal für die zu schätzenden Merkmale haben, was jedoch zum Zeitpunkt der Untersuchung nicht gegeben war (siehe Tabelle 2). Für die Betriebe, die über keine Ausprägung des Hilfsmerkmals aus der Datenbank HIT verfügen, werden die Registerwerte aus dem Zentralen Betriebsregister der Agrarstatistiken übernommen. Außerdem ist zu prüfen, ob die Erhebungs- und die Hilfsmerkmale in additiver und/oder multiplikativer Beziehung zueinander stehen. Für die Durchführung einer Differenzschätzung müssen die Hilfsmerkmale außerdem von gleicher Dimension, Art und Größenordnung¹⁴ wie die Erhebungsmerkmale sein. Besonders bei hoher (positiver) Korrelation können die gebundenen Hochrechnungsverfahren gewinnbringend eingesetzt werden. Die Beziehung zwischen Erhebungsmerkmal und Bezugsmerkmal wird anhand der Stichprobenbetriebe eines niedersächsischen Landkreises beispielhaft dargestellt.

Das Streudiagramm in Schaubild 3 veranschaulicht die hohe Korrelation zwischen dem Erhebungsmerkmal Viehbestand Schweine und dem Bezugsmerkmal. Außerdem lässt sich erkennen, dass sich die Punktwolke um eine Gerade durch den Ursprung herum konzentriert. Für den als Beispiel genutzten Landkreis beträgt der Wert des Korrelationskoeffizienten (siehe Formel 15) 0,86. Durchschnittlich liegt der Korrelationskoeffizient in den berechneten niedersächsischen Landkreisen bei 0,91. Die Beta-Koeffizienten (siehe Formel 13) liegen im Durchschnitt beim Merkmal Schweine in Niedersachsen bei 0,90 und 0,95 bei den Zuchtsauen. Die einzelnen Beta-Koeffizienten liegen nahe bei 1, wobei

Schaubild 3 Streudiagramm des Bezugs- und Erhebungsmerkmals für einen niedersächsischen Kreis



14 Siehe Krug, W./Nourney, M./Schmidt, J. (Fußnote 6), hier: Seite 194.

es vereinzelt starke Ausreißer gibt. Beim Bezugsmerkmal kann davon ausgegangen werden, dass es von gleicher Art, Dimension und Größenordnung wie das Erhebungsmerkmal ist. Das folgt aus der Lage der Punktwolke und der Dimension der horizontalen und vertikalen Achsen. Sind die gängigen Anforderungen bei Schätzungen ebenfalls erfüllt, so spricht nichts gegen die Anwendung der vorgestellten Verfahren der gebundenen Hochrechnung.

4 Darstellung der Ergebnisse

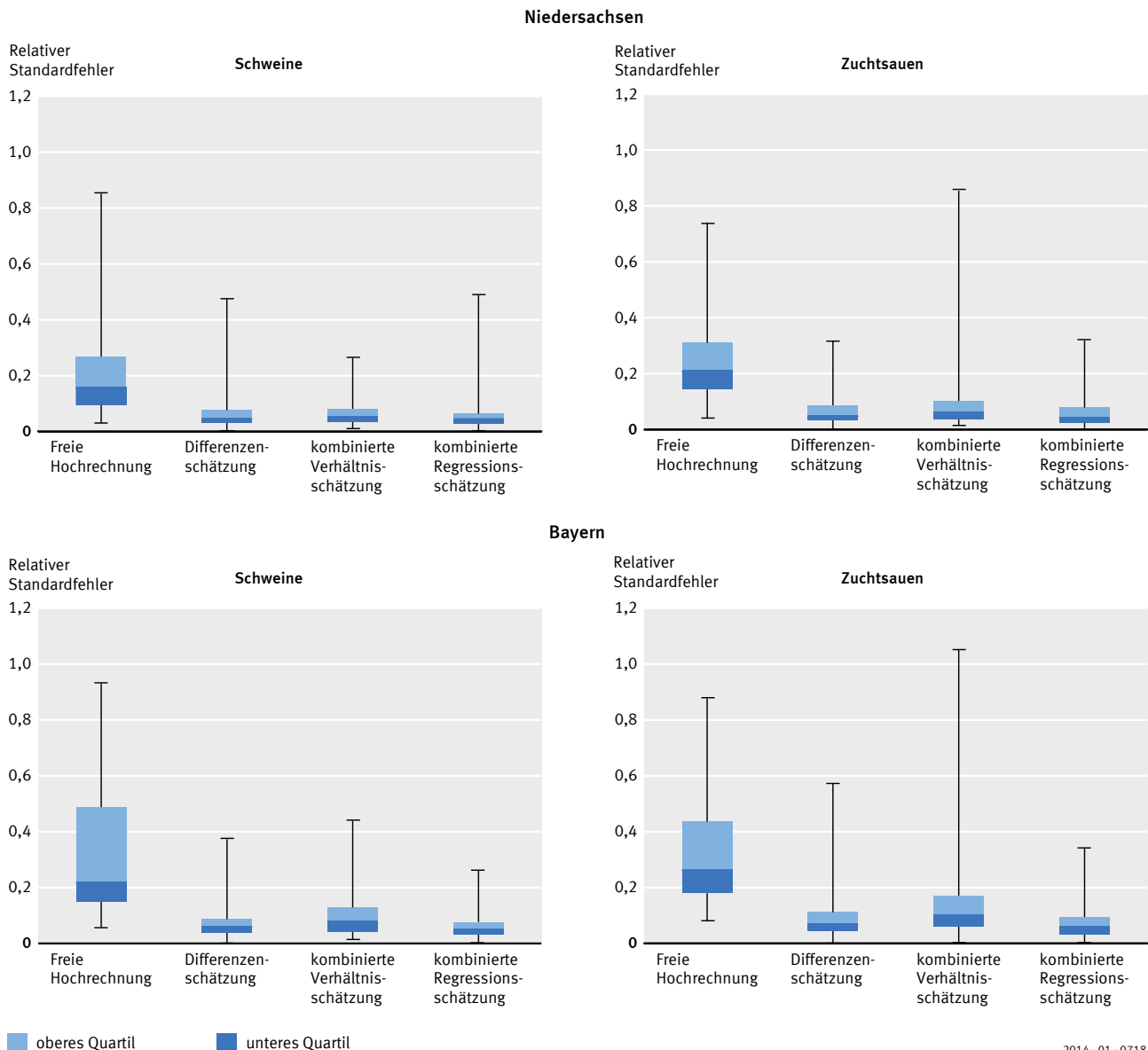
Zunächst werden für ausgewählte niedersächsische Landkreise die Schätzergebnisse für Schweinebestände und die zugehörigen relativen Standardfehler dargestellt. Anhand des relativen Standardfehlers werden die Schätzungen der einzelnen Verfahren miteinander verglichen. Um repräsentative Ergebnisse zu erhalten, wurden sowohl Kreise mit niedrigen relativen Standardfehlern als auch solche mit hohen relativen Standardfehlern ausgewählt. Tabelle 3 zeigt für alle durchgeführten Schätzverfahren die Schätzergebnisse der Totalwerte mit den zugehörigen relativen Standardfehlern. Dabei sind die Kreise aufsteigend nach der Größe des relativen Standardfehlers der freien Hochrechnung sortiert. Bei der kombinierten Verhältnisschätzung wurden die Schätzwerte um die Verzerrung korrigiert.

Der Tabelle 3 lässt sich entnehmen, dass der relative Standardfehler der freien Hochrechnung in den Kreisen mit hohen geschätzten Viehbeständen niedrig ist, wohingegen er in Kreisen mit niedrigen geschätzten Viehbeständen sehr hoch ist. Bei den Ergebnissen der gebundenen Hochrechnungsverfahren gilt diese Aussage nur mit Einschränkungen. Bei den Verfahren der gebundenen Hochrechnung hängt die Höhe des relativen Standardfehlers nicht mehr nur von der Streuung des Erhebungsmerkmals und den Auswahlätzen in den Schichten ab, sondern zusätzlich von der Korrelation zwischen Erhebungsmerkmal und Bezugsmerkmal. So ist es möglich, dass der Kreis J bei der freien Hochrechnung einen relativen Standardfehler von rund 56 % und bei den Verfahren der gebundenen Hochrechnung von teilweise unter 1 % hat. Beim Kreis J handelt es sich um einen Landkreis, der nur eine sehr geringe Anzahl an relevanten Betrieben aufweist. Die Erhebungsmerkmalswerte und die Bezugsmerkmalswerte weisen eine sehr hohe Korrelation auf, wodurch der relative Standardfehler drastisch reduziert wird. Dies ist allerdings ein extremes Beispiel und im Allgemeinen weisen die Ergebnisse weiterhin für Kreise mit hohen geschätzten Viehbeständen einen eher niedrigen relativen Standardfehler auf und umgekehrt. Ein Vergleich der relativen Standardfehler zeigt, dass die Verfahren der gebundenen Hochrechnung in der Tabelle 3 allesamt einen niedrigeren relativen Standardfehler aufweisen als die frei hochgerechneten Schätzergebnisse. Die geringsten relativen Standardfehler zeigen sich bei der kombinierten Regressionsschätzung. Die Differenzschätzung und die kombinierte Verhältnisschätzung weisen in etwa gleich hohe relative Standardfehler auf. Generell sind die Unterschiede aber marginal. Damit lassen sich durch die Anwendung der Verfahren der gebundenen Hochrechnung die Schätzungen für Kreisergebnisse deutlich stabilisieren.

Tabelle 3 Schätzergebnisse für ausgewählte Kreise in Niedersachsen für den Viehbestand Schweine

| | Freie Hochrechnung | | Differenzschätzung | | Kombinierte Verhältnisschätzung | | Kombinierte Regressionsschätzung | |
|-----------------|---------------------------|--------------------------|---------------------------|--------------------------|---------------------------------|--------------------------|----------------------------------|--------------------------|
| | Schweinebestand insgesamt | relativer Standardfehler | Schweinebestand insgesamt | relativer Standardfehler | Schweinebestand insgesamt | relativer Standardfehler | Schweinebestand insgesamt | relativer Standardfehler |
| | Anzahl | % | Anzahl | % | Anzahl | % | Anzahl | % |
| Landkreis A .. | 1 436 996 | 3,05 | 1 422 512 | 1,72 | 1 423 219 | 1,71 | 1 425 497 | 1,57 |
| Landkreis B .. | 1 010 492 | 3,72 | 1 002 748 | 1,56 | 1 002 607 | 1,58 | 1 002 982 | 1,56 |
| Landkreis C .. | 403 534 | 5,54 | 427 050 | 2,07 | 425 733 | 2,06 | 424 477 | 1,99 |
| Landkreis D .. | 96 113 | 11,94 | 98 750 | 4,41 | 99 040 | 5,55 | 98 863 | 4,38 |
| Landkreis E .. | 64 772 | 14,57 | 65 381 | 4,75 | 65 412 | 5,81 | 65 431 | 4,64 |
| Landkreis F ... | 58 151 | 15,32 | 55 687 | 6,24 | 55 823 | 6,11 | 55 878 | 6,10 |
| Landkreis G .. | 40 145 | 19,73 | 38 245 | 6,26 | 38 306 | 6,14 | 38 279 | 6,25 |
| Landkreis H .. | 26 276 | 23,85 | 23 218 | 4,12 | 23 200 | 4,12 | 23 289 | 4,06 |
| Landkreis I ... | 14 158 | 34,54 | 12 195 | 20,50 | 12 710 | 19,87 | 12 489 | 19,10 |
| Landkreis J ... | 4 129 | 55,99 | 5 652 | 0,59 | 5 633 | 1,07 | 5 633 | 0,26 |

Schaubild 4 Boxplots der relativen Standardfehler für Schweine und Zuchtsauen in Niedersachsen und Bayern



2014 - 01 - 0718

Die Verteilung der relativen Standardfehler über die Kreise für Schweine und Zuchtsauen wird für jedes berechnete Verfahren für Niedersachsen und Bayern mithilfe von Boxplots in Schaubild 4 veranschaulicht. Dabei handelt es sich um eine grafische Darstellung von Lage- und Streuungsmaßen wie etwa unteres und oberes Quartil, Median, Quartilsabstand und Spannweite. Hier stellt die untere Begrenzung der Box das untere Quartil der relativen Standardfehler dar, die Linie in der Mitte der Box den Median, sowie die obere Begrenzung der Box das obere Quartil. Die beiden Querstriche oben und unten an den sogenannten Antennen stellen Minimum und Maximum dar.

Die Boxplots bestätigen die in Tabelle 3 dargestellten Ergebnisse: Alle Verfahren der gebundenen Hochrechnung sind gemessen an der Höhe des relativen Standardfehlers deutlich vorteilhafter als die freie Hochrechnung. Das gilt sowohl für Niedersachsen als auch für Bayern und für die Merkmale Schweine und Zuchtsauen. Dem Schaubild 4 lässt sich auch entnehmen, dass der relative Standardfehler des kombinierten Regressionsschätzers in der Regel am niedrigsten ausfällt. Um die Darstellung der Boxplots zu verbessern, wurden die Daten um zwei extreme Ausreißer bereinigt. So gibt es in Bayern etwa einen Kreis, für den die Differenzschätzung einen negativen Totalwert für das Merkmal Zuchtsauen liefert. Das ist möglich, wenn beispielsweise nur eine sehr geringe Anzahl an relevanten Betrieben in dem Kreis liegt und die Differenz aus dem Kreistotalwert des Bezugsmerkmals und dem hochgerechneten Wert des Erhebungswertes in dem Kreis (siehe Formel 16). In einem anderen Fall hat der relative Standardfehler der Differenzschätzung eine extrem hohe Ausprägung, weil der hochgerechnete Wert des Bezugsmerkmals deutlich größer ist als der hochgerechnete Wert des Erhebungsmerkmals. Daraus resultiert ein sehr kleines Schätzergebnis der Differenzschätzung, wodurch der relative Standardfehler entsprechend hoch ausfällt. Dabei handelt es sich ebenfalls um eine Ausnahme und um einen Kreis, der nur über eine sehr geringe Anzahl an relevanten Betrieben verfügt. Tabelle 4 enthält die Mittelwerte der relativen Standardfehler für Bayern und Niedersachsen aufgelistet nach verwendeten Verfahren und geschätzten Merkmalen.

Die Mittelwerte der relativen Standardfehler zeigen, dass die Verfahren der gebundenen Hochrechnung im Mittel einen deutlich niedrigeren Standardfehler aufweisen als die freie Hochrechnung. Der kombinierte Regressionsschätzer hat im Mittel den niedrigsten relativen Standardfehler. Dies deckt sich mit den zuvor dargestellten Ergebnissen. Tendenziell sind die relativen Standardfehler bei bayerischen

Tabelle 4 Mittelwerte der relativen Standardfehler
Prozent

| | Schweine | | Zuchtsauen | |
|---|---------------|--------|---------------|--------|
| | Niedersachsen | Bayern | Niedersachsen | Bayern |
| Freie Hochrechnung | 22,41 | 31,74 | 25,4 | 33,65 |
| Differenzschätzung | 7,03 | 8,18 | 6,43 | 10,75 |
| Kombinierte Verhältnisschätzung | 11,92 | 10,78 | 9,72 | 15,88 |
| Kombinierte Regressions-schätzung | 6,18 | 6,42 | 5,76 | 7,15 |

Kreisen meist etwas höher als die relativen Standardfehler bei niedersächsischen Kreisen. Dies ist auch bei den relativen Standardfehlern der Landesergebnisse der Fall und resultiert im Wesentlichen aus den unterschiedlichen Verteilungen der Viehbestände in Bayern und Niedersachsen. Während sich in Niedersachsen sehr große Betriebe konzentrieren, gibt es in Bayern eine hohe Zahl an mittleren bis kleinen Betrieben.¹⁵ Aufgrund der höheren Auswahlsätze in den Schichten mit großen Betrieben (mit zum Beispiel 5 000 und mehr Schweinen) sind die Bedingungen in Niedersachsen bezüglich der Höhe der Standardfehler grundsätzlich vorteilhafter. Bei der Berechnung der Mittelwerte wurden die zuvor erwähnten beiden Kreise mit extremen Ausprägungen des relativen Standardfehlers nicht berücksichtigt, da sie das Ergebnis verzerren. Neben den Angaben zur Verteilung des relativen Standardfehlers über die Kreise interessiert, wie hoch der Anteil des gesamten Viehbestands ist, der mit soliden relativen Standardfehlern erklärt werden kann. Tabelle 5 liefert hierzu Informationen.

Tabelle 5 Kumulierte Anteile der geschätzten Viehbestände Schweine für Niedersachsen
Prozent

| | Freie Hochrechnung | Differenzschätzung | Kombinierte Verhältnisschätzung | Kombinierte Regressions-schätzung |
|--------------------------|--------------------|--------------------|---------------------------------|-----------------------------------|
| Relativer Standardfehler | | | | |
| unter 5 % ... | 26,93 | 90,50 | 84,17 | 92,66 |
| unter 10 % .. | 73,81 | 99,18 | 97,84 | 99,15 |
| unter 15 % .. | 84,21 | 99,68 | 99,69 | 100 |
| unter 30 % .. | 96,96 | 99,97 | 99,99 | 100 |
| unter 50 % .. | 98,94 | 100 | 99,99 | 100 |
| unter 70 % .. | 100 | 100 | 99,99 | 100 |
| unter 90 % .. | 100 | 100 | 99,99 | 100 |
| unter 225 % .. | 100 | 100 | 100 | 100 |

Die Tabelle verdeutlicht, dass mithilfe der Differenzschätzung oder der kombinierten Regressionsschätzung Kreisergebnisse mit relativen Standardfehlern von maximal 10 % möglich sind, die mehr als 99 % des Viehbestands Schweine in Niedersachsen abdecken. Die freie Hochrechnung deckt mit diesem Niveau nur etwa 74 % des Viehbestands ab.

Die Landes- und Bundesergebnisse, die bereits jetzt halbjährlich veröffentlicht werden, sind frei hochgerechnete Ergebnisse. Die Summe der frei hochgerechneten Kreisergebnisse entspricht dem frei hochgerechneten Landesergebnis. Das gilt allerdings nicht für die mit den Verfahren der gebundenen Hochrechnung ermittelten Kreisergebnisse. Tabelle 6 zeigt beispielhaft die Abweichungen zwischen dem frei hochgerechneten Landesergebnis und den gebunden hochgerechneten Summen der Kreisergebnisse für Niedersachsen.

Tabelle 6 weist die Viehbestände in Tausend mit einer Nachkommastelle nach, da es sich hier um die Darstellungsform bei Veröffentlichungen handelt. Die Abweichungen der

¹⁵ Siehe Fachserie 3 „Land- und Forstwirtschaft, Fischerei“, Reihe 2.1.3 „Viehhaltung der Betriebe“, 2013 (www.destatis.de, im Bereich Publikationen > Thematische Veröffentlichungen).

Tabelle 6 Abweichungen zwischen frei hochgerechnem Landesergebnis und Summe der Kreisergebnisse nach Verfahren und Merkmalen für Niedersachsen

| | Freie Hochrechnung | Differenzschätzung | Kombinierte Verhältnisschätzung | Kombinierte Regressionschätzung |
|------------------|--------------------|--------------------|---------------------------------|---------------------------------|
| | 1 000 | | | |
| Schweine | 8 717,7 | 8 713,5 | 8 707,5 | 8 714,0 |
| Zuchtsauen . . . | 516,8 | 519,9 | 519,7 | 520,7 |
| | Abweichung in % | | | |
| Schweine | - | - 0,048 | - 0,117 | - 0,042 |
| Zuchtsauen . . . | - | + 0,597 | + 0,573 | + 0,767 |

gebunden hochgerechneten Summen der Kreisergebnisse zum frei hochgerechneten Landesergebnis wurden mit den exakten Ergebnissen berechnet. Es zeigt sich bei den Schweinen eine Abweichung nach unten in einer Größenordnung von deutlich unter einem Prozent. Am geringsten ist die Abweichung beim kombinierten Regressionsschätzer. Bei den Zuchtsauen liegt die Abweichung nach oben in einer Größenordnung von ebenfalls unter einem Prozent. Diese Abweichungen müssen insbesondere bei Veröffentlichungen berücksichtigt werden. Bei der Differenzschätzung entspricht die Summe der Kreisergebnisse allerdings dem mit der Differenzschätzung berechneten Landesergebnis.

5 Zusammenfassung und Ausblick

Die steigende Nachfrage nach regionalen Daten, die knappen finanziellen Ressourcen der statistischen Ämter und die hohen Qualitätsanforderungen an die Daten der amtlichen Statistik machen es erforderlich, nach angemessenen Alternativen oder Ergänzungen zu kostenintensiven Vollerhebungen zu forschen. Diese Arbeit hatte zum Ziel, die klassischen Verfahren der gebundenen Hochrechnung als Möglichkeit zur Ermittlung von Kreisergebnissen in der Agrarstatistik im Rahmen des Viehbestands Schweine aufzuzeigen. Je nach Verfahren und Datenlage wurde dargestellt, dass es ein erwartungstreuere Verfahren gibt, das zu deutlich stabileren Kreisschätzungen führt als die freie Hochrechnung.

Gemessen am relativen Standardfehler hat sich der kombinierte Regressionsschätzer als präziseste Schätzmethode für Kreisergebnisse sowohl in Niedersachsen als auch in Bayern erwiesen – allerdings mit nur geringen Unterschieden zur Differenz- und auch zur Verhältnisschätzung. Da die Regressionsschätzung jedoch grundsätzlich verzerrte Schätzwerte liefert, die Abschätzung der Verzerrungen nicht ohne Weiteres möglich ist und auch die Standardfehlerabschätzung nicht problemlos ist, ist die Verwendung der Differenzschätzung in Betracht zu ziehen. Die Bedingungen, mit diesem Verfahren präzisere Ergebnisse zu erzielen als mit der freien Hochrechnung möglich, werden eindeutig erfüllt.

Die nächste agrarstatistische Vollerhebung findet im Rahmen der Allgemeinen Strukturserhebung der landwirtschaft-

lichen Betriebe im Jahr 2016 statt. Mit den Daten der Vollerhebung besteht die Möglichkeit, Mittelwerte, Varianzen und Kovarianzen zu berechnen. Damit entfällt die Notwendigkeit, diese zu schätzen. Außerdem ermöglicht der Datensatz der Vollerhebung, mittels Simulation die Verzerrungen abzuschätzen.

Neben den Verwaltungsdaten aus der Datenbank HIT werden künftig weitere Verwaltungsdatenquellen wie die Tierseuchenkassendatenbanken ausgewertet. Sollten diese Daten eine höhere Korrelation als die HIT-Daten zu den Erhebungsdaten aufweisen, können diese als Hilfsinformationen in den vorgestellten Verfahren verwendet werden.

Stärker modellbasierte Verfahren der Small-Area-Statistik¹⁶ sind aus zwei wesentlichen Gründen problematisch: Zum einen kann ihre Anwendung zu Akzeptanzproblemen bei den Datennutzern führen, weil diese den Entstehungsprozess der Schätzergebnisse nicht mehr nachvollziehen können und die Ergebnisse in der Regel nicht als unverzerrt angesehen werden können. Zum anderen ist das derzeitige Stichprobendesign der Viehbestandserhebung nicht auf die Schätzung von Kreisergebnissen ausgerichtet, wodurch sogenannte Unplanned Areas und Non-sampled-Areas entstehen¹⁷, also solche Kreise, in denen der Stichprobenumfang zufällig ist, oder solche, die mit keiner Stichprobeneinheit in der Stichprobe vertreten sind.

Die vorgestellten Verfahren sind eingängig und vermittelbar. Sie zeigen eine deutliche Verbesserung gegenüber den frei hochgerechneten Kreisergebnissen. Es bleibt festzuhalten, dass nur für Kreise mit Stichprobenbetrieben Schätzungen mit den vorgestellten Verfahren durchgeführt werden können. Aufgrund des beschriebenen Stichprobenziehungsprozesses handelt es sich bei den nicht abgedeckten Kreisen lediglich um solche, die für die zu schätzenden Merkmale von geringer Bedeutung sind. Für einige Bundesländer können im Laufe des nächsten Jahres Kreisergebnisse gebunden hochgerechnet werden. Ein automatisiertes Verfahren, welches für die meisten Länder auch aus Stichprobenerhebungen jährlich Ergebnisse auf Kreisebene bereitstellt, wäre frühestens 2017 realisierbar. [\[1\]](#)

¹⁶ Siehe Münnich, R./Burgard, J.P./Vogt, M. (Fußnote 8), hier: Seite 149 ff.

¹⁷ Siehe Münnich, R./Burgard, J.P./Vogt, M. (Fußnote 8), hier: Seite 158.

Auszug aus Wirtschaft und Statistik

Herausgeber

Statistisches Bundesamt, Wiesbaden

www.destatis.de

Schriftleitung

Dieter Sarreither,
Vizepräsident des Statistischen Bundesamtes

Redaktion: Ellen Römer
Telefon: + 49 (0) 6 11 / 75 23 41

Ihr Kontakt zu uns

www.destatis.de/kontakt

Statistischer Informationsservice

Telefon: + 49 (0) 6 11 / 75 24 05

Abkürzungen

| | | |
|----------|---|---|
| WiSta | = | Wirtschaft und Statistik |
| MD | = | Monatsdurchschnitt |
| VjD | = | Vierteljahresdurchschnitt |
| HjD | = | Halbjahresdurchschnitt |
| JD | = | Jahresdurchschnitt |
| D | = | Durchschnitt (bei nicht addierfähigen Größen) |
| Vj | = | Vierteljahr |
| Hj | = | Halbjahr |
| a. n. g. | = | anderweitig nicht genannt |
| o. a. S. | = | ohne ausgeprägten Schwerpunkt |
| St | = | Stück |
| Mill. | = | Million |
| Mrd. | = | Milliarde |

Zeichenerklärung

| | | |
|--------|---|--|
| p | = | vorläufige Zahl |
| r | = | berichtigte Zahl |
| s | = | geschätzte Zahl |
| – | = | nichts vorhanden |
| 0 | = | weniger als die Hälfte von 1 in der letzten besetzten Stelle, jedoch mehr als nichts |
| . | = | Zahlenwert unbekannt oder geheim zu halten |
| ... | = | Angabe fällt später an |
| X | = | Tabellenfach gesperrt, weil Aussage nicht sinnvoll |
| oder — | = | grundsätzliche Änderung innerhalb einer Reihe, die den zeitlichen Vergleich beeinträchtigt |
| / | = | keine Angaben, da Zahlenwert nicht sicher genug |
| () | = | Aussagewert eingeschränkt, da der Zahlenwert statistisch relativ unsicher ist |

Abweichungen in den Summen ergeben sich durch Runden der Zahlen.