

Evaluating asymptotic intervals of FH estimates with and without variance smoothing, using simulated data mirroring Kenya height-for-age

Alana McGovern

2025-11-14

The goal of this simulation study is to evaluate how well variance smoothing can improve interval properties under the FH model. We simulate data meant to mirror the height-for-age z-score in the 2022 Kenya DHS.

Simulation Design

1. Obtain weighted stratified mean estimates $(\boldsymbol{\mu}_R, \boldsymbol{\mu}_U)$, at the desired level, using the DHS data (height-for-age)
2. Generate a continuous response, for each sampled individual k in cluster c , in strata h , in area i , from either

- a. no cluster effect:

$$y_{ihck}^{(s)} \sim \mathcal{N}(\mu_{ih}, \sigma_{ih}^{2(s)})$$

- b. cluster effect:

$$y_{ihck}^{(s)} \sim \mathcal{N}(\mu_{ih} + \epsilon_c^{(s)}, \sigma_{ih}^{2(s)})$$

where $\sigma_{ih}^{2(s)}$ is generated from either:

- a. IID variance with constant strata effect:

$$\log(\sigma_{iR}^{2(s)}) \sim \mathcal{N}(\alpha, 1/\tau)$$

$$\sigma_{iU}^{2(s)} = (1 + \kappa)\sigma_{iR}^{2(s)}$$

- b. Structured variance with constant strata effect:

$$\log(\sigma_i^{2(s)}) = \alpha + Z\gamma + v_i$$

$$\sigma_{iU}^{2(s)} = (1 + \kappa)\sigma_{iR}^{2(s)}$$

where Z is a covariate matrix, and v can be distributed either IID or BYM2.

- c. Structured variance with varying strata effect:

$$\log(\sigma_i^{2(s)}) = \alpha + Z\gamma + v_i$$

$$\sigma_{iU}^{2(s)} = (1 + \kappa_i)\sigma_{iR}^{2(s)}$$

3. Take a two-stage stratified sample, with sampling design and sizes chosen to mirror the Kenya 2022 DHS.

4. Under DGPs a.a and a.b, we fit the implied model, and for a.b we also fit the model with IID variance (to see how much it hurts to not include relevant information). For b.a and b.b, we fit the models implied by a.a and a.b, respectively, to see if not directly accounting for cluster effect affects model performance. For a.c we fit the model implied by a.b to test how much assuming a fixed strata effect in the variance hinders results. For all of these models we use relevant covariates and BYM2 for the mean model.
5. Under each DGP, we compare it to the oracle model (where sampling variance estimate is correct), standard Fay-Herriot (no smoothing) and the naive variance smoothing model with $df = \#clusters - \#strata$.