

Vergleich des Fay-Harriot (FH) und des Battesse-Harter-Fuller (BHF) Models

Niklas, Lorenz

Institut für Statistik | Universität Bamberg



1) Motivation

Wir vergleichen die Schätzgenauigkeit von Area-Level- (FH) und Unit-Level-Modellen (BHF) bei sehr kleinen Stichproben. Ein Simulationsansatz ermöglicht den direkten Vergleich der Schätzungen mit der bekannten wahren Verteilung der Zielvariable.

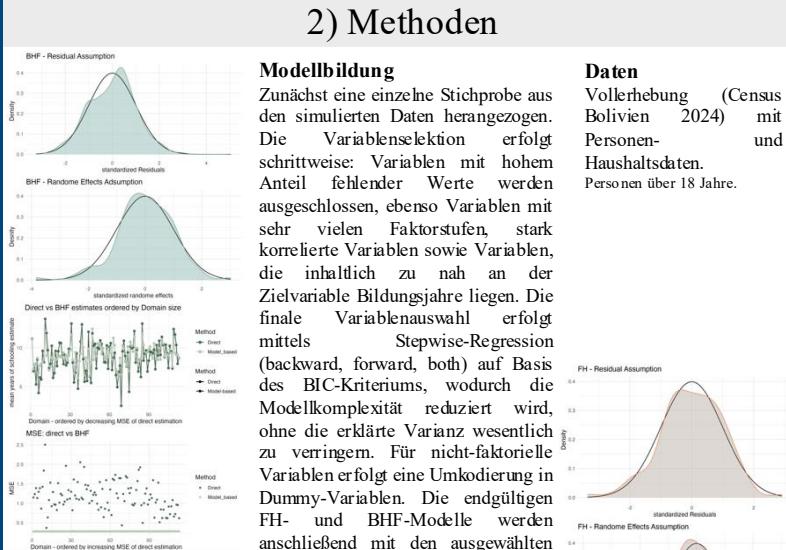
2) Methoden

Modellbildung

Zunächst eine einzelne Stichprobe aus den simulierten Daten herangezogen. Die Variabelselektion erfolgt schrittweise: Variablen mit hohem Anteil fehlender Werte werden ausgeschlossen, ebenso Variablen mit sehr vielen Faktorstufen, stark korrelierte Variablen sowie Variablen, die inhaltlich zu nah an der Ziellvariable Bildungsjahre liegen. Die finale Variableauswahl erfolgt mittels Stepwise-Regression (backward, forward, both) auf Basis des BIC-Kriteriums, wodurch die Modellkomplexität reduziert wird, ohne die erklärte Varianz wesentlich zu verringern. Für nicht-faktorielle Variablen erfolgt eine Umcodierung in Dummy-Variablen. Die endgültigen FH- und BHF-Modelle werden anschließend mit den ausgewählten Kovariaten geschätzt. Eingeschlossen sind sowohl individuelle Variablen wie Alter, Beruf und Lesefähigkeit, als auch haushaltsbezogene Indikatoren des Wohn- und Lebensstandards, darunter Urbanität, Krankenversicherung, Wohnqualität, Ausstattung der Wohnung (z. B. Küche, Warmwasserbereitung) und Autobesitz.

Daten

Vollerhebung (Census Bolivien 2024) mit Personen- und Haushaltsdaten. Personen über 18 Jahre.



Simulation

Simulation von 200 Stichproben mittels SRS mit $n = 20$ pro Domäne für 113 Provinzen. Die Formel, die für das erste Sample ermittelt wurden dann auf diese angewandt.

