

Small Area Schätzmethoden

Dr. Jan Pablo Burgard

Universität Trier
Fachbereich IV - Volkswirtschaftslehre
Wirtschafts- und Sozialstatistik

19.11.2018

Übersicht

Motivation

Grundbegriffe der Small Area-Statistik

Design-basierte Schätzverfahren

Modell-assistierte Schätzverfahren

Synthetische Schätzverfahren

Modell-basierte Schätzverfahren

Ausblick

Small Area Income and Poverty Estimates (SAIPE)

Das SAIPE Programm

- ▶ Wird vom U.S. Census Bureau durchgeführt
- ▶ Liefert jährliche Statistiken über Schätzer zu Einkommen und Armut für alle US Bundesstaaten, Counties und Schuldistrikte

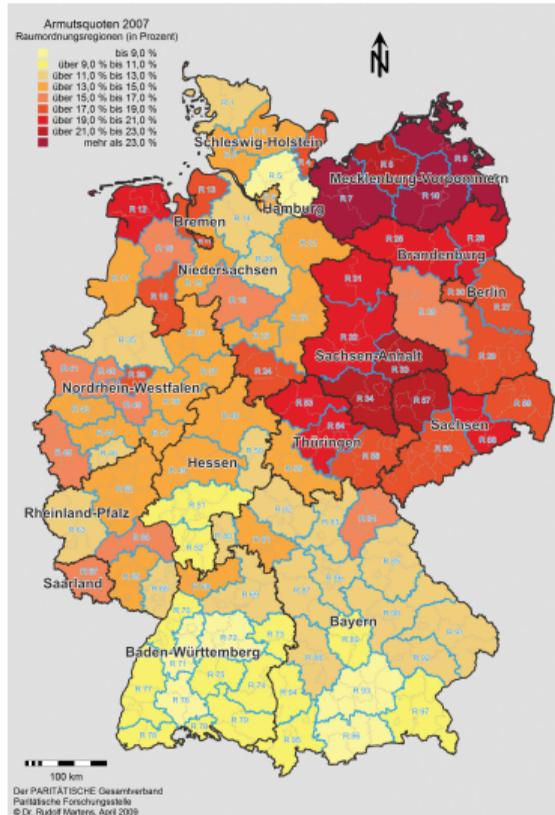
Publizierte Schätzwerte auf Ebene der Staaten und Counties:

- ▶ Absolute Zahl der Personen in Armut
- ▶ Zahl der Kinder unter 18 in Armut
- ▶ Median-Haushaltseinkommen

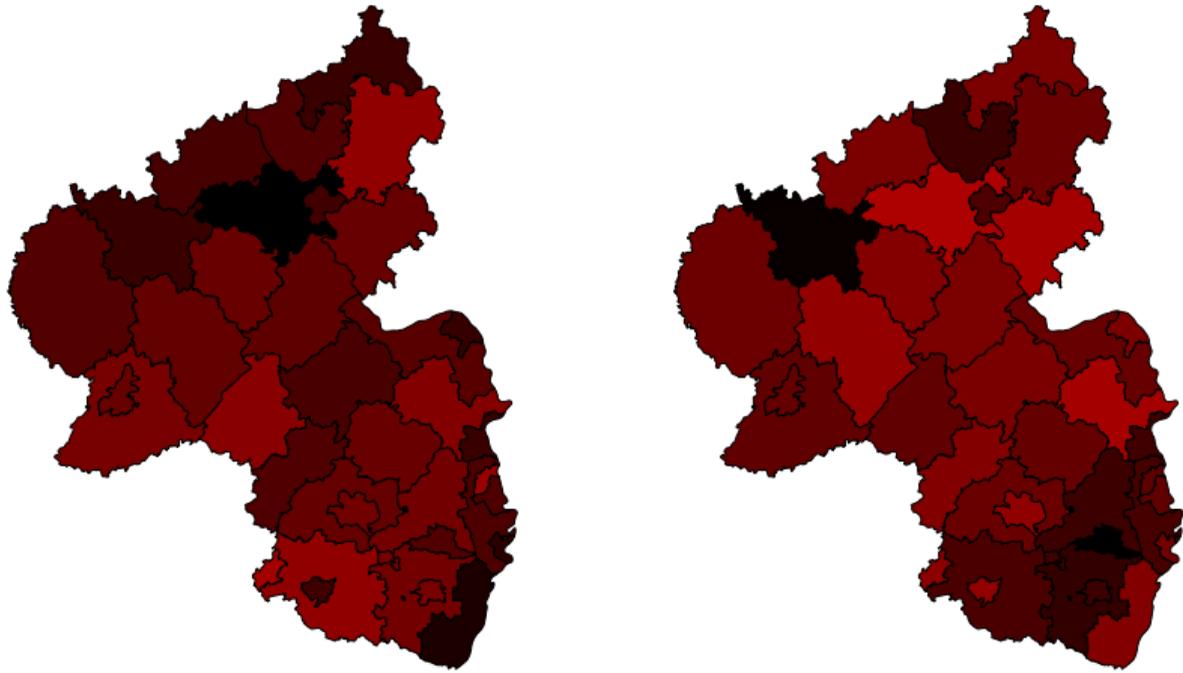
Quelle: <http://www.census.gov/did/www/saipe/about/index.html>

Armutsatlas

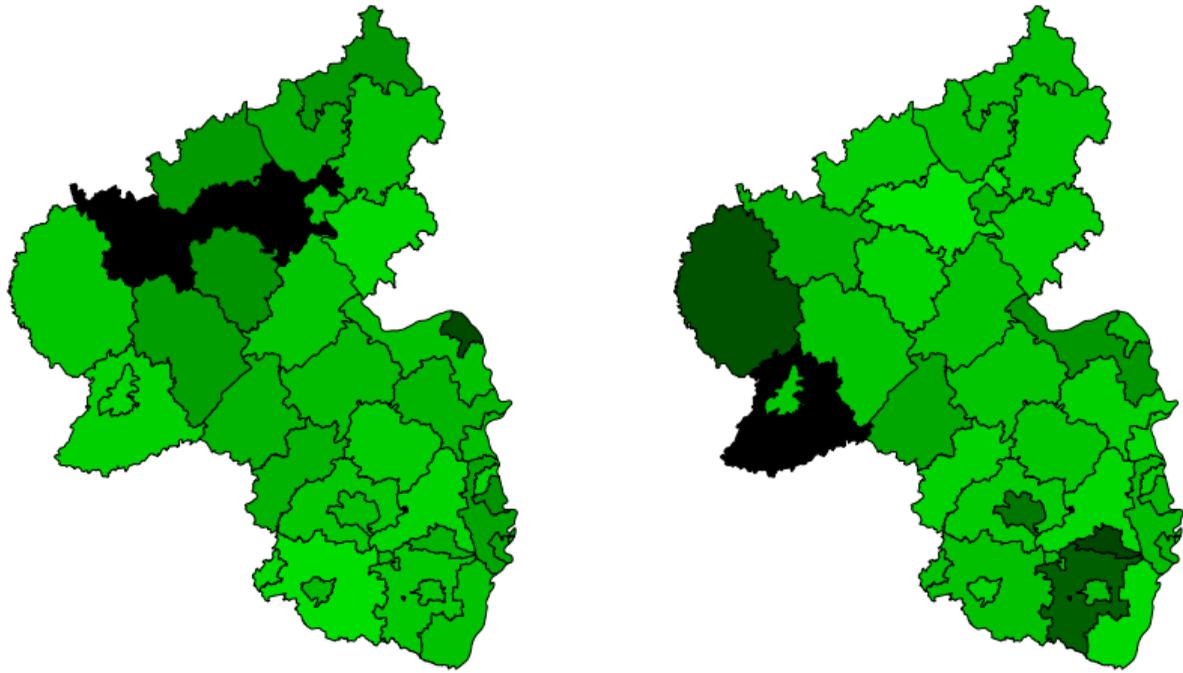
- ▶ Darstellung der At-Risk-of-Poverty Rate für die deutschen Raumordnungsregionen
- ▶ Datenbasis ist der Mikrozensus
- ▶ Keine Verwendung von Small Area Methoden
- ▶ Keine Publikation von detaillierteren Statistiken



Geschätzte Armutsgefährdungsquote bei Verwendung des MZ



Geschätzte QSR (S8020) bei Verwendung des MZ



Häufige Begriffspaare

- ▶ Direkte vs. indirekte Schätzverfahren
- ▶ Areas vs. Domains
- ▶ Small vs. large Areas
- ▶ Sampled vs. non-sampled Domains
- ▶ Planned vs. unplanned Domains
- ▶ Design-basierte vs. Modell-basierte Schätzverfahren
- ▶ Area-Level Modelle vs. Unit-Level-Modelle

Direkte Schätzer

Ein direkter Schätzer liegt vor, wenn nur Daten aus der Area selbst verwendet werden.

z.B. Mittelwertschätzer

Indirekte Schätzer

Auch Informationen außerhalb der Area werden verwendet.

Borrowing Strength => Indirekte Schätzer verwenden entweder implizite oder explizite Modelle.

z.B. National Sample Mean

Was heißt eigentlich klein?

Es gibt keine harte Definition, ab wann eine Area als klein gilt.

Eine Area ist dann als klein zu bezeichnen, wenn der Stichprobenumfang in einer Area nicht für eine ausreichend präzise Schätzung mit Hilfe eines direkten Schätzers ausreicht (vgl. Rao, 2003, S. 1).

Verwendung von zusätzlichen Informationen außerhalb der Area wird zur verlässlichen Schätzung benötigt.

Design-basierte Schätzer

- ▶ Nehmen Bezug auf fixe und finite Population
- ▶ Zufall entsteht aus dem Stichprobendesign
- ▶ Wünschenswerte Eigenschaften: Design-Konsistenz und Design-Unverzerrtheit

Modellbasierte Schätzer

- ▶ Endliche Population ist zufällige Ausprägung einer Superpopulation
- ▶ Schätzungen über Vorhersage der unbeobachteten Werte unterm Modell
- ▶ Einfluss der Modellannahmen

Warum Modell-basierte Schätzungen?

	Design-basierte Methoden GREG / Kalibrierung	Modell-basierte Methoden Synthetische und EBLUP Methoden
Bias	Design-unverzerrt (appr.) auf Grund der Konstruktion	Bias möglich
Varianz	hoch, wenn n_d klein	gering, n_d mit geringerem Einfluss
MSE	$MSE \approx$ Varianz	Modell-Bias evtl. dominant
KI	(weitgehend) unproblematisch	???

Lehtonen und Veijanen (2009)

Unit-Level Modelle

Individuelle Information auf Populationsebene vorhanden für \mathbf{x} und auf Stichprobenebene für (\mathbf{x}, \mathbf{y}) .

Bei linearen Modellen reichen die Area-Randwerte für \mathbf{x} .

Area-Level Modelle

Alle Informationen liegen nur auf Area-Ebene vor.

Sampled Areas

Die Ziehung von Beobachtungen innerhalb jeder interessierenden Area wurde im Stichproben-Design berücksichtigt.

Non-Sampled Areas

Es gibt Areas in denen keine Beobachtungen durch die Stichprobe gemacht wurde.

Planned/Unplanned Domains sind analog zu verstehen.

Horvitz-Thompson-Schätzer

Ein Design-unverzerrter Schätzer des Totalwertes τ_d ist der HT-Schätzer

$$\hat{\tau}_d^{\text{HT}} = \sum_{k \in S_d} \frac{y_k}{\pi_k} = \sum_{k \in S_d} w_k \cdot y_k, \quad d = 1, \dots, D. \quad (1)$$

Unter der Annahme eines Stichprobendesigns, bei dem die Inklusionswahrscheinlichkeiten erster Ordnung positiv sind, ist

$$V(\hat{\tau}_d^{\text{HT}}) = \sum_{k \in U_d} \pi_k (1 - \pi_k) \cdot \left(\frac{y_k}{\pi_k} \right)^2 + 2 \cdot \sum_{\substack{k, l \in U_d \\ k < l}} (\pi_{kl} - \pi_k \cdot \pi_l) \cdot \frac{y_k}{\pi_k} \cdot \frac{y_l}{\pi_l} \quad (2)$$

die Varianz der Schätzfunktion $\hat{\tau}_d^{\text{HT}}$.

Horvitz-Thompson-Schätzer (Fortsetzung)

Diese kann unverzerrt geschätzt werden durch

$$\widehat{V}(\widehat{\tau}_d^{\text{HT}}) = \sum_{k \in S_d} (1 - \pi_k) \cdot \left(\frac{y_k}{\pi_k}\right)^2 + 2 \cdot \sum_{\substack{k, l \in S_d \\ k < l}} \left(1 - \frac{\pi_k \cdot \pi_l}{\pi_{kl}}\right) \cdot \frac{y_k}{\pi_k} \cdot \frac{y_l}{\pi_l} \quad . \quad (3)$$

In der Praxis kann die Bestimmung der Inklusionwahrscheinlichkeiten zweiter Ordnung problematisch sein (etwa bei π ps Designs), weshalb Approximationen benötigt werden.

Für einige bekannte Designs wie die uneingeschränkte oder eine geschichtete Zufallsstichprobe innerhalb der Areas vereinfachen sich die Ausdrücke (2) und (3) enorm.

Generalisierte Regressionsschätzer (GREG)

Der GREG ist durch folgenden Ausdruck gegeben

$$\hat{\tau}_d^{\text{GREG}} = \sum_{k \in U_d} \hat{y}_k + \sum_{k \in S_d} w_k (y_k - \hat{y}_k), \quad (4)$$

mit den \hat{y}_k als den Prädiktionen aus einem zu wählenden Modell. Alternativ kann der zweite Term mit dem Faktor N_d/\hat{N}_d multipliziert werden.

Der GREG besteht aus zwei Komponenten:

- ▶ der Summe der Prädiktionen für alle Einheiten, die zur Domain d gehören, welche um die
- ▶ gewichtete Summe der Residuen korrigiert wird.

Durch diese Korrektur ist der GREG asymptotisch Design-unverzerrt, unabhängig davon ob das Modell korrekt spezifiziert wurde oder nicht.

Synthetische Schätzverfahren

- ▶ Annahme: Areas habe gleiche Charakteristiken wie übergeordnete Ebene.
- ▶ „Gleichen Charakteristiken “ müssen durch ein statistisches Modell beschreibbar sein.
- ▶ z.B.: Proportionale Zusammenhänge oder identische konditionale Verteilungen
- ▶ Es sind implizite wie explizite Modellierungen möglich
- ▶ Erlauben Schätzwerte für Areas ohne Stichprobenumfang

Regressionssynthetische Schätzung

- ▶ Angenommen, Hilfsinformationen \mathbf{x}_k seien für die in die Stichprobe gekommenen Elemente verfügbar
- ▶ Überdies, seien Area-spezifische Totalwerte jener Hilfsinformation $\tau_{X,d}$ verfügbar (bspw. aus dem Zensus / Register)
- ▶ Betrachte regressionssynthetischen Schätzer

$$\hat{\tau}_d^{\text{SYN}} = \sum_{k \in U_d} \hat{y}_k = \tau'_{X,d} \hat{\beta} \quad (5)$$

- ▶ $\hat{\beta}$ wird aus kombinierter Regression der y_k auf \mathbf{x}_k ermittelt
- ▶ Der Schätzer funktioniert gut, sofern kombiniertes Regressionsmodell adäquat ist

MSE-Schätzung I

$$\begin{aligned}\text{MSE}(\widehat{\tau}_{d,\text{syn}}) &= \mathbf{E} \left[(\widehat{\tau}_{d,\text{syn}} - \tau_d)^2 \right] \\ &= \mathbf{E} \left[(\widehat{\tau}_{d,\text{syn}} - \widehat{\tau}_{d,\text{dir}} + \widehat{\tau}_{d,\text{dir}} - \tau_d)^2 \right] \\ &= \mathbf{E} \left[(\widehat{\tau}_{d,\text{syn}} - \widehat{\tau}_{d,\text{dir}})^2 \right] + \mathbf{E} \left[(\widehat{\tau}_{d,\text{dir}} - \tau_d)^2 \right] \\ &\quad + 2 \mathbf{E} \left[(\widehat{\tau}_{d,\text{syn}} - \widehat{\tau}_{d,\text{dir}}) (\widehat{\tau}_{d,\text{dir}} - \tau_d) \right]\end{aligned}$$

mit

$$\mathbf{E} \left[(\widehat{\tau}_{d,\text{syn}} - \widehat{\tau}_{d,\text{dir}}) (\widehat{\tau}_{d,\text{dir}} - \tau_d) \right] \approx -\mathbf{V}(\widehat{\tau}_{d,\text{dir}})$$

MSE-Schätzung II

ergibt sich

$$\text{MSE}(\hat{\tau}_{d,\text{syn}}) \approx \text{E} \left[(\hat{\tau}_{d,\text{syn}} - \hat{\tau}_{d,\text{dir}})^2 \right] - \text{V}(\hat{\tau}_{d,\text{dir}}) \quad (6)$$

$\text{MSE}(\hat{\tau}_{d,\text{syn}})$ kann durch

$$\widehat{\text{MSE}}(\hat{\tau}_{d,\text{syn}}) = (\hat{\tau}_{d,\text{syn}} - \hat{\tau}_{d,\text{dir}})^2 - \widehat{\text{V}}(\hat{\tau}_{d,\text{dir}}) \quad (7)$$

geschätzt werden. Allerdings ist diese Schätzung instabil und kann Werte kleiner 0 liefern.

MSE-Schätzung III

Alternativen

- ▶ Schätzung des mittleren MSEs über alle Areas. Diese Schätzung ist allerdings nicht mehr area-spezifisch
- ▶ Verwendung der Identität $\text{MSE}(\hat{\tau}_{d,\text{syn}}) = V(\hat{\tau}_{d,\text{syn}}) + B^2(\hat{\tau}_{d,\text{syn}})$
Unter der Annahme dass $B^2(\hat{\tau}_{d,\text{syn}}) \approx \frac{1}{D} \sum_d B^2(\hat{\tau}_{d,\text{syn}})$ gilt, folgt

$$\text{MSE}(\hat{\tau}_{d,\text{syn}}) \approx V(\hat{\tau}_{d,\text{syn}}) + \frac{1}{D} \sum_d B^2(\hat{\tau}_{d,\text{syn}})$$

Vorteile Modell-basierter Verfahren

- ▶ Wie bei synthetischen Schätzern ist die Varianz nicht so sehr vom Stichprobenumfang abhängig
- ▶ Area-spezifische Präzision kann für die Schätzwerte angegeben werden
Achtung: Präzisionsmessung beruht auf Annahme der Gültigkeit des Modells

Bedingungen

Modell-basierte Methoden setzen die Verfügbarkeit guter Hilfsvariablen voraus:

- ▶ Modelldiagnostiken benötigt zur Bestimmung eines geeigneten Modells
- ▶ Datenebene: Unit Level-Modelle versus Area Level-Modelle

Relevanz des Modellbaus

- ▶ **Überragende Bedeutung bei Modell-basierten Small Area-Verfahren**
 - ▶ Eigenschaften der Schätzer **nur** unter Gültigkeit der Modelle
→ *given the model is true*
 - ▶ Schätzer funktionieren genau dann gut, wenn das Modell die Daten gut beschreibt und die Annahmen erfüllt sind
- ▶ Bei Modell-assistierten Verfahren ist Modellbau ebenfalls wichtig, allerdings robuster gegen Fehlspezifikationen

Modelle für Small Area-Schätzungen

- ▶ Es werden meistens Zwei-Ebenen Modelle eingesetzt. Ein einfacher Fall ist

$$y_{dj} = \mathbf{x}_{dj}^T \boldsymbol{\beta} + v_d + e_{dj}, \quad (8)$$

mit \mathbf{x}_{dj} als dem Zeilenvektor der Kovariablen, $\boldsymbol{\beta}$ als dem Zeilenvektor der Regressionskoeffizienten, v_d als dem Area-Effekt für Area $d = 1, \dots, D$ und e_{dj} als dem individuellen Störterm.

- ▶ In der Regel $v_d \stackrel{iid}{\sim} N(0, \sigma_v^2)$, $e_{dj} \stackrel{iid}{\sim} N(0, \sigma_e^2)$ und $v_d \perp e_{dj}$
- ▶ v_d wird dann als *Random Effect* bezeichnet.
- ▶ Damit werden zufällige Einflüsse auf zwei Ebenen modelliert:
 - Ebene 1** Die individuellen Störgrößen e_{dj} und
 - Ebene 2** Den area-spezifischen Random Effect v_d .

Prädiktion der \mathbf{v}

Empirical Best Linear Unbiased Predictor (EBLUP)

Man sucht den Prädiktor $\hat{\mathbf{v}}$ für \mathbf{v} mit der kleinsten Varianz gegeben:

1. die Beobachtungen \mathbf{y} gehen linear ein
2. $E(\hat{\mathbf{v}}) = E(\mathbf{v})$.

Achtung: Die Unverzerrtheitsbedingung bezieht sich nur auf den Erwartungswert der \mathbf{v} , da diese eine Zufallsvariable beschreiben.

Empirical Best/Bayes Predictor (EBP)

Ähnliches Konzept, dass aber auch auf nicht-linearen Modellen Anwendung finden kann.

Unit-Level Modell von Battese et al., 1988 I

Es wird das folgende Modell betrachtet:

$$y_{dj} = \mathbf{x}_{dj}^T \boldsymbol{\beta} + v_d + e_{dj}, \quad d = 1, \dots, D, j = 1, \dots, N_d, \quad (9)$$

$$v_d \stackrel{iid}{\sim} (0, \sigma_v^2)$$

$$e_{dj} \stackrel{iid}{\sim} (0, \sigma_e^2)$$

mit $\mathbf{x}_{dj} = (x_{dj1}, \dots, x_{djp})^T$ als $p \times 1$ Spaltenvektor der Hilfsvariablen für die j -te Einheit in der d -ten Area. Ferner wird angenommen, dass die Area-spezifischen Zufallseffekte v_d unabhängig vom Sampling-Fehler e_{dj} sind.

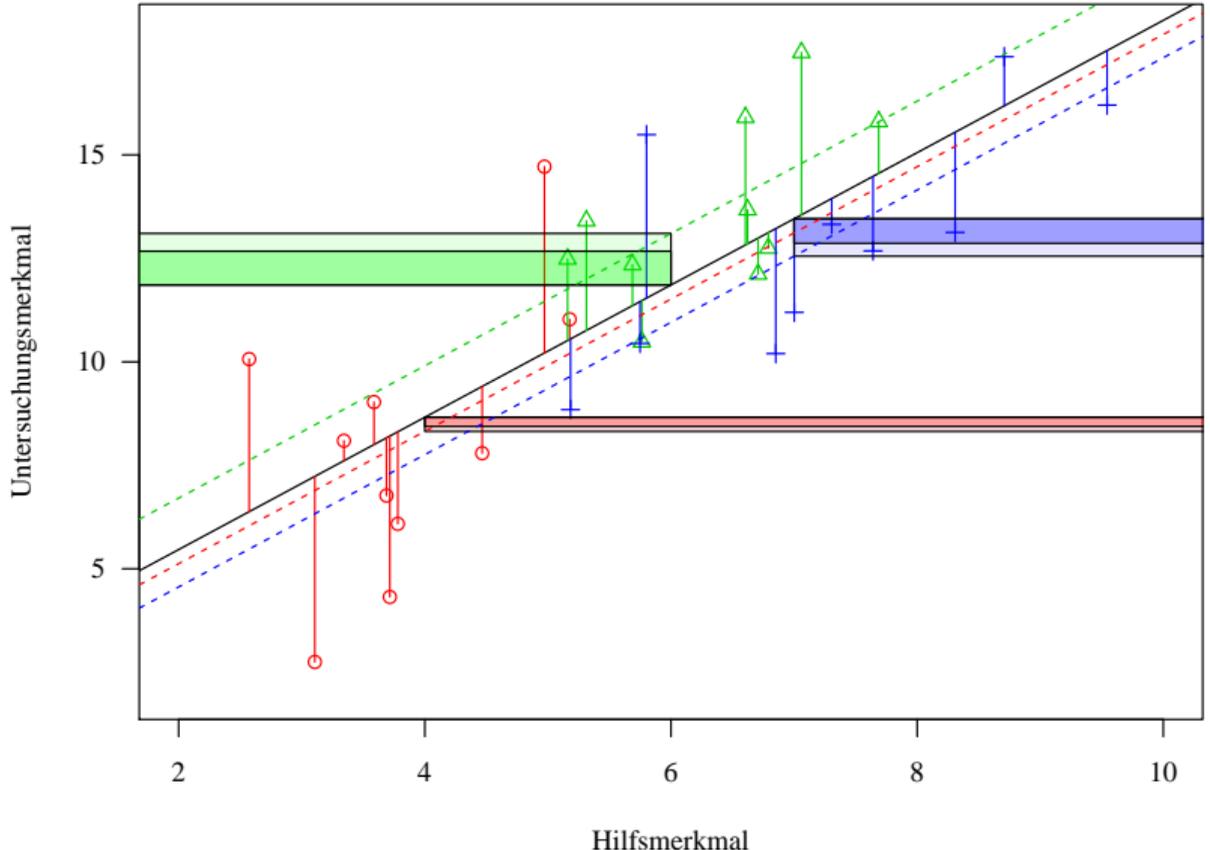
- ▶ Man setzt i.d.R. $x_{dj1} \equiv 1 \forall j, d$.
- ▶ Random-Intercept Modell bei Unit-Level Informationen
- ▶ Für $\sigma_v^2 = 0$ ergibt sich das klassische lineare Regressionsmodell

Unit-Level Modell von Battese et al., 1988 II

Unter der Annahme, dass das Modell (9) auch für die Stichprobe gilt und die Auswahlanteile n_d/N_d klein sind, kann der Mittelwert durch $\mu_d \approx \bar{\mathbf{X}}_d^T \boldsymbol{\beta} + v_d$ für alle Areas $d = 1, \dots, D$ approximiert werden (vgl. Battese et al., 1988).

$$\begin{aligned}\hat{\mu}_d^{\text{BHF}} &= \bar{\mathbf{X}}_d^T \hat{\boldsymbol{\beta}} + \hat{v}_d \\ \hat{v}_d &= \hat{\gamma}_d \left(\bar{y}_d - \bar{\mathbf{x}}_d^T \hat{\boldsymbol{\beta}} \right) \\ \hat{\gamma}_d &= \frac{\hat{\sigma}_v^2}{\hat{\sigma}_v^2 + \hat{\sigma}_e^2/n_d}.\end{aligned}\tag{10}$$

In (10) ist $\hat{\boldsymbol{\beta}}$ fixe Effekte eines gemischten linearen Modells, $\bar{\mathbf{X}}_d$ ist der $p \times 1$ -Vektor der bekannten Mittelwerte der Hilfsvariablen in den Areas und $\hat{\gamma}_d$ gibt den *Shrinkage-Faktor* an.



Anmerkungen

- ▶ Wenn $\hat{\sigma}_v^2$ und/oder n_d klein $\rightarrow \hat{\gamma}_d$ klein \rightarrow mehr Gewicht auf indirekte Informationen
- ▶ Wenn $\hat{\sigma}_v^2$ und/oder n_d groß $\rightarrow \hat{\gamma}_d$ groß \rightarrow mehr Gewicht auf direkte Informationen
- ▶ Der BHF-Schätzer ist nicht Design-konsistent für allgemeine Stichprobendesigns / berücksichtigt keine Design-Gewichte
- ▶ Der EBLUP ist Modell-unverzerrt im Sinne von $E(\hat{\mu}_d^{BHF}) = E(\mu_d)$, aber verzerrt bedingt der Realisation von v_d (vgl. Pfeffermann 2013).

Endliche Populationen

Bei endlichen Populationen mit höheren Auswahlsätzen ist Gleichungen (10) anzupassen

$$\hat{\mu}_d^{BHF} = \frac{1}{N_d} \left[\sum_{j \in S_d} y_{dj} + \sum_{j \notin S_d} \hat{y}_{dj} \right] \quad \text{mit} \quad (11)$$

$$\hat{y}_{dj} = \mathbf{x}_{dj}^T \hat{\boldsymbol{\beta}} + \hat{v}_d \quad (12)$$

Gleichung (11) ist auch die allgemeine Darstellung für einen Empirical Best Predictor. Dieser setzt sich allgemein aus zwei Komponenten zusammen

1. der Summe der beobachteten Werte in einer Area sowie
2. der Summe der geschätzten Werte für die unbeobachteten Werte einer Area.

Area Level-Modelle

Unit Level-Modelle leider nicht immer anwendbar.

- ▶ keinen Zugriff auf Unit Level-Informationen;
- ▶ geeignete Hilfsinformation liegen nicht auf Mikroebene vor oder
- ▶ Unit Level-Modelle zu rechenaufwändig.

In diesen Fällen können Area Level-Modelle Abhilfe leisten.

Darstellung des Area Level-Modell nach Jiang und Lahiri (2006)

$$\begin{aligned}\widehat{\mu}_d^{Dir} &= \bar{\mathbf{X}}_d^T \boldsymbol{\beta} + v_d + e_d, \quad d = 1, \dots, D, \\ v_d &\stackrel{iid}{\sim} (0, \sigma_v^2) \\ e_d &\stackrel{[ind]}{\sim} (0, \psi_d).\end{aligned}\tag{13}$$

Area Level-Modelle (Forts.)

Anders als im Unit Level-Modell (9) werden im Area-Level-Modell (13) nicht die Beobachtungen selbst, sondern die Mittelwerte der direkten Schätzer modelliert.

Alternativ lässt sich das Modell auch wie folgt darstellen

$$\begin{aligned}\mu_d^{Dir} | \mu_d &\stackrel{ind}{\sim} (\mu_d, \psi_d), & d = 1, \dots, D, & \text{Sampling-Modell} \\ \mu_d &\stackrel{ind}{\sim} (\bar{\mathbf{X}}_d^T \boldsymbol{\beta}, \sigma_v^2), & d = 1, \dots, D, & \text{Linking-Modell}\end{aligned}$$

Dabei bezeichnet man σ_v^2 auch als Modellvarianz und ψ_d als Stichprobenvarianz des direkten Schätzers. Diese wird als bekannt vorausgesetzt, hängt von den Teilstichprobenumfängen n_d ab und unterscheidet sich daher zwischen den Areas.

EBLUP

Der EBLUP unter dem Modell (13) ist gegeben durch (vgl. Fay und Herriot, 1979):

$$\hat{\mu}_d^{FH} = \bar{\mathbf{X}}_d^T \hat{\beta} + \hat{v}_d \quad (14)$$

$$\hat{\beta} = \left(\sum_{d=1}^D \frac{\bar{\mathbf{X}}_d \bar{\mathbf{X}}_d^T}{(\psi_d + \hat{\sigma}_v^2)} \right)^{-1} \left(\sum_{d=1}^D \frac{\bar{\mathbf{X}}_d \hat{\mu}_d^{Dir}}{(\psi_d + \hat{\sigma}_v^2)} \right)$$

$$\hat{v}_d = \frac{\hat{\sigma}_v^2}{(\psi_d + \hat{\sigma}_v^2)} (\hat{\mu}_d^{Dir} - \bar{\mathbf{X}}_d^T \hat{\beta})$$

Durch Umformung ergibt sich

$$\hat{\mu}_d^{FH} = \hat{\gamma}_d \hat{\mu}_d^{Dir} + (1 - \hat{\gamma}_d) \bar{\mathbf{X}}_d^T \hat{\beta} \quad \text{mit} \quad (15)$$

MSE-Schätzung

- ▶ Keine geschlossene Lösung für den MSE, jedoch Approximationen möglich
- ▶ Generell können die MSE-Schätzungen der EBLUPs wie folgt zerlegt werden:

$$\widehat{\text{MSE}}(\widehat{\mu}_d) \approx g_{1d}(\widehat{\boldsymbol{\theta}}) + g_{2d}(\widehat{\boldsymbol{\theta}}) + 2g_{3d}(\widehat{\boldsymbol{\theta}})$$

wobei $\widehat{\boldsymbol{\theta}}$ den Vektor der Varianzkomponenten darstellt.

- ▶ Im FH-Modell: $\widehat{\boldsymbol{\theta}} = \widehat{\sigma}_v^2$
- ▶ Im BHF-Modell: $\widehat{\boldsymbol{\theta}} = (\widehat{\sigma}_v^2, \widehat{\sigma}_e^2)^T$

Interpretation der Komponenten

- ▶ Die erste Komponente $g_{1d}(\hat{\theta})$ geht aus der Vorhersage der Random-Effects \hat{v}_d hervor.
- ▶ Die zweite Komponente $g_{2d}(\hat{\theta})$ berücksichtigt die Variabilität resultierend aus der Schätzung von β .
- ▶ Die dritte Komponente $g_{3d}(\hat{\theta})$ reflektiert die Unsicherheit durch die Schätzung der Varianz-Komponenten θ .
- ▶ Der dominante Term $g_{1d}(\hat{\theta})$ ist von höchster Ordnung.

MSE-Schätzung

FH-Modell

Es kann gezeigt werden, dass

$$g_{1d}(\hat{\sigma}_v^2) = \hat{\gamma}_d \psi_d$$

$$g_{2d}(\hat{\sigma}_v^2) = (1 - \hat{\gamma}_d)^2 \bar{\mathbf{X}}_d^T \left[\sum_{d=1}^D \frac{\bar{\mathbf{X}}_d \bar{\mathbf{X}}_d^T}{(\psi_d + \hat{\sigma}_v^2)} \right]^{-1} \bar{\mathbf{X}}_d. \quad (17)$$

Der Term $g_{3d}(\hat{\sigma}_v^2)$ hängt von der verwendeten Schätzmethode ab.

MSE-Schätzung

BHF-Modell

Es kann gezeigt werden, dass

$$g_{1d}(\hat{\sigma}_v^2, \hat{\sigma}_e^2) = \hat{\gamma}_d \hat{\sigma}_e^2 / n_d$$

$$g_{2d}(\hat{\sigma}_v^2, \hat{\sigma}_e^2) = (\bar{\mathbf{X}}_d - \hat{\gamma}_d \bar{\mathbf{x}}_d)^T \left(\sum_{d=1}^D \mathbf{X}_d^T \mathbf{V}_d^{-1} \mathbf{X}_d \right)^{-1} (\bar{\mathbf{X}}_d - \hat{\gamma}_d \bar{\mathbf{x}}_d). \quad (18)$$

Der Term $g_{3d}(\hat{\sigma}_v^2, \hat{\sigma}_e^2)$ hängt von der verwendeten Schätzmethode ab.

Ausblick

- ▶ Forschung an neuen Small Area Schätzmethoden sehr dynamisches
- ▶ Aktuelle Forschungsarbeiten
 - ▶ EBP unter Berücksichtigung von Fehlenden Werten in den abhängigen Variablen
 - ▶ EBP unter bekannten und unbekanntem Fehlern in den Kovariablen
 - ▶ EBP unter Mischverteilungsannahmen
 - ▶ EBP mit Designkonsistenz
- ▶ Bisher in Deutschland wenige empirische Anwendungen => Viel Potential!

Ausblick

- ▶ Forschung an neuen Small Area Schätzmethoden sehr dynamisches
- ▶ Aktuelle Forschungsarbeiten
 - ▶ EBP unter Berücksichtigung von Fehlenden Werten in den abhängigen Variablen
 - ▶ EBP unter bekannten und unbekanntem Fehlern in den Kovariablen
 - ▶ EBP unter Mischverteilungsannahmen
 - ▶ EBP mit Designkonsistenz
- ▶ Bisher in Deutschland wenige empirische Anwendungen => Viel Potential!

Vielen Dank für Ihre Aufmerksamkeit!

Wichtige Literatur I

- ▶ Battese, G. E., Harter, R. M. und Fuller, W. A. (1988): An Error-Components Model for Prediction of County Crop Areas Using Survey and Satellite Data. In: Journal of the American Statistical Association, 83 (401), S. 28-36.
- ▶ Fay, R. E. und Herriot, R. A. (1979): Estimates of Income for Small Places: An application of James Stein Procedures to Census Data. In: Journal of the American Statistical Association, 74 (366), S. 269-277.
- ▶ Jiang, J. und Lahiri, P. (2006): Mixed Model Prediction and Small Area Estimation (with Discussion). In: Test, 15 (1); S. 1-96.
- ▶ Lehtonen, R. und Veijanen, A. (2009): Design-based Methods of Estimation for Domains and Small Areas. Kapitel 31 in Pfeiffermann, D. und Rao, C. R. (Hrsg.): Handbook of Statistics, Vol. 29B, New York: Elsevier.

Wichtige Literatur II

- ▶ Münnich, R.; Burgard, J. P. und Vogt, M. (2013): Small Area-Statistik: Methoden und Anwendungen. In: AStA Wirtschafts- und Sozialstatistisches Archiv, Nr. 3/4, S. 149 – 191.
- ▶ Pfeffermann, D. (2013): New important developments in small area estimation. In: Statistical Science, 28, S. 40-68.
- ▶ Prasad, N. G. N. und Rao, J. N. K. (1990): The Estimation of the Mean Squared Error of Small-Area Estimators. In: Journal of the American Statistical Association, 85, 409, S. 163 – 171.
- ▶ Rao, J.N.K. (2003): Small Area Estimation. New York: Wiley. *Es gibt eine neue Auflage!*