Stage 2 simulation results

Austin Schumacher

2022-03-04

Overview

We will perform a simulation study to explore the recovery of parameters

Data generating mechanism

Let $r = 1, \ldots, R$ index region and $c = 1, 2$ index outcomes. The region-level means will be generated as

$$
\begin{aligned}
\mu_{r1} &= \beta_1 + v_{r1} + u_{r1} + \lambda(v_{r2} + u_{r2}) \\
\mu_{r2} &= \beta_2 + v_{r2} + u_{r2} \\
v_{r1}|\sigma_1^2 &\overset{iid}{\sim} N(0, \sigma_1^2) \\
v_{r2}|\sigma_2^2 &\overset{iid}{\sim} N(0, \sigma_2^2) \\
\boldsymbol{u}_1 &\sim ICAR(1) \\
\boldsymbol{u}_2 &\sim ICAR(1)
\end{aligned}
$$

with β_c set to be the estimated intercept and σ_c^2 set to be the estimated variances of the random effects from a model fit to the 2014 KDHS HAZ and WAZ data. The random effects will only be simulated once and the same ones will be used across all simulations.

We will use a BYM2 parameterization for the spatial and IID random effects in each model such that, for $c = 1, 2$, we let

$$
v_{rc} + u_{rc} = \sigma_c(\sqrt{1 - \rho_c} v_{cr}^* + \sqrt{\rho_c} u_{cr}^*)
$$

where $v_{cr}^* \overset{iid}{\sim} N(0, 1)$ is the unstructured random effect with fixed standard deviation 1 and $\boldsymbol{u}_r^*$ is the ICAR model scaled so $\text{Var}(u_{cr}^*) \approx 1$, which is done by scaling the model so the geometric mean of these variances is 1, using the adjacency matrix to calculate the inverse precision of the ICAR model as in Riebler et al. (2016).

Once we have these μ_{rc} values, we will simulate the area-level sample means as

$$
\hat{\boldsymbol{y}}_r | \boldsymbol{\mu}_r, \boldsymbol{V}_r \sim N_2(\boldsymbol{\mu}_r, \boldsymbol{V}_r)
$$

where the $\boldsymbol{V}_r$ are set to be equal to the estimated asymptotic design-based covariance matrix of the area-level mean HAZ and WAZ from the 2014 KDHS data.

Now, for an SRS, the sampling distribution of the area-level sample covariance matrices is

$$(n_r - 1)\hat{\boldsymbol{V}}_r^{srs}|V_r, n_r \sim \text{Wishart}\left(\boldsymbol{V}_r, (n_r - 1)\right)$$

where n_r is the sample size in area r. The 2014 KDHS uses a stratified cluster design rather than an SRS, which has a different sampling variance than an SRS. The ratio of the variance for a statistic calculated using a specific survey design to the variance of that statistic calculated using an SRS of the same sample size is called the design effect,

$$d^2 = \frac{\hat{V}}{\hat{V}^{srs}}.$$

In the a typical DHS, the average design effect across all indicators is $d^2 = 1.5^2 = 2.25$ (taken from this answer on the DHS user forum: https://userforum.dhsprogram.com/index.php?t=msg&goto=3448&S=Google). Then, the effective sample size in the stratified design is approximately equal to $n_r^* = n_r/(d^2)$, since the sample variance is linear with respect to the sample size. Thus, we calculate n_r^* as the observed sample size in region r divided by 2.25 and use this to adjust for the survey design effect, giving us

$$\hat{\boldsymbol{V}}_r|\boldsymbol{V}_r, n_r^* \sim \frac{1}{(n_r^* - 1)}\text{Wishart}\left(\boldsymbol{V}_r, (n_r^* - 1)\right).$$

In each simulation, we will have the same values of μ_r and V_r, and we use these to simulate $\hat{\boldsymbol{y}}_r$ and $\hat{\boldsymbol{V}}_r$. We will treat the $\hat{\boldsymbol{y}}_r$ as pseudo-direct estimates and calculate asymptotic 95% confidence intervals as $\hat{y}_{rc} \pm z_{0.975}\sqrt{\hat{V}_{r,cc}}$ with $\hat{V}_{r,cc}$ the diagonal entries of the sampled covariance matrix.

We will also fit six different smoothing models to the pseudo-direct estimates to estimate the latent means in each region with corresponding 95% credible intervals. They will be:

1. Univariate IID
2. Univariate BYM
3. Bivariate nonshared IID
4. Bivariate nonshared BYM
5. Bivariate shared IID
6. Bivariate shared BYM

All data will be the same size as the Kenya 2014 DHS (47 regions). The DGM parameters will be set equal to those estimated from models fit to the Kenya 2014 DHS HAZ/WAZ data. We run 100 simulations.

Results

Table 1: Model:Direct estimates

parameter	bias	abs bias	relative abs bias	80% coverage	80% width	95% coverage	95% width
HAZ latent means	0	0.074	0.082	0.797	0.239	0.950	0.365
WAZ latent means	0	0.067	0.086	0.800	0.215	0.945	0.328

Table 2: Model:Univariate IID

parameter	bias	abs bias	relative abs bias	80% coverage	80% width	95% coverage	95% width
beta[1]	-0.004	0.011	0.012	1.000	0.110	1.000	0.171
beta[2]	-0.007	0.011	0.013	1.000	0.112	1.000	0.174
sigma[1]	0.091	0.091	0.492	0.000	0.084	0.000	0.130
sigma[2]	0.036	0.036	0.145	0.550	0.083	0.930	0.128
rho[1]	NA	NA	NA	NA	NA	NA	NA
rho[2]	NA	NA	NA	NA	NA	NA	NA
lambda	NA	NA	NA	NA	NA	NA	NA
HAZ latent means	0.008	0.075	0.083	0.785	0.225	0.940	0.344
WAZ latent means	0.001	0.065	0.084	0.786	0.205	0.944	0.314

Table 3: Model:Univariate BYM

parameter	bias	abs bias	relative abs bias	80% coverage	80% width	95% coverage	95% width
beta[1]	-0.006	0.012	0.012	0.980	0.067	1.000	0.108
beta[2]	-0.006	0.011	0.012	0.960	0.060	1.000	0.097
sigma[1]	0.092	0.092	0.499	0.000	0.091	0.000	0.141
sigma[2]	0.020	0.021	0.084	0.920	0.086	1.000	0.133
rho[1]	-0.148	0.148	0.171	0.970	0.530	1.000	0.723
rho[2]	-0.143	0.143	0.157	0.880	0.472	1.000	0.666
lambda	NA	NA	NA	NA	NA	NA	NA
HAZ latent means	0.007	0.073	0.081	0.778	0.221	0.942	0.339
WAZ latent means	0.003	0.063	0.082	0.786	0.201	0.947	0.308

Table 4: Model:Bivariate nonshared IID

parameter	bias	abs bias	relative abs bias	80% coverage	80% width	95% coverage	95% width
beta[1]	-0.002	0.011	0.011	1.000	0.105	1.000	0.162

parameter	bias	abs bias	relative abs bias	80% coverage	80% width	95% coverage	95% width
beta[2]	-0.003	0.010	0.011	1.000	0.107	1.000	0.165
sigma[1]	0.076	0.076	0.408	0.000	0.079	0.040	0.123
sigma[2]	0.020	0.021	0.084	0.840	0.079	1.000	0.122
rho[1]	NA	NA	NA	NA	NA	NA	NA
rho[2]	NA	NA	NA	NA	NA	NA	NA
lambda	NA	NA	NA	NA	NA	NA	NA
HAZ latent means	0.010	0.078	0.087	0.756	0.218	0.920	0.333
WAZ latent means	0.006	0.069	0.090	0.748	0.197	0.922	0.302

Table 5: Model:Bivariate nonshared BYM

parameter	bias	abs bias	relative abs bias	80% coverage	80% width	95% coverage	95% width
beta[1]	-0.002	0.011	0.011	0.990	0.061	1.000	0.098
beta[2]	-0.002	0.010	0.011	0.960	0.054	1.000	0.088
sigma[1]	0.072	0.072	0.390	0.000	0.084	0.090	0.130
sigma[2]	0.000	0.012	0.047	1.000	0.079	1.000	0.122
rho[1]	-0.116	0.116	0.134	0.980	0.496	1.000	0.692
rho[2]	-0.118	0.118	0.130	0.950	0.445	1.000	0.640
lambda	NA	NA	NA	NA	NA	NA	NA
HAZ latent means	0.010	0.076	0.086	0.745	0.211	0.914	0.323
WAZ latent means	0.006	0.067	0.088	0.740	0.191	0.916	0.292

Table 6: Model:Bivariate shared IID

parameter	bias	abs bias	relative abs bias	80% coverage	80% width	95% coverage	95% width
beta[1]	-0.002	0.011	0.011	1.000	0.110	1.000	0.171
beta[2]	-0.003	0.010	0.011	1.000	0.111	1.000	0.172
sigma[1]	0.013	0.014	0.074	0.940	0.060	1.000	0.093
sigma[2]	0.035	0.035	0.140	0.570	0.083	0.960	0.128
rho[1]	NA	NA	NA	NA	NA	NA	NA
rho[2]	NA	NA	NA	NA	NA	NA	NA
lambda	-0.119	0.119	0.152	0.670	0.293	1.000	0.448
HAZ latent means	0.011	0.075	0.083	0.777	0.224	0.939	0.342
WAZ latent means	0.006	0.067	0.086	0.777	0.204	0.940	0.311

Table 7: Model:Bivariate shared BYM

parameter	bias	abs bias	relative abs bias	80% coverage	80% width	95% coverage	95% width
beta[1]	-0.003	0.011	0.011	0.990	0.066	1.000	0.105
beta[2]	-0.003	0.010	0.011	0.960	0.059	1.000	0.096
sigma[1]	-0.003	0.008	0.044	1.000	0.059	1.000	0.092
sigma[2]	0.017	0.019	0.075	0.940	0.084	1.000	0.129
rho[1]	-0.165	0.165	0.190	0.910	0.502	0.990	0.687
rho[2]	-0.144	0.144	0.158	0.830	0.460	1.000	0.649
lambda	-0.008	0.043	0.054	1.000	0.300	1.000	0.459
HAZ latent means	0.010	0.073	0.081	0.777	0.220	0.941	0.336
WAZ latent means	0.006	0.064	0.083	0.778	0.199	0.943	0.305