



Anwendung von « small area estimation » Methoden und experimentelle Statistiken

Schätzung der Erwerbsbeteiligung anhand der 3-Jahres Poolingdaten der Strukturerhebung

Marie Dupraz, Anne Massiani, Christoph Freymond, Daniela Wanner

Bundesamt für Statistik

Schweizer Statistiktage | 29.08.2018



TEIL 1 : METHODISCHE ASPEKTE



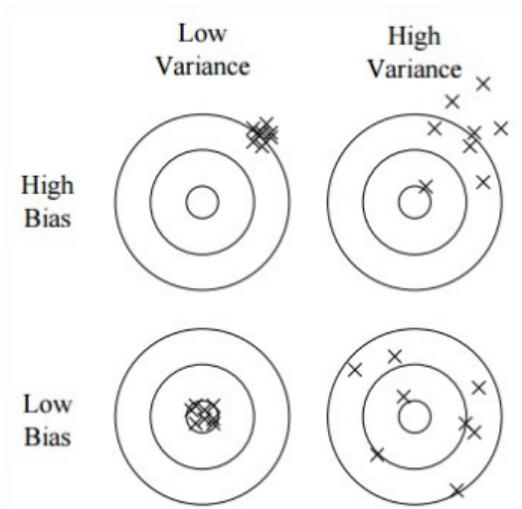
1. Einführung

- ▶ **Ziel:** Schätzung der Erwerbsbeteiligung auf Gemeindeebene in der Strukturerhebung.
Oft kleine Stichprobengrösse \Rightarrow Zuhilfenahme von Methoden des Typs “Small Area Estimation” (SAE).
- ▶ **Erster Schritt: Mandat an die Universidad Carlos III de Madrid (UC3M)**
 \Rightarrow Unter bestimmten Bedingungen, zuverlässige jährliche Schätzungen für 2012 (SST 2016).
- ▶ **Zweiter Schritt: Fortführung der Arbeiten im BFS**
 - ▶ Aneignung des Know-hows im BFS: Produktion jährlicher Schätzungen (2012-2013-2014), Analyse der Stabilität.
 - ▶ Pooling über drei Jahre und Analyse des Einflusses auf die Resultate.



2. Grundsätze der SAE und Resultate des UC3M Mandats

Zwei Grundbegriffe, um die Herausforderungen der SAE zu verstehen: Bias und Varianz eines Schätzers





Die SAE Schätzung basiert auf :

- ▶ Hilfsvariablen, die für die ganze Population bekannt sind,
- ▶ ein Modell zwischen interessierender Variablen und dieser Hilfsvariablen.

Unter den Hypothesen des Modells: Varianzgewinn im Vergleich zu “traditionellen” Schätzern (Verwendung von mehr Information).

Aber Biasrisiko, wenn die Hypothesen nicht erfüllt sind!

Beispiel: wesentliche Resultate aus der Sturkturerhebung

- ▶ Zuverlässige Schätzungen der Erwerbsbeteiligung und deren Genauigkeit für Gemeinden mit einer Stichprobengrösse von mindestens 100 Personen.
- ▶ Median-Gewinn der Genauigkeit: 78%.
- ▶ Dies entspricht einer Multiplikation der Stichprobengrösse mit 8.



Vergleich “traditionelle” Schätzung vs SAE bezüglich Genauigkeit

	Bias	Varianz
Traditioneller Schätzer	Im Prinzip, kleiner Bias	Hängt von der Stichprobengrösse in der Domaine ab
SAE Schätzer	Mehr oder weniger grosses Risiko in Abhängigkeit der Modellvalidität	Generell klein

Verwendete Genauigkeitsschätzung, um die Schätzer zu beurteilen:

$$\text{Design Mean Squared Error (MSE)} = \text{Varianz} + \text{Bias}^2$$

ou

$$\text{RMSE} = \sqrt{\text{MSE}} \text{ (gleiche Einheit wie Punktschätzung)}$$

Genauigkeitsschätzung extrem komplex!



3. Modell

► **Datenquellen**

1. Erhebungsdaten: Strukturerhebung;
2. Hilfsinformationen: Statistik der Bevölkerung und der Haushalte (STATPOP) und AHV Daten des Vorjahres.

► **Modell:** Linear Mixed Model mit Zufallseffekt für die Gemeinde.

► **Schätzer, der sich daraus ergibt:** Kombination eines Schätzers vom Typ “traditionell” und eines synthetischen Schätzers (d.h. auf einem “klassischen” linearen Modell ohne spezifischen Gemeindefeffekt basierend).

Je grösser die Stichprobengrösse in der Gemeinde, desto wichtiger wird der Anteil des “traditionellen” Schätzers (Verringerung des Biasrisikos).



4. Einfluss des Poolings auf die Anzahl publizierbarer Gemeinden

Die Verlässlichkeit der Resultate (Schätzung und Genauigkeit) wird für Gemeinden mit einer Stichprobengrösse von mindestens 100 Personen als zufriedenstellend erachtet.

⇒ Einschränkung der Anzahl Gemeinden, für welche Resultate publiziert werden.

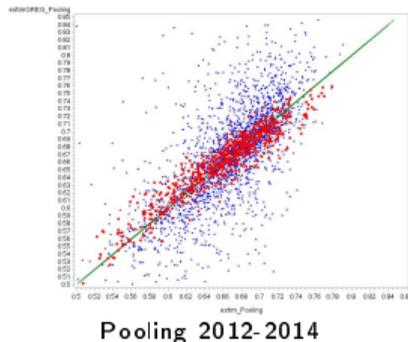
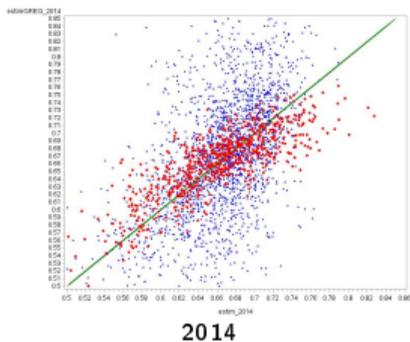
- ▶ Jährliche Schätzungen: ca. 30% der Gemeinden.
- ▶ Pooling über drei Jahre: ca. 60% der Gemeinden.
- ▶ Pooling über fünf Jahre: ca. 70% der Gemeinden?



5. Einfluss des Poolings (3 Jahre) auf das Bias-Risiko

Grafiken: (Punkt-)Schätzungen SAE versus traditionelle Schätzungen.

- ▶ Erlaubt einen eventuellen Bias der SAE Schätzungen zu entdecken, wenn die Punkte nicht uniform um die Gerade $y = x$ verteilt sind.
- ▶ Rote Punkte: Gemeinden, mit einer jährlichen Stichprobe von mindestens 100 Personen.

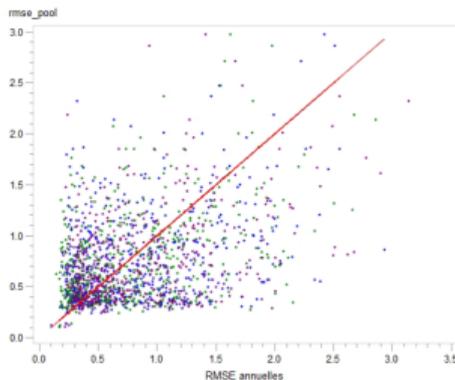


- ▶ “Traditionelle” Schätzung stabiler mit Pooling.
- ▶ Je grösser die Stichprobengrösse, desto mehr nähert sich der SAE Schätzer dem “traditionellen” an.



6. Einfluss des Poolings (3 ans) auf Genauigkeitsschätzung

Jährliche RMSE vs. RMSE Pooling. Eine Farbe pro Jahr.



Diese Frage:

- ▶ muss noch vertieft analysiert werden,
- ▶ unterstreicht die Komplexität der (design)-RMSE Schätzung.

Die Punkte sind um die Gerade $y = x$: keine flagrante Verbesserung für die **Schätzung** der Genauigkeit.

Hypothesen:

- ▶ entweder verbessert das Pooling die Genauigkeit wenig,
- ▶ oder Schätzmethode der RMSE ist nicht genügend genau, um den Pooling-Effekt auf den RMSE zu entdecken.



7. Vorteile und Einschränkungen der Verwendung des Poolings

Vorteile

- ▶ Anzahl publizierbarer Gemeinden wird erhöht.
- ▶ Positiver Einfluss auf Biasrisiko.

Einschränkungen

- ▶ Populationsdefinition: Mittlere Population, Gemeindemutationen.
- ▶ Glättung der Entwicklungen aufgrund der Verwendung einer mittleren Population.

Vorteile und Einschränkungen bezüglich SAE Methoden kommen noch dazu: mögliche grosse Genauigkeitsgewinne, aber Wahl und Validation eines Modells, Komplexität der design-RMSE Schätzung, Herausforderung für die Kommunikation, etc.



8. Methodische Perspektiven

- ▶ Pooling über 5 Jahre?
- ▶ Verbesserungen oder Anpassungen des Modells, z.B. Integration neuer Hilfsvariablen?
- ▶ Analyse des Poolingeffekts auf die Schätzung des design-RMSE.
- ▶ Einbezug des Feedbacks nach der Publikation der experimentellen Statistiken.



TEIL 2 : POTENZIELLE VERWENDUNG, SCHLUSSFOLGERUNGEN UND PERSPEKTIVEN

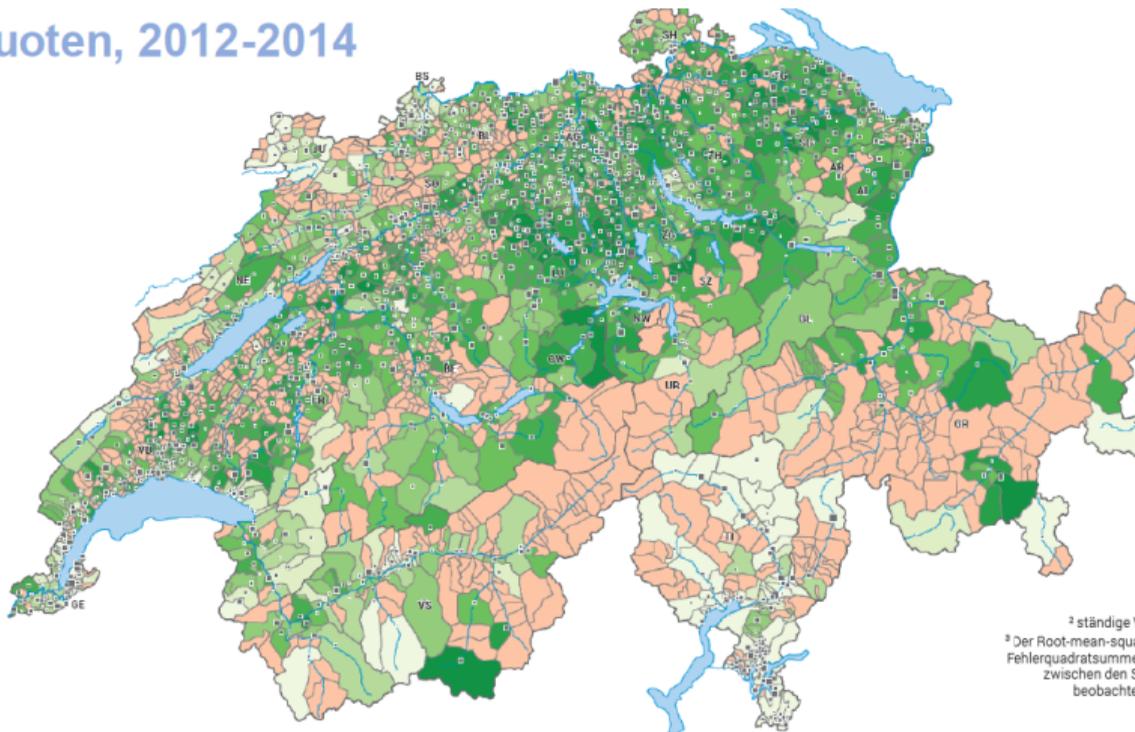


Erwerbsquoten, 2012-2014

Anteil Erwerbspersonen an der Bevölkerung², in %



Design Root-mean-squared error³ (*100)

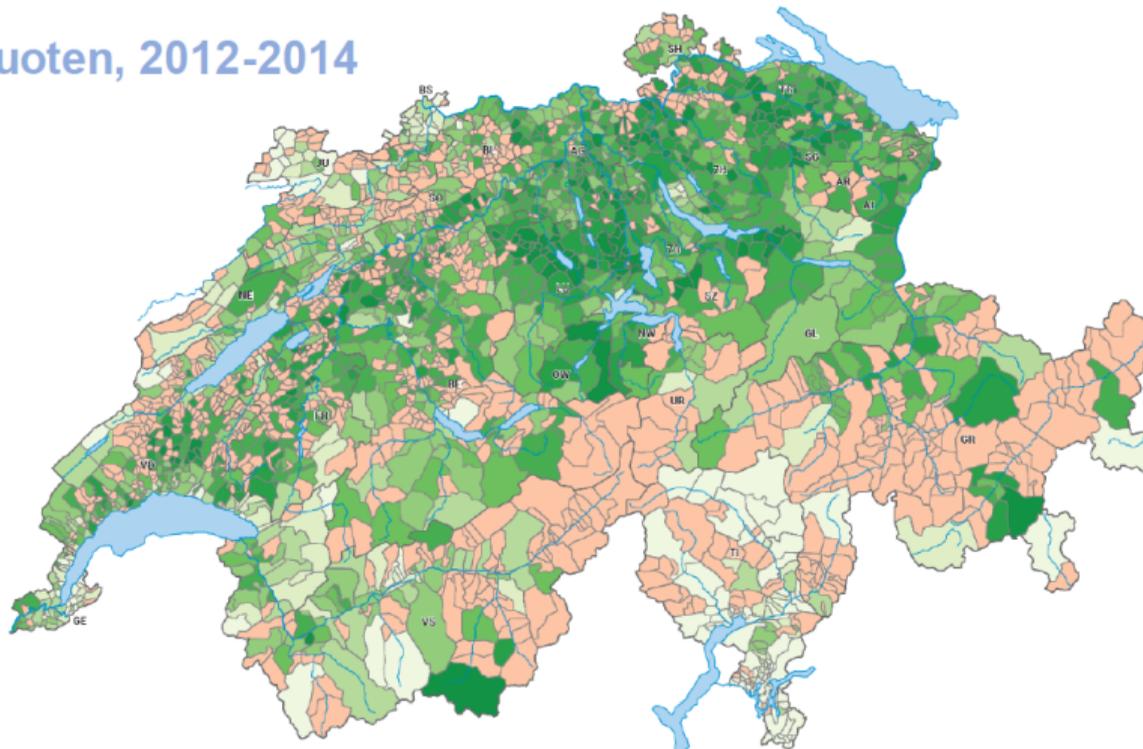
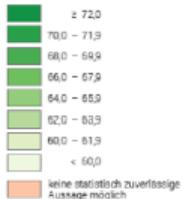


² ständige Wohnbevölkerung ab 15 Jahren
³ Der Root-mean-squared Error (Wurzel der mittleren Fehlerquadratsumme) ist ein Mass, um die Differenz zwischen den Schätzungen eines Modells und beobachteten Messungen auszudrücken.



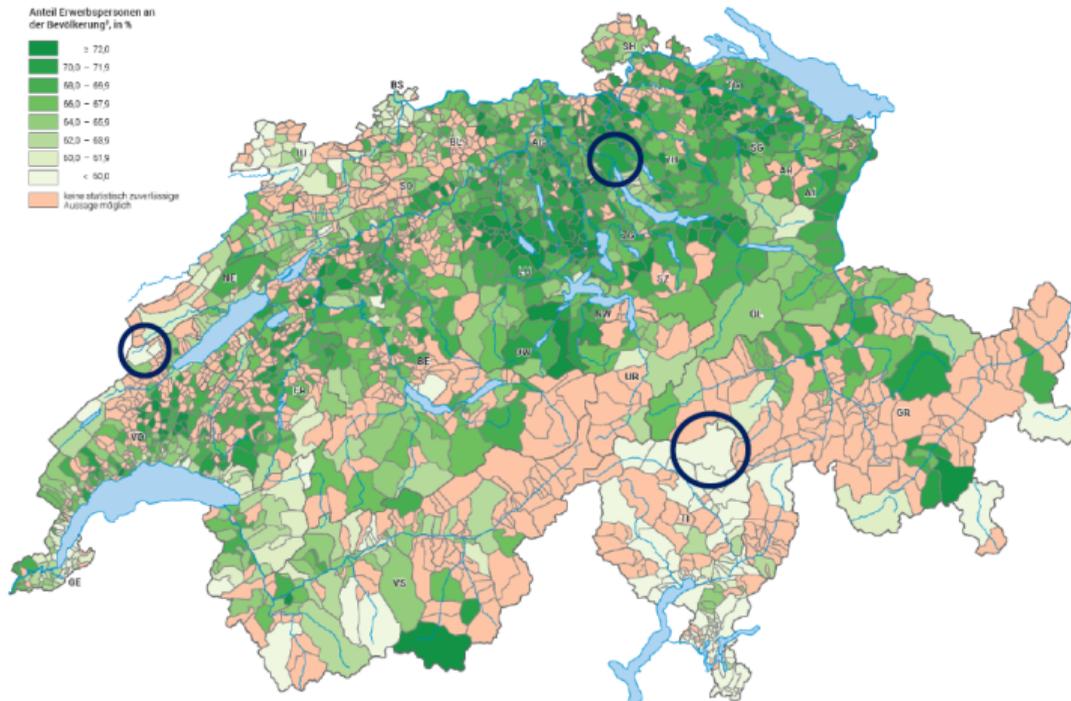
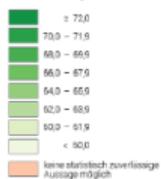
Erwerbsquoten, 2012-2014

Anteil Erwerbspersonen an
der Bevölkerung¹, in %





Anteil Erwerbspersonen an der Bevölkerung¹, in %



	Blenio	Sainte-Croix	Zürich
Geschätzte standardisierte Erwerbsquote (in %) gemäss Modell	54	54	71
design-RMSE	1.11	0.52	0.11
Einwohner 2014	1'726	4'766	391'359
dawon (in %)			
0-14 Jahre	12	15	13
15-64 Jahre	62	61	71
65 Jahre und mehr	27	24	16
Beobachtungen Pooling 2012-2014 der SE	269	711	28'558
Gesamtfläche (in km ²)	202.2	39.3	87.9
dawon Siedlungsfläche (in %)	1.3	7.0	61.9



Experimentelle Statistiken

Was heisst das?

- ▶ neben der normalen Statistikproduktion
- ▶ neue Methoden / Datenquellen
- ▶ noch nicht vollständig ausgereift

Ziele?

- ▶ frühere Einbindung unserer Partner
- ▶ optimale Berücksichtigung derer Bedürfnisse
- ▶ insgesamt agiler reagieren

Was bedeutet das für dieses Projekt?

1. Tests mit einem Pooling über 5 Jahre
2. weitere Erkenntnisse und Empfehlungen
3. Übergang in laufende Statistikproduktion?



Die Reise geht weiter ...

www.experimental.bfs.admin.ch

... zusammen mit Ihnen

EXPERIMENTAL STATISTICS