

# Das Stichproben- design des registergestützten Zensus 2011

# The Sample Design for the Register-Assisted Census 2011

*Ralf Münnich, Siegfried Gabler, Matthias Ganninger,  
Jan Pablo Burgard und Jan-Philipp Kolb*

## *Zusammenfassung*

Im Rahmen der europaweiten Zensus-Erhebungsrunde im Jahr 2011 wird zum ersten Mal seit 1987 auch im vereinigten Deutschland wieder eine Volkszählung stattfinden, diesmal allerdings nicht in Form einer Vollerhebung, sondern in Form einer kosten- und ressourcenschonenden registergestützten Erhebung. Diese wird flankiert durch eine Haushaltsstichprobe, aus der erstens in den Registern nicht erfasste Informationen gewonnen werden sollen und zweitens eine Abschätzung der Zahl der Karteileichen (KAL) und Fehlbestände (FEB) in den Melderegistern erfolgen soll. Aus den Register- und Stichprobendaten sollen möglichst verlässliche und genaue Schätzungen der Totalwerte vorgenommen werden. Ziel des von DESTATIS eingesetzten Stichprobenforschungsprojektes ist es, Antworten auf die Frage zu geben, welches Stichprobendesign unter den gegebenen Restriktionen empfohlen werden kann. Darüber hinaus sollen Schätzstrategien entwickelt werden, die zur Verwendung im Zensus 2011 vorgeschlagen werden können. Der vorliegende Aufsatz stellt einige wichtige Erkenntnisse aus dem Forschungsprojekt dar, wobei ein Schwerpunkt auf der Darstellung eines optimalen Stichprobendesigns liegt.

## *Abstract*

Within the context of the Europe-wide census elicitation in 2011 there will be the first population census in reunified Germany. In contrast to the last German census in 1987, where all households were interviewed, the new census will be conducted by means of a cost- and resource-effective register-assisted census. In addition to the register information, a household sample will be drawn. On the one hand this sample will provide information that is not included in the register, on the other hand it will allow for the estimation of over- and undercounts in the register. Reliable estimates for total values of interest are to be derived from the register and sample data. The aim of the research project, which was initiated by DESTATIS, is to elaborate an efficient sample design as well as to develop estimation strategies which allow accurate estimates for the census 2011. This article presents some important findings from the research project. However, one focus is on the description of an optimal sample design.

## 1 Ausgangssituation

Im Rahmen der europaweiten Zensus-Erhebungsrunde im Jahr 2011 wird zum ersten Mal seit 1987 auch im vereinigten Deutschland wieder eine Volkszählung stattfinden. Diese wird jedoch nicht in Form einer Vollerhebung erfolgen, sondern wird in Form einer kosten- und ressourcenschonenden registergestützten Erhebung vollzogen. Zudem wird eine Haushaltsstichprobe gezogen, aus der zum einen eine Abschätzung der Zahl der Karteileichen (KAL) und Fehlbestände (FEB) in den Melderegistern erfolgen soll und zum anderen in den Registern nicht erfasste Informationen gewonnen werden.

Ziel ist es insbesondere, die amtliche Einwohnerzahl zu ermitteln, aber auch eine Schätzung von Totalwerten zusätzlicher in den Registern nicht enthaltener Merkmale zu erhalten. Zur Erarbeitung eines geeigneten Stichprobendesigns hat das Statistische Bundesamt (DESTATIS) ein Forschungsprojekt in Auftrag gegeben, das von einem Konsortium unter der Leitung von Ralf Münnich (Uni Trier) in Zusammenarbeit mit GESIS bearbeitet wird. Das Forschungsprojekt soll vor allem Antworten auf die Frage geben, welches Stichprobendesign unter den gegebenen Restriktionen empfohlen werden kann. Daneben sollen Schätzstrategien entwickelt werden, die zur Verwendung im Zensus 2011 vorgeschlagen werden können.

Um diese Ziele zu erreichen, wurde eine geeignete Simulationsumgebung erarbeitet, mit deren Hilfe der Beantwortung der Forschungsfragen nachgegangen wird. Ebenso wichtig ist es, eine geeignete Darstellungsform für die Visualisierung und Evaluation der hochdimensionalen Ergebnisse zu finden.

Der vorliegende Aufsatz gibt einige wichtige Erkenntnisse aus dem Forschungsprojekt wieder, wobei ein Schwerpunkt auf der Darstellung eines optimalen Stichprobendesigns liegt.

## 2 Methodische und statistische Grundlagen

Mit Hilfe der Daten aus der 2011 zu realisierenden Stichprobe des Zensus sollen nach Maßgabe von DESTATIS zwei Ziele erfüllt werden

*Ziel 1: Die Ermittlung der amtlichen Einwohnerzahl*

*Ziel 2: Die Schätzung von Kennzahlen bei Zusatzvariablen*

Um diese Ziele zu erreichen, werden geeignete Stichprobendesigns untersucht. Von Seiten des Auftraggebers (DESTATIS) existieren hierzu einige Vorgaben, die in jedem Fall eingehalten werden müssen. So werden im Zensus 2011 komplette Anschriften

aus dem Anschriften- und Gebäuderegister (AGR) gezogen, dessen Aufbau in Kleber et al. (2009) beschrieben ist. In einer ausgewählten Anschrift wird dann die Anzahl an Personen ermittelt, die eine bestimmte interessierende Eigenschaft aufweisen. Bei dieser Eigenschaft handelt es sich im Fall von Ziel 1 um die Existenz einer Person in der Anschrift (*Ziel 1 Variable*) und bei Ziel 2 um die konkrete Ausprägung einer interessierenden Variablen (*Ziel 2 Variable*). Im AGR ist die Anschriftengröße, also die Anzahl der an der Anschrift registrierten Personen, enthalten. Es liegt nahe, diese Information durch eine Schichtung schon in das Auswahlverfahren einfließen zu lassen, da für etliche interessierende Merkmale, insbesondere Anzahl der Karteileichen und Fehlbestände, ein Zusammenhang mit der Anschriftengröße vermutet wird. Hierzu wird die Variable Anschriftengröße in Klassen eingeteilt und die so entstandene Anschriftengrößenklasse (ADK) als Schichtungsvariable verwendet. Allem Folgenden liegen die in Tabelle 1 dargestellten drei Schichtungsvarianten zugrunde, wobei SMP für *Sampling Point* steht (zur Definition der Sampling Points siehe Abschnitt 2.2).

Tabelle 1 Schichtungsvarianten

Bezeichner	Beschreibung
ADK1	6 Schichten mit Schichtgrenzen: 1, 2, 3, 4-6, 7-10, 11+ registrierte Personen in der Anschrift
ADK2	4 mit registrierten Personen gleich stark besetzte Schichten pro SMP
ADK3	8 mit registrierten Personen gleich stark besetzte Schichten pro SMP

Die Definitionen der Schichtgrenzen der Varianten ADK2 und ADK3 lassen sich auch verstehen als eine Bestimmung der Quartile (ADK2) beziehungsweise der Oktile (ADK3) in Bezug auf die Variable Anschriftengröße.

Neben der Festlegung der Schichten ist eine wesentliche Aufgabe des Forschungsprojekts, eine geeignete Aufteilung des Stichprobenumfangs auf die Schichten zu finden. Wünschenswert ist die Einhaltung von Präzisionsanforderungen an die Schätzungen (in unterschiedlichen Graden je nach Ziel und Untergliederungstiefe; siehe Aufstellung in Abschnitt 2.3) sowie das Nicht-Überschreiten eines vorab festgelegten Gesamtstichprobenumfangs bezogen auf in Deutschland registrierte Personen.

## 2.1 Simulationsumgebung

Bei der Entwicklung eines geeigneten Stichprobendesigns und der Bewertung von Schätzverfahren wird auf unterschiedliche Datenbestände zurückgegriffen. Dem

zeitlichen Vorliegen dieser Daten folgt die chronologische Einteilung des Projektes in drei Phasen. In der Phase 0 liegen rein synthetische Daten zugrunde, die sich aber an der Realität orientieren. Diese synthetische Grundgesamtheit wurde im Rahmen des DACSEIS-Projektes auf Basis des Vorgehens von Devroye (1986) erzeugt.<sup>1</sup> In Phase 1 wurden zum einen ausgewählte anonymisierte Melderegisterdaten aus vier Bundesländern um synthetisch generierte Variablen angereichert sowie erste gesamtdeutsche aggregierte Melderegisterdaten zur Planung herangezogen. In Phase 2 stehen die anonymisierten Melderegisterdaten für Gesamtdeutschland (im Folgenden MR-Daten) zur Verfügung.

Die Simulationsgesamtheiten, die der ersten und zweiten Phase zugrunde liegen, setzen sich aus verschiedenen Teilen zusammen.<sup>2</sup> Da die MR-Daten als deterministischer Block angesehen werden können, dienen die 85.790.381 Einträge als Rahmen der Simulation. Es werden also keine neuen Personen hinzu simuliert, sondern lediglich weitere Variablen für diesen schon vorhandenen Block synthetisch generiert. Zudem werden einzelne Personen verschiedenen Modellen zufolge als nicht im Melderegister erfasste (Fehlbestände) oder fälschlicherweise im Melderegister erfasste Personen (Karteileichen) ausgewiesen (Münnich et al. 2009; Burgard 2009; Burgard/Münnich 2010). Die Grundlage dieser Karteileichen- und Fehlbestandsmodelle liefert ein anonymisierter Datensatz, der aus dem Zensusstest 2001 resultiert.<sup>3</sup>

Als dritte wichtige Datenquelle zur Erzeugung der synthetischen Variablen dient der anonymisierte Mikrozensus aus dem Jahr 2006. Ziel der synthetischen Datengenerierung in den Phasen 1 und 2 war es, heterogenere Strukturen über die administrativen Einheiten zu erzeugen, als dies noch für die Phase 0 der Fall war.

Bei der synthetischen Datengenerierung werden nun aufbauend auf bereits vorhandenen Variablen weitere Variablen generiert, welche auch in Gruppen (blockweise) erzeugt werden können. Zunächst wird hier der Block der Registervariablen verwendet. Die Erzeugung erfolgt nun rekursiv auf Basis von Kreuztabellen oder Modellen. Die Kreuztabellen werden im Mikrozensus auf Kreis-Ebene berechnet, da tiefer gegliederte Ebenen bei der Auszählung kaum noch valide Ergebnisse liefern. Darüber hinaus müssen die Hochrechnungsfaktoren aus dem Mikrozensus 2006 (Variable EF951) bei der Auszählung berücksichtigt werden. Damit wird eine Kon-

1 Zum genauen Vorgehen zur Erzeugung des Phase 0 Datensatzes siehe Münnich und Schürle (2003).

2 Die beiden Simulationsgesamtheiten unterscheiden sich hinsichtlich des Umfangs (Phase 1 – anonymisierte Melderegisterauszüge aus vier Bundesländern, Phase 2 – anonymisierte Melderegisterdaten aus allen Bundesländern) und weniger hinsichtlich der verwendeten Methodik. Deshalb wird im Folgenden nur von einer Simulationsgesamtheit die Rede sein.

3 Für nähere Informationen zum Zensusstest 2001 siehe Schäfer 2004.

sistenz zwischen den hochgerechneten Stichprobenergebnissen aus dem Mikrozensus und der synthetischen Population erzwungen. Wichtig ist hierbei, dass die Verteilungen der Variablen, die aus den MR-Daten resultieren, sich auch in der Kreuztabelle wiederfinden lassen.

Allgemein kann die gemeinsame Verteilung einer  $n$ -variaten Verteilung rekursiv über ihre bedingten Verteilungen erzeugt werden. Es gilt:

$$F(x_1, \dots, x_n) = F(x_1) \cdot F(x_2 | x_1) \cdots F(x_n | x_1, \dots, x_{n-1})$$

Speziell kann man hieraus für die Erzeugung eines neuen Blockes  $x_{k+1}, \dots, x_n$  bei bereits gegebenen oder erzeugten Variablen  $x_1, \dots, x_k$  die Verteilungsfunktion

$$F(x_{k+1}, \dots, x_n | x_1, \dots, x_k) = \frac{F(x_1, \dots, x_n)}{F(x_1, \dots, x_k)}$$

verwenden. Durch die blockweise Generierung der Daten werden innerhalb der Blöcke die beobachteten Strukturen erhalten, was beispielsweise bei Ausbildungsvariablen sinnvoll ist. Die Reihenfolge der Erzeugung der Variablen spielt damit keine Rolle, solange auf alle bereits erzeugten Variablen konditioniert wird.

Ein weiteres Problem bei der Datenerzeugung ist die Unterscheidung zwischen Stichprobennullen und strukturellen Nullen in den Kreuztabellen. Bei den Erstgenannten handelt es sich um vorhandene, aber durch die MZ-Stichprobe nicht beobachtete Ausprägungen. Strukturelle Nullstellen sind Ausprägungen, die tatsächlich nicht vorkommen. Insbesondere bei der Kombination von Alter, Bildungsvariablen sowie Erwerbsvariablen ergeben sich strukturelle Nullen. Verwendet man bei der Erzeugung synthetischer Daten Modelle oder benutzt nicht alle bereits vorhandenen Variablen für die Konditionierung, kann es sein, dass strukturelle Nullen nicht korrekt erkannt werden. In solchen Fällen muss nach der Erzeugung noch ein Editing durchgeführt werden.

## 2.2 Datenstruktur und Erhebungseinheiten

Die nachfolgend dargestellte hierarchische Struktur von Zusammenfassungen regionaler Einheiten ist derart gestaltet, dass die einzelnen Ebenen der Präzisionsanforderungen widerspruchsfrei und eindeutig berücksichtigt werden. Diese Struktur dient als Basis, um eine regionale Aufteilung des Gesamtstichprobenumfangs zu ermöglichen. Eine anschließende Schichtung zur Erhöhung der Präzision der Schätzungen bleibt davon unberührt.

Eine *Stichprobenbasiseinheit*<sup>4</sup> ist als regionale Einheit definiert, aus der eine Stichprobe gezogen wird, wobei der Stichprobenumfang noch festzulegen ist. Die Einteilung der SMPs in vier Typen erfolgt nach einem streng hierarchischen Schlüssel, bei dem ein Typ höherer Ordnung sich nur noch auf den Rest bezieht, der von dem Typ niedriger Ordnung verblieben ist. Sind daher in einer Verbandsgemeinde eine oder mehrere Gemeinden mit mehr als 10.000 Einwohnern (EW) vorhanden, so sind diese großen Gemeinden Typ 1 zugehörig. Die größten deutschen Städte werden in Stadtteile unterteilt, diese Stadtteile sind dem Typ 0 zugehörig. Die restlichen kleinen Gemeinden der Verbandsgemeinde werden dem Typ 2 zugeordnet, wenn sie zusammen mehr als 10.000 Einwohner haben. Ansonsten gehören sie zum Typ 3.

Tabelle 2 Verteilung der SMP-Typen in den Bundesländern

Bundesland	SMP-Typ (absolute Häufigkeiten)				Summe
	SDT	GEM	VBG	KRS	
Baden-Württemberg	2	244	126	35	407
Bayern	8	216	30	71	325
Berlin	12	0	0	0	12
Brandenburg	0	71	5	14	90
Bremen	3	1	0	0	4
Hamburg	7	0	0	0	7
Hessen	3	168	0	21	192
Mecklenburg-Vorpommern	0	24	30	12	66
Niedersachsen	2	205	68	34	309
Nordrhein-Westfalen	12	339	0	17	368
Rheinland-Pfalz	0	46	122	20	188
Saarland	0	40	0	5	45
Sachsen	4	69	13	22	108
Sachsen-Anhalt	0	60	27	11	98
Schleswig-Holstein	0	53	52	11	116
Thüringen	0	33	6	17	56
Deutschland	53	1569	479	287	2391

4 Mit der Wahl dieser Bezeichnung soll vermieden werden, dass diese mit den so genannten *Small Areas* beziehungsweise den Schichten, die innerhalb dieser Einheiten gebildet werden, direkt in Verbindung gebracht werden.

**Typ 0 (SDT):** Stadtteile ab 200.000 EW aus Gemeinden mit mindestens 400.000 EW

**Typ 1 (GEM):** Gemeinden mit mindestens 10.000 EW, sofern sie nicht zum Typ 0 gehören

**Typ 2 (VBG):** Kleine Gemeinden (unter 10.000 EW) innerhalb eines Gemeindeverbands beziehungsweise einer Verbandsgemeinde werden zusammengefasst, sofern sie in der Summe mindestens 10.000 EW betragen

**Typ 3 (KRS):** Zusammenfassung aller Gemeinden eines Kreises, die bis dahin noch keinem Typ zugeordnet wurden

Insgesamt lassen sich auf diese Weise 2.391 SMPs bilden<sup>5</sup>, die sich auf die Bundesländer wie in Tabelle 2 dargestellt verteilen.

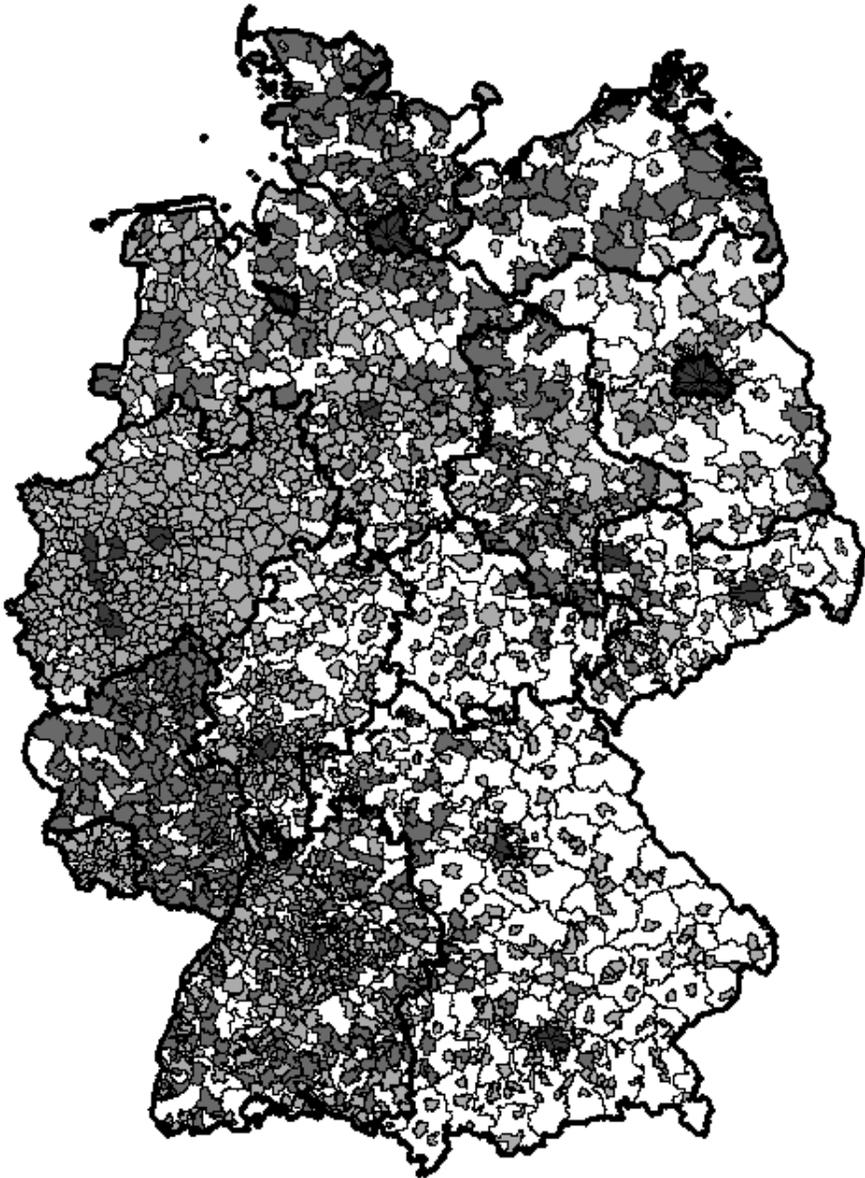
In Abbildung 1 ist eine Darstellung der geografischen Verteilung der SMPs zu finden.<sup>6</sup> Die Einfärbung in Abbildung 1 orientiert sich dabei an der Klassifikation der SMPs (GEM: dunkelstes Grau; SDT: dunkles Grau; KRS: helles Grau; VBG: hellstes Grau); diese sind durch feine schwarze Linien abgegrenzt.

Es sei noch erwähnt, dass von den in Tabelle 1 beschriebenen Schichtgrenzen jene von ADK1 in allen SMPs gleichermaßen definiert sind, wohingegen die Schichtgrenzen von ADK2 und ADK3 von der Verteilung der Personen innerhalb einer SMP abhängen. Da in jeder SMP eine Stichprobe gezogen wird und alle Anschriften des Bundesgebietes vollständig auf die SMPs aufgeteilt sind, muss auch die SMP als Schichtungsvariable betrachtet werden. Durch die Kreuzung von SMP und der Anschriftengrößenklassen ergeben sich schließlich insgesamt 14.346 SMPxADK1, 9.564 SMPxADK2 beziehungsweise 19.128 SMPxADK3 Schichten.

5 Es gilt zu beachten, dass diese Zahl auf den Melderegisterdaten der Phase 1 beruhen, die endgültige Zahl der SMPs zum Zensus-Stichtag kann hiervon noch geringfügig abweichen.

6 Die Abbildung stammt aus einer Forschungsarbeit des Forschungszentrums für Regional- und Umweltstatistik zum Thema *Wirkung der Verwendung von Verbandsgemeinden beim Zensus 2011 in Rheinland-Pfalz*. Die Darstellung wurde mit Hilfe des Datensatzes VG250 vom 31.12.2006 des Geodatenzentrums erstellt und kann kleinere Abweichungen zum aktuellen Registerauszug aufweisen, welche jedoch für die Aussagen keine Auswirkungen haben. Den verwendeten Karten liegen die Vektordaten in den Verwaltungsgrenzen 1:250.000 des Bundesamtes für Kartographie und Geodäsie zugrunde.

Abbildung 1 Darstellung der Stichprobenbasiseinheiten in Deutschland



## 2.3 Präzisionsanforderungen und Schätzer

Zentrales Ziel der Forschungsarbeiten ist es, Vorschläge für Stichprobendesign und Schätzer zu entwickeln, welche gegebene Präzisionsanforderungen erfüllen. Im Folgenden werden nun zunächst diese Präzisionsanforderungen genauer dargestellt. In einem weiteren Abschnitt wird schließlich auf das Referenzschätzverfahren sowie weitere verwendete Schätzer näher eingegangen. Neben dem Horvitz-Thompson (HT) Schätzer und dem modellunterstützten verallgemeinerten Regressionsschätzer (GREG) werden synthetische Schätzer und empirisch beste lineare unverzerrte Prädiktoren (EBLUP) verwendet.

### Präzisionsanforderungen

Im Vordergrund des Forschungsprojekts steht die Frage, wie die folgenden drei zentralen Anforderungen an ein Auswahlverfahren erfüllt werden können

1. Einhaltung der gestellten Präzisionsanforderungen
2. Eingrenzung der Variation der Designgewichte
3. Berücksichtigung eines maximalen Stichprobenumfangs

Tatsächlich lassen sich diese Anforderungen nicht alle gleichzeitig erfüllen, da es sich um konkurrierende Ziele handelt. Vor der Beantwortung obiger Frage stehen Überlegungen zu einer geeigneten regionalen Gliederung von Gemeinden, Verbandsgemeinden und Stadtteilen in die oben beschriebenen Stichprobenbasiseinheiten (SMP) (im Folgenden indiziert durch  $\langle \text{area} \rangle$ ), aus denen die Teilstichproben von Adressen mit zu bestimmenden Umfängen in den Schichten gezogen werden. Zentrales Bewertungskriterium der konkurrierenden Stichprobendesigns ist der relative root mean square error (RRMSE). Es wird also die Präzision der Schätzung von interessierenden Variablen für verschiedene Schätzverfahren und Stichprobendesigns verglichen. Im Rahmen des Forschungsprojekts wurden für die zwei oben genannten Ziele Präzisionsanforderungen formuliert, welche eingehalten werden müssen. Dabei bezeichnet  $\hat{\tau}_{Z, \langle \text{area} \rangle}$  den mit den Stichprobendaten geschätzten Totalwert der in  $\text{SMP}_{\langle \text{area} \rangle}$  lebenden Personen und  $\hat{\tau}_{Y, \langle \text{area} \rangle}$  den mit den Stichprobendaten geschätzten Totalwert einer Untersuchungsvariable  $Y$  der in  $\text{SMP}_{\langle \text{area} \rangle}$  tatsächlich lebenden Personen. Es gelten insbesondere die folgenden Anforderungen bei<sup>7</sup>

7 Siehe hierzu auch: Gesetz zur Anordnung des Zensus 2011 sowie zur Änderung von Statistikgesetzen, §7.

Ziel 1: Schätzungen ausschließlich in Gemeinden ab 10.000 EW

- Stadtteile von Großstädten (SMP-Typ 0):  
RRMSE( $\hat{\tau}_{Z,<area>}$ )  $\leq 0,5\%$
- Gemeinden ab 10.000 EW (SMP-Typ 1):  
RRMSE( $\hat{\tau}_{Z,<area>}$ )  $\leq 0,5\%$

Ziel 2: Betrachtet wird bei

$$\frac{\tau_{Y,<area>}}{\tau_{Z,<area>}} \approx p \quad (\text{mit } p \geq 1/15) \text{ in } \%$$

Es gilt jeweils  $\text{RRMSE}(\hat{\tau}_{Y,<area>}) \leq \frac{1}{p}$  in % für:

- Stadtteile von Großstädten (SMP-Typ 0)
- Gemeinden ab 10.000 EW (SMP-Typ 1)
- VBG in Rheinland-Pfalz (SMP-Typ 2)
- Kreise (SMP-Typ 3)

Bei der Ermittlung der amtlichen Einwohnerzahl (Ziel 1) interessiert ausschließlich die Präzision der Schätzer in so genannten *großen Gemeinden*, also Gemeinden und Stadtteilen von Großstädten mit 10.000 und mehr Einwohnern. Wird die Präzision der Schätzung von Ziel 2 Variablen betrachtet, so unterscheiden sich die Präzisionsanforderungen danach, wie hoch der Anteil  $p$  der interessierenden Merkmalsausprägung ist. Bei Merkmalsausprägungen von einem Anteil unter  $p = 1/15 \approx 6,67\%$  wird keine explizite Anforderung an die Präzision der Schätzer gestellt. Liegt der Anteil einer Ziel 2 Variablen über  $p = 1/15$ , so muss der RRMSE kleiner sein als  $1/p$  in % und zwar bezogen auf die Schätzung in den vier oben genannten Gebiets-typen. Bei der Schätzung einer Ziel 2 Variablen, bei der eine Merkmalsausprägung beispielsweise mit 25 % von Interesse ist, müsste der RRMSE also kleiner oder gleich 4 % sein, wird eine Merkmalsausprägung von 50 % betrachtet, so muss der RRMSE kleiner gleich 2 % sein.

In den bisherigen Untersuchungen hat sich gezeigt, dass die Präzisionsanforderungen für Ziel 1 von vielen Auswahlverfahren und Schätzern eingehalten werden können. Als komplexer zu beurteilen hat sich die Einhaltung der für Ziel 2 formulierten Präzisionsanforderungen erwiesen, da hier sowohl die Anteile als auch deren Verteilung auf die Schichten zu erheblich unterschiedlichen Ergebnissen führen können. In Abschnitt 4 wird an einem Beispiel verdeutlicht, wie unterschiedlich dann auch die Schätzergebnisse beurteilt werden müssen.

## Schätzer

Im Rahmen des Forschungsprojekts werden acht Schätzer für den Totalwert interessierender Variablen untersucht. Eine ausführlichere Diskussion dieser Schätzer findet sich in Münnich et al. (2007). Allgemein bezeichnet  $d_k$  das Designgewicht, also die Inverse der Inklusionswahrscheinlichkeit für Einheit  $k$ . Bei dem verwendeten geschichteten Auswahlverfahren ist die Inklusionswahrscheinlichkeit des  $i$ -ten Elements in Schicht  $h$  gegeben durch

$$\pi_k = \pi_{h,i} = \frac{n_{A,h}}{N_{A,h}} ,$$

wobei  $n_{A,h}$  die Anzahl der zu ziehenden und  $N_{A,h}$  die Gesamtzahl der Anschriften in Schicht  $h$  bezeichnet. Allgemein verzichten wir im Folgenden auf das Subskript  $A$  und verwenden es nur, wenn eine explizite Abgrenzung der Anzahl der Anschriften von der Anzahl der Personen (Subskript  $P$ ) notwendig ist. Somit ist das Designgewicht des  $i$ -ten Elements in Schicht  $h$  gegeben durch

$$w_k = w_{h,i} = \frac{1}{\pi_{h,k}} = \frac{N_h}{n_h} .$$

Die untersuchten Schätzer lassen sich unterteilen in *designbasierte* sowie *modellbasierte* Schätzer. Aus der Klasse der designbasierten Schätzer werden der HT Schätzer und drei Varianten des verallgemeinerten Regressionsschätzers (GREG-Schätzer) betrachtet.

Es bezeichne  $d$  die SMP,  $y_i$  die Ausprägung der  $i$ -ten Einheit auf der Untersuchungs- und  $x_i$  die Ausprägung der  $i$ -ten Einheit auf der Hilfsvariable. Als Hilfsvariable wird in der Regel die Anschriftengröße aus dem AGR verwendet. Die Qualität der Schätzer hängt natürlich von den verfügbaren Hilfsmerkmalen und deren Korrelationen mit der Untersuchungsvariable ab (siehe Abbildung 2).

Darüber hinaus ist  $\tau_{x,d}$  der bekannte Totalwert der Hilfsvariable in der SMP  $d$  und  $\hat{\tau}_{x,d}$  der aus der Stichprobe  $s$  geschätzte Totalwert der Hilfsvariable in der SMP  $d$ . Außerdem ist  $\hat{\tau}_{y,d}$  der aus der Stichprobe  $s$  geschätzte Totalwert der Untersuchungsvariable. Damit ist der HT- und der GREG-Schätzer für den Totalwert gegeben durch

$$\begin{aligned}
 \text{HT Schätzer} \quad & \hat{\tau}_Y^{\text{HT}} = \sum_{i \in S} w_k \cdot Y_k \\
 \text{GREG (Small Area, SA)} \quad & \hat{\tau}_{Y,d}^{\text{GREG,SA}} = \hat{\tau}_{Y,d}^{\text{HT}} + (\tau_{X,d} - \hat{\tau}_{X,d})' \hat{\beta}, \\
 \text{mit} \quad & \hat{\beta} = \left( \sum_{i \in S} w_i x_i x_i^T \right)^{-1} \sum_{i \in S} w_i x_i y_i.
 \end{aligned}$$

Bei den modellbasierten Schätzern lassen sich bei Small Area-Schätzern synthetische und zusammengesetzte Schätzer unterscheiden. Letztere sind in unserem Fall empirisch beste lineare unverzerrte Prädiktoren (EBLUP). Innerhalb dieser zwei Schätzer-Familien kann weiterhin zwischen den den Schätzern zugrunde liegenden Modellen unterschieden werden. Die folgende Aufstellung gibt eine Übersicht über die untersuchten modellbasierten Schätzer.

## Synthetische Schätzer

Modell A:

Angenommen, unit-level Daten von Hilfsvariablen  $x_{di} = (x_{di1}, \dots, x_{di p})^T$  seien für jede Einheit  $i$  in der Small Area Gesamtheit  $d$  verfügbar. Weiter sei die Zielvariable  $y_{di}$  durch ein lineares Regressionsmodell mit  $x_{di}$  und hierarchischen Fehlertermen verbunden

$$y_{di} = x_{di}^T \beta + u_d + \varepsilon_{di},$$

wobei  $\beta$  der Regressionskoeffizientenvektor ist,  $u_d$  der Area-spezifische Effekt mit  $E(u_d) = 0$ ,  $\text{var}(u_d) = \sigma_u^2$  und  $\text{var}(\varepsilon_{di})$  der unabhängige Zufallsfehler mit  $E(\varepsilon_{di}) = 0$  und  $\text{var}(\varepsilon_{di}) = \sigma_\varepsilon^2$  ist. Der synthetische Schätzer ist durch

$$\hat{\mu}_{Y,d}^{\text{SYNA}} = \mu_{X,d}^T \hat{\beta}$$

definiert mit bekanntem Area-level Kovariatenvektor  $\mu_{X,d} = (\mu_{X,d1}, \dots, \mu_{X,d p})^T$ , dem Vektor der wahren Mittelwerte von  $p$  Kovariaten ( $x_i$  ist  $p$ -dimensional) in der Area  $d$ .  $\hat{\beta}$  ist wie beim GREG-Schätzer definiert. Beispiele für Kovariate finden sich in Kapitel 4.

Modell B:

Der synthetische Schätzer B verwendet ein lineares (normalverteiltes) Modell mit Area-level Kovariaten und einer gepoolten Schätzung der Varianz innerhalb der Areas. Das Modell ist

$$\bar{y}_d = \mu_{Xd}^T \beta + \xi_d ,$$

wobei  $\xi_d$  der Area-spezifische Effekt mit  $E(\xi_d) = 0$  ist und  $\text{var}(\xi_d) = \sigma_u^2 + \psi_d$  gilt. Die Stichprobenvarianz ist  $\psi_d = \sigma_\varepsilon^2 / n_d$ , wobei  $n_d$  den Stichprobenumfang in der Area  $d$  bezeichnet.

## Empirische beste linear unverzerrte Prädiktoren (EBLUP)

Modell A:

$$\hat{\mu}_{Yd}^{\text{EBLUPA}} = \hat{\gamma}_d \hat{\mu}_{Yd}^{\text{GREG}} + (1 - \hat{\gamma}_d) \hat{\mu}_{Yd}^{\text{SYNA}} = \hat{\gamma}_d (\hat{\mu}_{Yd}^{\text{GREG}} - \hat{\mu}_{Xd}^T \hat{\beta}) + \mu_{Xd}^T \hat{\beta} ,$$

mit

$$\hat{\gamma}_d = \frac{\hat{\sigma}_u^2}{\hat{\sigma}_u^2 + \hat{\sigma}_\varepsilon^2 / n_d} .$$

Modell B:

$$\hat{\mu}_{Yd}^{\text{EBLUPB}} = \hat{\gamma}_d \hat{\mu}_{Yd}^{\text{GREG}} + (1 - \hat{\gamma}_d) \hat{\mu}_{Yd}^{\text{SYNB}} = \hat{\gamma}_d \hat{\mu}_{Yd}^{\text{GREG}} + (1 - \hat{\gamma}_d) \mu_{Xd}^T \hat{\beta} ,$$

mit

$$\hat{\gamma}_d = \frac{\hat{\sigma}_u^2}{\hat{\sigma}_u^2 + \hat{\psi}_d} .$$

Für alle Modelle gilt, dass  $\hat{\sigma}_u^2$ ,  $\hat{\sigma}_\varepsilon^2$  und  $\hat{\psi}_d$  adäquate Schätzer für  $\sigma_u^2$ ,  $\sigma_\varepsilon^2$  und  $\psi_d$  sind.

Eine eingehende Übersicht über Small Area-Verfahren kann der Monografie von Rao (2003) oder Münnich, Burgard und Vogt (2011) entnommen werden. Sowohl bei den modellunterstützten als auch bei den modellbasierten Verfahren werden die Daten der Melderegister als Hilfsvariablen herangezogen, insbesondere die Anschriftengröße.

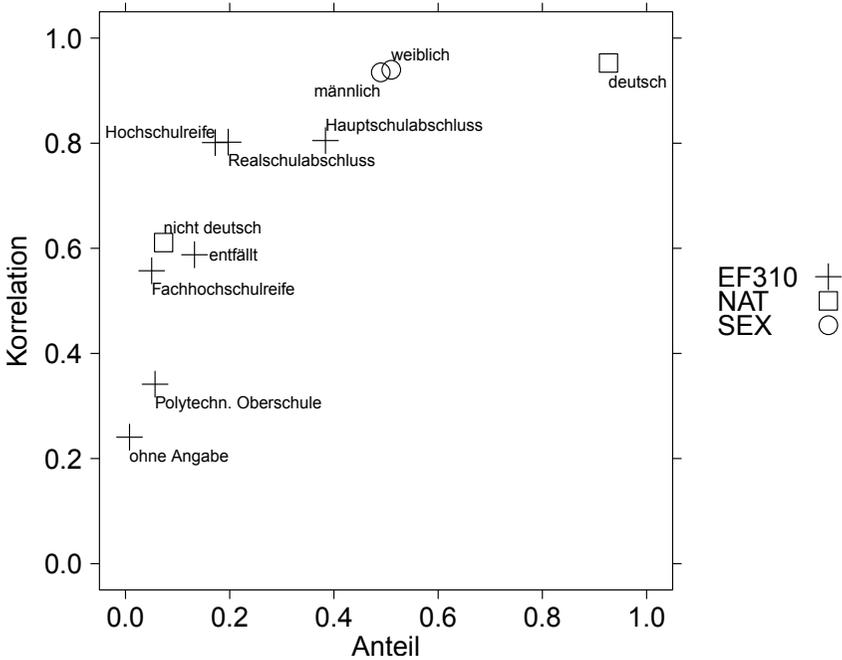
Von Interesse ist neben der Punktschätzung natürlich die Genauigkeit, welche im Rahmen dieser Untersuchungen durch den RRMSE der Schätzer quantifiziert wird. Die Schätzer hierfür lassen sich mitunter nicht in geschlossener Form

ausdrücken. Es existieren jedoch Approximationen, die zur Varianzschätzung benutzt werden können. Für eine Übersicht über einige wichtige Methoden sei auf Wolter (2007) und Münnich (2008) verwiesen. Als Benchmark dient die Varianz des GREG-Schätzers

$$\text{var}(\hat{\tau}_Y^{\text{GREG}}) = \sum_{h=1}^H N_h^2 \cdot \frac{S_{h,Y}^2}{n_h} \cdot \left(1 - \frac{n_h}{N_h}\right) \cdot (1 - \rho^2),$$

wobei  $S_{h,Y}^2$  die Varianz der Untersuchungsvariable in der Schicht  $h$  ist und  $\rho = 0,993$  die Korrelation zwischen der Hilfsvariable (hier: Anschriftengröße im Melderegister) und dem Totalwert einer Merkmalsausprägung in einer Anschrift einer interessierenden Ziel 1 oder Ziel 2 Variablen bezeichnet. Ist diese Korrelation hoch, so ist auch der Gewinn an Präzision durch Einbeziehen der Anschriftengröße als Hilfsvariable hoch. Voruntersuchungen im Zensusstest haben gezeigt, dass selbst Ziel 2 Variablen mitunter sehr hohe Korrelationen aufweisen, wie Abbildung 2 verdeutlicht.

Abbildung 2 Korrelation von Ausprägungen ausgewählter Ziel 2 Variablen mit der Anschriftengröße



Hier sind die Anteile der Ausprägungen der Ziel 2 Variablen Geschlecht (SEX), Nationalität (NAT) sowie höchster Bildungsabschluss (EF310) gegen deren jeweilige Korrelationen mit der Anschriftengröße dargestellt.<sup>8</sup>

Es ist leicht ersichtlich, dass mit steigendem Anteilswert tendenziell auch die Korrelation mit der Anschriftengröße zunimmt. Insgesamt bewegen sich die Korrelationen aller Ausprägungen dieser Ziel 2 Variablen jedoch ohnehin auf einem hohen Niveau, so dass ein merklicher Gewinn an Präzision durch Verwendung des GREG-Schätzers zu erwarten ist.

### 3 Optimale Allokation unter Nebenbedingungen

Auf Basis der oben definierten Anforderungen und Grundlagen empfiehlt sich ein geschichtetes Auswahlverfahren. Aufgrund der Tatsache, dass die Schichtvarianz  $S_{R,h,Y}^2$  der Anschriftengröße aus dem Melderegister zur Verfügung steht, liegt es nahe, die optimale Allokation nach Neyman-Tschuprov (Tschuprov 1923; Neyman 1934) zur Aufteilung des Gesamtstichprobenumfangs  $n$  auf die  $H$  Schichten zu verwenden. Dies ist wünschenswert, da optimale Allokationsverfahren zu Stichproben mit minimaler Varianz führen. Allerdings gilt dies nur insoweit, als ein hoher Zusammenhang zwischen der bei der Allokation verwendeten  $S_{R,h,Y}^2$  und den tatsächlichen Schichtvarianzen im Zensus  $S_{Z,h,Y}^2$  besteht. Da es sich bei den  $S_{R,h,Y}^2$  um Melderegisterdaten handelt, kann hier von einer hohen Übereinstimmung mit  $S_{Z,h,Y}^2$  ausgegangen werden und daher  $S_{R,h,Y}^2 = S_{Z,h,Y}^2 = S_{h,Y}^2$  angenommen werden.

Die Neyman-Tschuprov Allokation eines Gesamtstichprobenumfangs  $n$  auf  $H$  Schichten ist gegeben durch

$$n_h^{\text{opt,NT}} = n \cdot \frac{N_h S_{h,Y}}{\sum_{\ell=1}^H N_\ell S_{\ell,Y}} = n \cdot \frac{d_h}{\sum_{\ell=1}^H d_\ell} .$$

Bei identischen Schichtvarianzen entspricht die optimale Allokation der proportionalen Aufteilung. Doch anders als bei der proportionalen Aufteilung kann es bei der optimalen Aufteilung vorkommen, dass in einer Schicht mehr Anschriften ausgewählt werden sollen als Elemente in der Population existieren, also  $n_h > N_h$

8 Die sehr hohe Korrelation etwa der Ausprägung männlich beziehungsweise weiblich der Ziel 2 Variablen Geschlecht bedeutet inhaltlich, dass tendenziell in einer Anschrift gleich viele Frauen wie Männer leben.

resultiert. Dies kann passieren, wenn die Varianz  $S_{h,y}^2$  in einer Schicht besonders groß ist. Je nach Schichtungsvariante kommt dies in der praktischen Umsetzung durchaus vor, vor allem in den größten Anschriftengrößenklassen. Ein Teilstichprobenumfang mit  $n_h > N_h$  kann im Ziehungsmodell ohne Zurücklegen natürlich nicht realisiert werden und sollte von einem geeigneten Allokationsverfahren von vornherein vermieden werden. Darüber hinaus sollte auch kein zu geringer Auswahlatz innerhalb einer Schicht vorkommen, da hiermit möglicherweise sehr hohe RRMSEs verbunden sind. Verwendet man modellbasierte Verfahren, dann sollte man auch darauf achten, dass die Designgewichte nicht zu unterschiedlich ausfallen. In einem diskutierten Papier stellt Gelman (2007) die Problematik von Designgewichten bei statistischer Modellbildung heraus. Diese Problematik spielt bei der Anwendung von Small Area-Verfahren durchaus eine Rolle und führt zu einer weiteren Einschränkung der Gewichte in Form von oberen und unteren Grenzen, so genannten Box-Constraints.

Diese Restriktionen führen dazu, dass die folgenden Box-Constraints als Anforderung an die zu bestimmenden Stichprobenumfänge  $n_h$  definiert werden

$$m_h \leq n_h \leq M_h$$

mit bekannten Constraints  $m_h$  und  $M_h$ . Darüber hinaus darf ein vorab festgelegter Gesamtumfang an Personen nicht überschritten werden. Aufgrund der Tatsache, dass in einer Schicht Anschriften von verschiedener Größe zusammengefasst sind, kann die Personenzahl in einer Schicht von Stichprobe zu Stichprobe variieren. Daher kann auch nur eine mittlere erwartete Personenzahl bei gegebenem  $n_{A,h}$  angegeben werden. Die mittlere erwartete Personenzahl in Schicht  $h$  ist gegeben durch  $n_{A,h} \cdot \frac{N_{P,h}}{N_{A,h}}$ . Soll insgesamt der Anteil  $\theta$  von Personen in Deutschland ausgewählt werden, erhalten wir als zusätzliche Nebenbedingung

$$\sum_{h=1}^H n_{A,h} \cdot \frac{N_{P,h}}{N_{A,h}} = N_P \cdot \theta .$$

Diese Anforderungen können von der naiven Neyman-Tschuprov-Allokation im Rahmen geschichteter Zufallsstichproben nicht mehr ohne weiteres erfüllt werden. Das so gestellte Problem einer nicht-linearen Optimierung unter Nebenbedingungen kann jedoch zufriedenstellend gelöst werden, wie Gabler et al. (2010) zeigen. Ein einfacher Algorithmus ermöglicht die optimale Aufteilung eines Gesamtstich-

probenumfangs  $n$  auf  $H$  Schichten, wobei sowohl untere als auch obere Grenzen für die Stichprobenumfänge in den Schichten eingehalten werden und auch die oben genannte zusätzliche Nebenbedingung erfüllt wird. Dabei macht sich der Algorithmus die Tatsache zunutze, dass die Schichten exakt drei Klassen zugeordnet werden können. In Schichten, die der ersten Klasse  $U_1$  angehören, wird der Stichprobenumfang exakt auf die untere Schranke  $m_h$  gesetzt, in Schichten der zweiten Klasse  $U_2$  wird  $n_h$  exakt auf die obere Schranke  $M_h$  gesetzt und in der dritten Klasse  $U_3$  wird der verbleibende Stichprobenumfang  $n - \sum_{h \in U_1} m_h - \sum_{h \in U_2} M_h$  optimal im Sinne von Neyman aufgeteilt. Das Problem besteht somit darin, diejenige Zusammensetzung der Klassen zu finden, für die insgesamt eine bestimmte Zielfunktion minimiert wird. Der Algorithmus löst dieses Problem dadurch, dass zunächst zwei geordnete Reihen gebildet werden, in denen die Schichten entsprechend ihrer Ausprägung auf  $N_h \cdot S_{h,Y}$  in aufsteigender, beziehungsweise absteigender Reihenfolge angeordnet sind. Anschließend werden die Kombinationen dieser Ordnungen Schritt für Schritt abgearbeitet. Die erste Lösung, bei der alle Elemente aus  $U_3$  die Nebenbedingungen erfüllen, ist die angestrebte Lösung.

Details des Algorithmus inklusive Beispiel sind in Gabler et al. (2010) aufgeführt. Der Algorithmus wird für die Aufteilung des Gesamtstichprobenumfangs auf die Schichten im Rahmen des Forschungsprojekts verwendet. Im folgenden Abschnitt wird der Einfluss verschiedener Schichtungsvarianten und Box-Constraints auf die zu erwartende Präzision des GREG-Schätzers untersucht.

## 4 Ergebnisse

Der Grad der zu erwartenden Einhaltung der Präzisionsanforderungen spielt eine zentrale Rolle in der Bewertung der Eignung der potentiellen Allokations- und Auswahlverfahren. Die folgenden Darstellungen beruhen auf den Ergebnissen von Phase 1-Melderegisterdaten. Die in den Abbildungen dargestellten Box-Plots verdeutlichen die zu erwartenden Verteilungen des RRMSE nach SMP-Typen (SMP-Typ 0, SMP-Typ 1, SMP-Typ 2 und SMP-Typ 3). Bei den auf Bundesländern bezogenen Darstellungen kann es vorkommen, dass für verschiedene Typen keine Box-Plots ausgewiesen sind. Dies ist auf den Umstand zurückzuführen, dass in diesen Bundesländern gewisse SMP-Typen nicht vorkommen (siehe hierzu auch Tabelle 2).

Neben den oben beschriebenen Schichtungsvarianten (ADK1, ADK2 und ADK3) werden in den Simulationen darüber hinaus drei Varianten der Kombination von unterer und oberer Grenze der Schichtumfänge sowie unterschiedliche Grade

des Zusammenhangs  $\rho$  der Untersuchungsvariable  $Y$  mit den Anschriften aus den Melderegisterdaten angenommen. Allen Simulationen liegt ein beispielhafter Wert von  $\theta = 9,15\%$  zugrunde. Schließlich werden noch drei Varianz-Varianten (minimal, mittel, maximal) angenommen, welche die Art und Weise beeinflussen, wie die Untersuchungsvariable auf die Anschriften verteilt ist.<sup>9</sup>

In Abbildung 3 ist die Verteilung des relativen RMSE für eine Ziel 1 Fragestellung (also die amtliche Einwohnerzahl) dargestellt, wobei die mittlere Varianz-Variante und eine Korrelation der Anschriftengröße aus dem Melderegister mit der tatsächlichen Anschriftengröße von  $\rho = 0,993$  angenommen wird.<sup>10</sup> Innerhalb eines Panels verdeutlichen die vier Box-Plots die Verteilung des RRMSEs nach SMP-Typen. Der Box-Plot des SMP-Typs 1 setzt sich dabei zum Beispiel aus 1.621, der Box-Plot des SMP-Typs 0 dagegen aus 52 Datenpunkten zusammen. Der durchgezogene vertikale Strich markiert den höchstens erlaubten RRMSE, die gestrichelte vertikale Linie den Durchschnitt aller RRMSEs. Darüber hinaus markiert die durchgezogene Linie in einem Boxplot den Mittelwert der RRMSEs innerhalb eines SMP-Typs, der Punkt den Median. Die drei Panels in der unteren Reihe der Abbildung beziehen sich auf die Variante mit 1 % und 50 % als untere bzw. obere Grenze der Anteile  $n_{A,h} / N_{A,h}$  in Prozent, die Panels der mittleren Reihe auf die Variante mit 2 % und 40 % und die Panels der oberen Reihe auf die Variante mit 5 % und 20 %. Analog beruhen die Verteilungen, die in den Panels der ersten Spalte dargestellt sind, auf der Schichtungsvariante ADK1, diejenigen der zweiten Spalte auf ADK2 und die der dritten Spalte auf ADK3.

Aus Abbildung 3 wird ersichtlich, dass sich die Verteilungen der RRMSEs zwischen den einzelnen Panels zwar im Niveau unterscheiden, sich gewisse Muster aber in einer Mehrzahl der Panels finden. So ist etwa der Median der RRMSEs von SMPs vom Typ 2 in den Panels der unteren beiden Reihen der Abbildung 3 jeweils am höchsten, der Median der RRMSEs von SMPs vom Typ 1 dagegen am geringsten. Legt man das Kriterium an, dass der mittlere RRMSE über alle SMP-Typen am kleinsten sein soll, so ist von den Schichtungsvarianten ADK1 zu bevorzugen. Mit dem gleichen Kriterium würde innerhalb von ADK1 die Box-Constraints Variante mit 2 %-40 % favorisiert werden. Auch unter der Maßgabe, dass bei Ziel 1 ausschließlich die SMP-Typen 0 und 1 eine Rolle spielen, ist diese Box-Constraints-Variante zu bevorzugen.

9 Alle drei Schichtungsvarianten wurden zu Testzwecken untersucht. Letztendlich fiel die Entscheidung für acht Schichten in Personen-gleicher Allokation (ADK3).

10 Der Wert 0,993 geht auf Abschätzungen des Auftraggebers zurück.

Abbildung 3 RRMSE bei Ziel 1 Fragestellung; mittlere Varianz-Variante;  $\rho = 0,993$

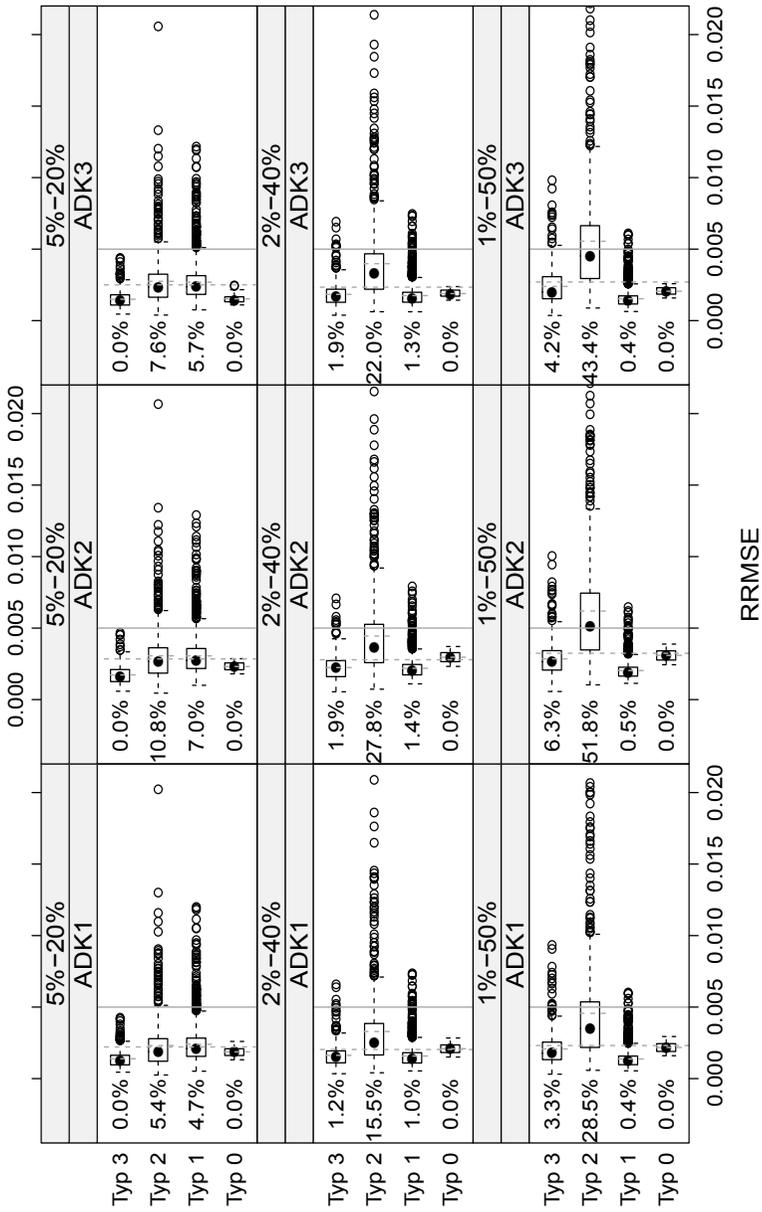


Abbildung 4 zeigt die Verteilung der RRMSEs analog bei einer (synthetischen) Ziel 2 Variable mit 50 % Anteilswert einer interessierenden Ausprägung und einer nur geringen Korrelation mit der Anschriftengröße von  $\rho = 0,6$  und mittlerer Varianz-Variante.

Nachfolgend werden nun die RRMSEs der Schätzung auf Basis von Phase 2-Daten grafisch verdeutlicht. In Abbildung 5 sind die RRMSEs der Erwerbstätigen-Schätzung für Rheinland-Pfalz und Nordrhein-Westfalen kartografisch aufgearbeitet dargestellt. Die schwarzen Linien stellen dabei die Begrenzungen der SMPs dar. Das Ausmaß der RRMSEs ist farblich gekennzeichnet. Dabei stehen hellgraue Färbungen für niedrige RRMSEs. Je dunkler die Einfärbung, desto höher ist der RRMSE in dem jeweiligen SMP. SMPs mit RRMSEs über 10 % sind in sehr dunklem Grau gekennzeichnet.

Dargestellt werden die Ergebnisse für die in Abschnitt 2.3 beschriebenen vier Schätzer. Unten links der HT Schätzer, unten rechts der Small Area GREG-Schätzer, oben links ein Schätzer beruhend auf einem synthetischen Unit-Level Modell und oben rechts der EBLUPA. Als Hilfsvariablen für die Modelle wurden jeweils die Anschriftengröße, der Anteil an Ausländern sowie die Anschriftengrößenklasse verwendet (unter 3 Personen, 3 bis unter 7 Personen und ab 7 Personen).

Beim Vergleich der beiden designbasierten Verfahren (HT und GREG) schneidet der GREG leicht besser als der HT ab. Insbesondere die beim HT dunkleren SMPs sind beim GREG etwas heller. Auch weist der GREG keinen SMP mit einem RRMSE über 10 % auf. Jedoch ist die Verbesserung des GREG gegenüber dem HT nicht besonders stark ausgeprägt. Dies rührt daher, dass die verwendeten Hilfsvariablen keine hohe Erklärungskraft für die Untersuchungsvariable aufweisen. Da als Hilfsvariablen lediglich Informationen verwendet werden können, die für ganz Deutschland vorhanden sind, können nur Melderegistervariablen als Hilfsinformation herangezogen werden. Dazu gehören Variablen wie Alter, Geschlecht, Nationalität oder Wohnstatus. Für viele Ziel 2 Fragestellungen bieten diese Variablen wenig Erklärungskraft. Weitere Hilfsvariablen aus anderen Registern, etwa erwerbsstatistischen Registern, versprechen hierbei jedoch enorme Verbesserungen der Schätzqualität.

Abbildung 4 RRMSE bei Ziel 2 Fragestellung; mittlere Varianz-Variante; 50 % Anteilswert;  $\rho = 0,6$

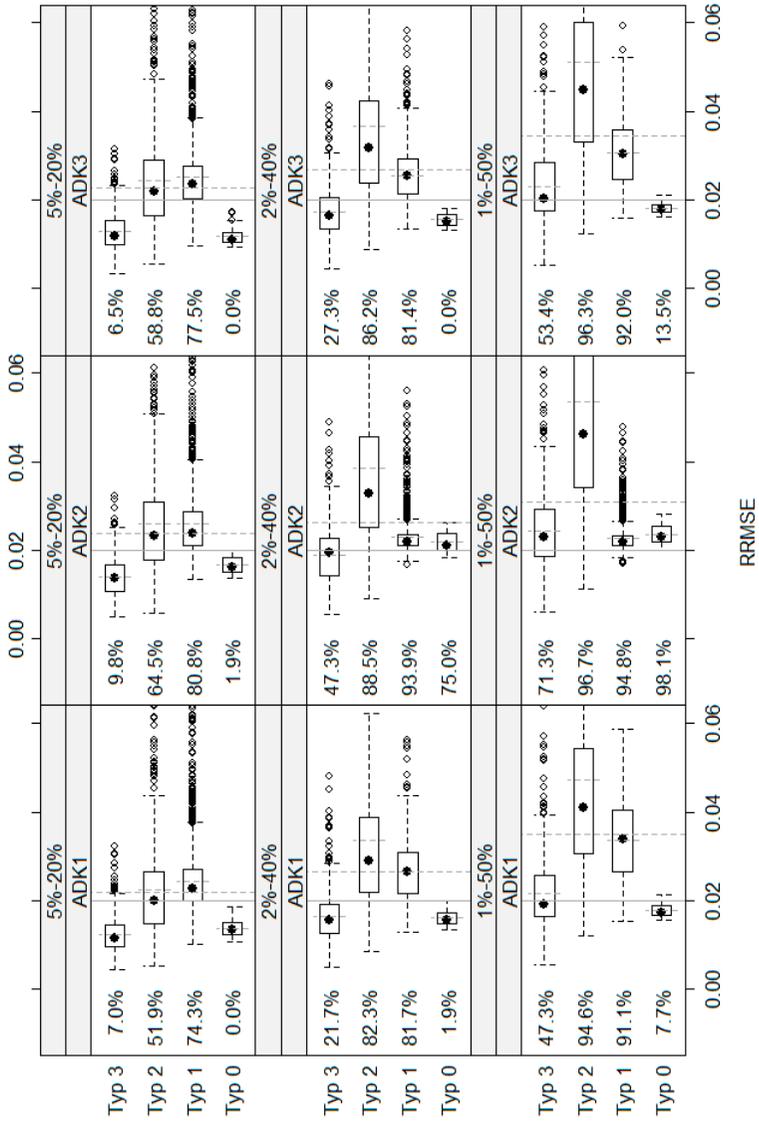
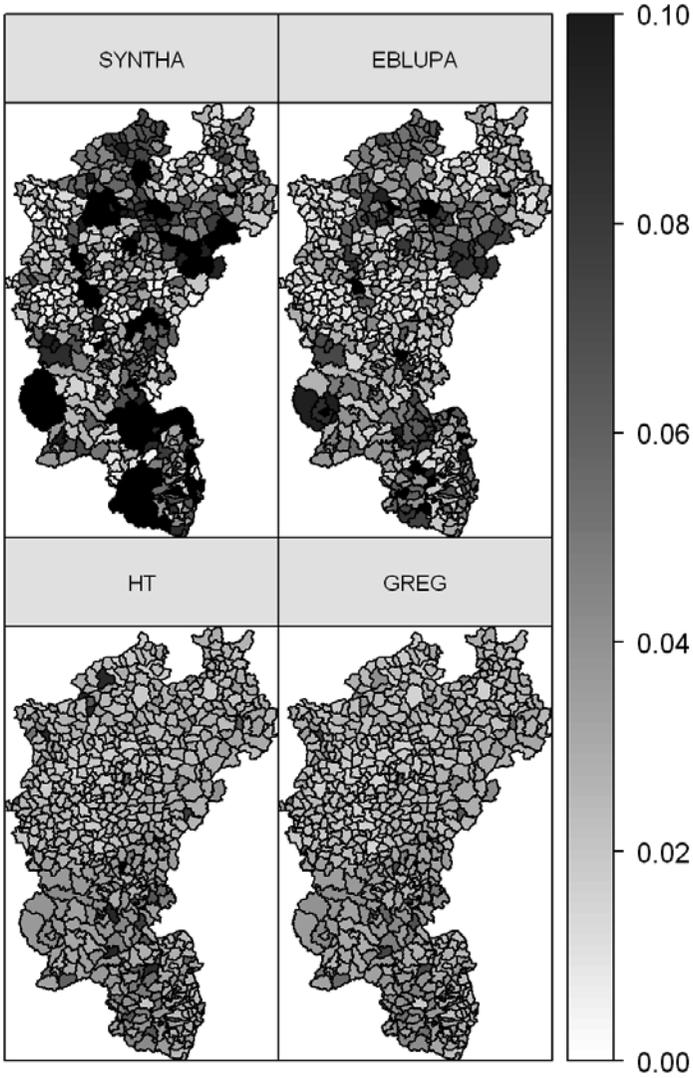


Abbildung 5 Ergebnisse der Small Area-Schätzungen in Rheinland-Pfalz und Nordrhein-Westfalen



Beim GREG werden die Hilfsvariablen lediglich als Unterstützung des design-basierten Schätzers verwendet. Die Erklärungskraft dieser Variablen in Bezug auf die Untersuchungsvariable wirkt sich fast ausschließlich positiv auf die Schätzqualität aus. Dagegen hängt die Qualität der Small Area-Schätzer fast ausschließlich von den Hilfsvariablen ab. Dies bedeutet auch, dass ungünstige Hilfsvariablen einen erheblichen Bias in den Schätzungen hervorrufen können. Daher ist in dieser Situation nicht weiter verwunderlich, dass der synthetische Unit-Level Schätzer bei diesem Vergleich am schlechtesten abschneidet. Die Variablen können nicht die Variabilität zwischen den einzelnen SMPs erklären.

Hierin begründen sich auch die Probleme des EBLUPA, mit welchem versucht wird, die Area-Schätzwerte aufgrund eines synthetischen Schätzers zu ermitteln, dabei aber korrigiert um beobachtete Abweichungen von diesem Modell. Hierbei fällt auf, dass der EBLUPA zwar in einigen Areas erwartungsgemäß sehr niedrige RRMSEs ausweist, jedoch in zahlreichen SMPs aufgrund der nicht passenden Modellierung schlechtere Ergebnisse als der GREG liefert.

Aus diesen einfachen Resultaten lässt sich unmittelbar das Potential der Small Area Modelle erkennen. Allerdings bergen sie auch die Gefahr, erhebliche Effizienzverluste zu verursachen, jedenfalls bei unsachgemäßem Einsatz, etwa wenn keine guten Hilfsinformationen verwendet werden. Gerade bei sehr kleinen (Teil-)Stichprobenumfängen wird man jedoch nicht mehr auf die Small Area Modelle verzichten können.

In zahlreichen Simulationen kristallisiert sich heraus, dass je nach Fragestellung entweder der GREG-Schätzer oder der EBLUPA verwendet werden soll. Mittlerweile steht auch eine erweiterte Version des EBLUPA zur Verfügung, welche bei der Small Area-Schätzung Gewichte verwendet. Diese Methode erweist sich beim derzeit umgesetzten Stichprobendesign bisher als besser geeignet als der EBLUPA.

## 5 Schlussfolgerungen

Der vorliegende Artikel gibt einen Überblick über die Herausforderungen und erste Ergebnisse innerhalb des Projektes *Zensus 2011 – Projekt zur methodischen Grundlagenforschung* und die gewählte Herangehensweise, um diesen Herausforderungen zu begegnen. Bis Ende des Projektes werden noch einige Erweiterungen der Schätzmethodik ausgebaut und evaluiert. Unter anderem muss die MSE-Schätzung von Small Area Methoden noch verfeinert werden.

Das vorgestellte Stichprobendesign verwendet eine neue Variante einer optimalen Allokation unter Nebenbedingungen. Hierdurch werden verschiedene

Zielsetzungen realisiert. Auf der einen Seite können Mindeststichprobenumfänge formuliert werden, auf der anderen Seite kann die Variabilität von Designgewichten eingeschränkt werden, wodurch der statistische Modellbau erleichtert wird, der auch bei der Verwendung von Small Area Modellen benötigt wird. Dies kommt auch dem inhaltlich interessierten Forscher zugute, der später eigene Analysen mit den Zensusdaten durchführt, sofern Scientific Use Files zur Verfügung gestellt werden. Ganz nebenbei verhindern die Box-Constraints auch eine allzu ungleiche Befragungswahrscheinlichkeit der Bevölkerung.

Nach bisherigen Ergebnissen werden unter Verwendung der Vorgaben vom Auftraggeber die Qualitätsziele für das Ziel 1 erreicht. Eine Verfeinerung der Methoden wird eventuell auch noch einzelne kleinere Verbesserungen ermöglichen. Im Falle von Ziel 2 gestaltet sich die Beurteilung erheblich problematischer, da effiziente Hilfsvariablen kaum zur Verfügung stehen. Ebenso variieren die Genauigkeiten der Schätzergebnisse vielmehr auf Grund der Tatsache, dass in den Sampling Points und Schichten bei unterschiedlichen Variablen die Beobachtungsanteile viel stärker variieren und innerhalb der Schichten heterogene Subgruppen auftreten. Insbesondere bei modellbasierten Schätzverfahren können nur geeignete Modelle diese Strukturen geeignet kompensieren. Der überwiegende Teil der Randschätzungen von Ziel 2-Variablen wird aber voraussichtlich die Präzisionsziele erfüllen, sofern nicht zu geringe Anteile einzelner Ausprägungen auftreten.

## Literatur

- Burgard, J. P., 2009: Erstellung von Karteileichen und Fehlbestandsmodellen durch Multi-level Modelle. Diplomarbeit an der Universität Trier.
- Burgard, P. und R. Münnich, 2010: On the Impact of Over and Undercounts on Small Area Estimates in Register-based Censuses. *Computational Statistics and Data Analysis*, <http://dx.doi.org/10.1016/j.csda.2010.11.002>.
- Devroye, L., 1986: *Non-Uniform Random Variate Generation*. Springer.
- Gabler, S., M. Ganninger und R. Münnich, 2010: Optimal Allocation of the Sample Size to Strata Under Box Constraints. *Metrika*, DOI: 10.1007/s0018401003193.
- Gelman, A., 2007: Struggles With Survey Weighting and Regression Modeling. *Statistical Science* 22 (2): 153–164.
- Kleber, B., A. Maldonado, D. Scheuregger und K. Ziprik, 2009: Aufbau des Anschriften und Gebäuderegisters für den Zensus 2011. *Wirtschaft und Statistik* 7: 629–640.
- Münnich, R., 2008: Varianzschätzung in komplexen Erhebungen. *Austrian Journal of Statistics* 37: 319–334.
- Münnich, R., P. Burgard und M. Vogt, 2009: Small Area Estimation for Population Counts in the German Census 2011. Section on Survey Research Methods JSM 2009.
- Münnich, R., S. Gabler und M. Ganninger, 2007: Some Remarks on the Registerbased Census 2010/2011 in Germany. Proceedings of the Workshop Innovative Methodologies for Censuses in the New Millennium Southampton.
- Münnich, R. und J. Schürle, J., 2003: On the Simulation of Complex Universes in the Case of Applying the German Microcensus. Technical report, DACSEIS research paper series 4.

- Münnich, R. und M. Vogt, 2011: Small Area Methoden: Modelle, Anwendungen und Praxis. Wirtschafts- und Sozialstatistisches Archiv. In Fertigstellung.
- Neyman, J., 1934: On the two Different Aspects of the Representative Method: The Method of Stratified Sampling and the Method of Purposive Selection. *Journal of the Royal Statistical Society* 97: 558–606.
- Rao, J. N. K., 2003: Small Area Estimation. Wiley Series in Survey Methodology. New York: John Wiley and Sons.
- Schäfer, J., 2004: Ergänzende Verfahren für einen künftigen registergestützten Zensus. *Statistische Analysen und Studien NRW*, 17: 20–27.
- Tschuprov, A. A., 1923: On the Mathematical Expectation of the Moments of Frequency Distributions in the Case of Correlated Observations. *Metron*, 2: 461–493, 646–683.
- Wolter, K. M., 2007: Introduction to Variance Estimation. New York: Springer, 2 edition.

## Anschrift der Autoren

Prof. Dr. Ralf Münnich  
Fachbereich IV – VWL  
Lehrstuhl für Wirtschafts- und Sozialstatistik  
Universitätsring 15  
54286 Trier  
muennich@uni-trier.de

PD Dr. Siegfried Gabler  
GESIS – Leibniz-Institut für Sozialwissenschaften  
B 2, 1  
68159 Mannheim  
siegfried.gabler@gesis.org

Dr. Matthias Ganninger  
GESIS – Leibniz-Institut für Sozialwissenschaften  
B 2, 1  
68159 Mannheim  
matthias.ganninger@gesis.org

Jan Pablo Burgard  
Fachbereich IV – VWL  
Lehrstuhl für Wirtschafts- und Sozialstatistik  
Universitätsring 15  
54286 Trier  
jpburgard@uni-trier.de

Jan-Philipp Kolb  
Fachbereich IV – VWL  
Lehrstuhl für Wirtschafts- und Sozialstatistik  
Universitätsring 15  
54286 Trier  
kolb@uni-trier.de