



# Small Area Schätzungen-Personen

P. Eichenberger, A. Massiani, J.-M. Nicoletti

Sektion Statistische Methoden METH

Bundesamt für Statistik

*Informationsveranstaltung VZ*

*10. Oktober 2013*



## Das Projekt Small Area Schätzungen-Personen

- Ziel : Schätzer für kleine Gebiete verbessern. Es wurde entschieden, mit der Schätzung der Anzahl Erwerbstätiger in Gemeinden oder Gruppen von Gemeinden der Schweiz zu beginnen.
- Die Kleingebiete sind hier somit die Gemeinden oder Gruppen von Gemeinden.
- Schwierigkeit : Direkte Schätzmethoden sind für Schätzungen in kleinen Gemeinden nicht effizient.
- Lösung : Schätzmethoden basierend auf Kleingebieten verwenden.



## Mandat an Universität Trier

- Mandat hat 2010 begonnen und wurde anfangs 2012 beendet.
- Verwendete Daten : Volkszählung 2000.
- Verschiedene Modelle analysiert, darunter insbesondere das Modell :  
Die abhängige Variable (Antwortvariable)

$$Y_i = \begin{cases} 1 & \text{falls das Individuum } i \text{ erwerbstätig ist,} \\ 0 & \text{sonst.} \end{cases}$$

Das lineare Modell :

$$Y_i = \alpha + \text{Alter}_{i,1} + \text{Alter}_{i,2} + \text{Alter}_{i,3} + \text{Geschlecht}_i + \text{GrossRegionen}_{i,1} + \dots + \text{GrossRegionen}_{i,8} + \varepsilon_i$$



## Mandat an Universität Trier

- Resultate : Eine deutliche Verbesserung der Genauigkeit in einer grossen Anzahl Gemeinden, allerdings :
  - ▶ Für einige Gemeinden, ist der Bias gross, ohne dass dieser erklärt werden kann.
  - ▶ Kein Vorschlag, um problematische Gemeinden zu entdecken.
  - ▶ Keine Modelldiagnose.
- Riskant Resultate zu publizieren, die in einigen Fällen einen grossen Bias aufweisen können und nicht glaubwürdig sind.



## Erster Ansatz zur Verbesserung des Modells

- Verbesserung des Modells : Vorschlag von Methoden zur Modelldiagnose und hinzufügen von Interaktionen zwischen den Variablen.
- Man basiert sich immer auf Daten der Volkszählung 2000.
- Resultate : Das Modell konnte mit den erklärenden Variablen der Volkszählung 2000 nicht signifikant verbessert werden.
- Schlussfolgerung : Suche von neuen Hilfsvariablen.



## Vergleich der Auswertungen der Modelle

Modell Uni Trier	Neues Modell
82.26%	82.72%

**TABLE :** Prozentsatz der Personen, die korrekt als erwerbstätig oder nicht-erwerbstätig nachgewiesen wurden.



## Verwendung der AHV Daten

- Entwickeln eines Modells basierend auf den Daten der Strukturerhebung (SE2012), verwenden der STATPOP2012 Daten.
- Zur Verfügung : zwei Variablen aus der AHV, eine die angibt, ob eine Person im Jahr 2011 einen AHV-Beitrag entrichtet hat und eine zweite mit der AHV-pflichtigen Lohnsumme.
- Termin → April 2014.



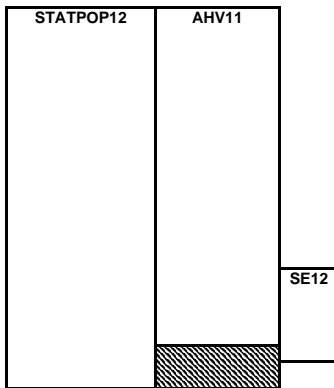
## Verwendung der AHV Daten







## Verwenden der AHV Daten





## Verwenden der AHV Daten

- Zurzeit berücksichtigte Modelle :

$$M_1 : Y_i = \alpha + AHVBeitragVorhanden_i + \varepsilon_i$$

$$M_2 : Y_i = \alpha + AHVBeitragVorhanden_i + Alter_{i,1} \\ + Alter_{i,2} + Alter_{i,3} + Geschlecht_i + Zivilstand_i \\ + Nationalitaet_i + \varepsilon_i,$$



## Erste Resultate

$M_1$	$M_2$
90.73%	90.74%

**TABLE :** Prozentsatz der Personen, die korrekt als erwerbstätig oder nicht-erwerbstätig nachgewiesen wurden.



## Schlussfolgerung

- Ein Modell, welches nur übliche Variablen verwendet ist ungenügend.
- Die Variable Vorhandensein eines AHV-Beitrags verbessert das Modell deutlich. Allerdings kann diese Variable beispielsweise unter Berücksichtigung des Lohnes verbessert werden.