

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

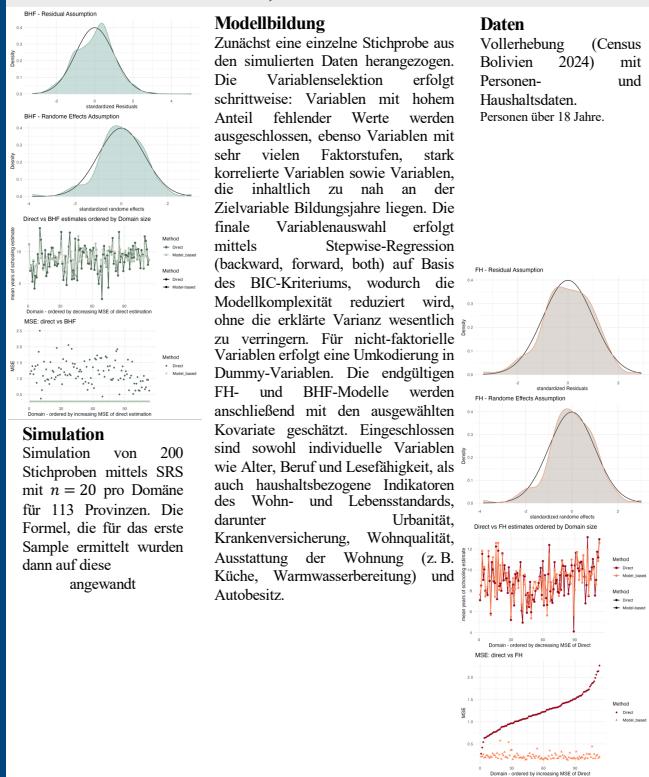
Niklas, Lorenz

Institut für Statistik | Universität Bamberg

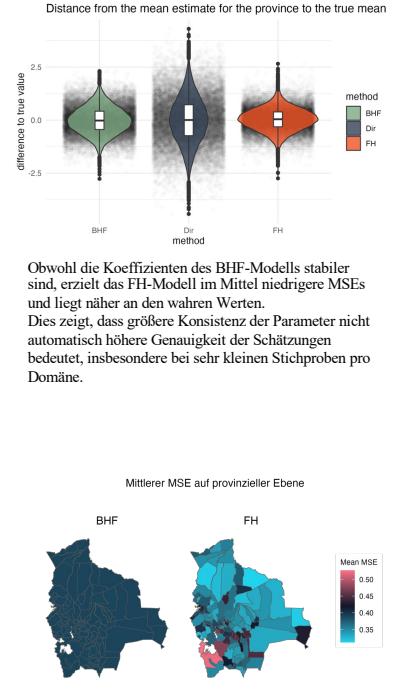
1) Motivation

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



3) Ergebnisse



Obwohl die Koeffizienten des BHF-Modells stabiler sind, erzielt das FH-Modell im Mittel niedrigere MSEs und liegt näher an den wahren Werten. Dies zeigt, dass größere Konsistenz der Parameter nicht automatisch höhere Genauigkeit der Schätzungen bedeutet, insbesondere bei sehr kleinen Stichproben pro Domäne.

4) Diskussion

Obwohl die Koeffizienten des BHF-Modells stabiler sind, erzielt das FH-Modell im Mittel niedrigere MSEs und liegt näher an den wahren Werten. Dies zeigt, dass größere Konsistenz der Parameter nicht automatisch höhere Genauigkeit der Schätzungen bedeutet, insbesondere bei sehr kleinen Stichproben pro Domäne. Der FH-Ansatz profitiert von der direkten Aggregation auf Domänenebene, während das BHF-Modell auf individuelle Daten angewiesen ist, die bei kleinen n stärker schwanken.

5) Limitationen

Die Verteilung der Residuen ist nicht optimal, und auch die gewählten Variablen weisen keine ideale Verteilung auf. Insbesondere bei korrelierten Variablen wäre es sinnvoll gewesen, zu prüfen, welche Variablen einen besseren Modell-Fit ermöglichen. Darüber hinaus würde ein aussagekräftiger Vergleich der Modelle erfordern, dass für jedes Modell eine separate Modellsuche durchgeführt wird, um die besten Prädiktoren zu identifizieren.

Die Verteilung der Residuen ist nicht optimal, und auch die gewählten Variablen weisen keine ideale Verteilung auf. Insbesondere bei korrelierten Variablen wäre es sinnvoll gewesen, zu prüfen, welche Variablen einen besseren Modell-Fit ermöglichen. Darüber hinaus würde ein aussagekräftiger Vergleich der Modelle erfordern, dass für jedes Modell eine separate Modellsuche durchgeführt wird, um die besten Prädiktoren zu identifizieren.

References

- Battese, G. E., Harter, R. M., & Fuller, W. A. (1988). An Error-Components Model for Prediction of County Crop Areas Using Satellite and Census Data.
- Ford, R. E., & Herren, R. A. (1979). Estimates of Income for Small Places: An Application of James-Stein Procedures to Census Data.
- Harmeling, S., Kreutzmann, A.-K., Schmidt, S., Salvati, N., & Schmid, T. (2023). A framework for producing small area estimates based on unit-level models in R.
- INE Bolivia. (2024). Census [Dataset].
- Kreutzmann, A.-K., Pannier, S., Rojas-Perilla, N., Schmid, T., Tempi, M., & Tzavidis, N. (2019). The R package emdi for estimating and mapping regularly disaggregated administrative boundaries. *Open Statistical Methods*, 1(1).
- Worr, N. (2023). *seeTris: Transformations for unit-level small area models* [Manual].