Small Area Estimation: Advanced Area-Level Models

Jon Wakefield and Victoria Knutson

Department of Statistics,
University of British Columbia

Departments of Statistics and Biostatistics
University of Washington

CEMA SAE Short Course
November 10–12th, 2026

Outline

Variance Modification

Nested Spatial Models

Variance Modification

A Tweak on the Fay-Herriot Model

- The successful use of the Fay Herriot model hinges on the availability of a weighted estimator $\widehat{p}_i^{\mathrm{w}}$ and its standard error $\sqrt{\widehat{V}_i}$.
- For sparse data, $\widehat{p}_i^{\mathrm{w}}$ may be on the boundary and $\sqrt{\widehat{V}_i}$ undefined or equal to zero.
- In **?** an automatic variance modification procedure is developed.

Variance Modeling: What's the Issue?

The variance of the Hajék estimator is

$$\widehat{V}(\widehat{p}_i^{\text{HJ}}) = \sum_{h_2=1}^{2} \widehat{V}(\widehat{p}_{ih_1[i]h_2}^{\text{HJ}}),$$

where,

$$\widehat{V}(\widehat{p}_{ih_1[i]h_2}^{\text{HJ}}) = \frac{1}{v_{ih_1[i]\cdots}^2} \frac{n_{ih_1[i]h_2}}{n_{ih_1[i]h_2} - 1} \sum_{c \in S_{ih_1[i]h_2}} \left[v_{ih_1[i]h_2c\cdot} (\widehat{p}_{ih_1[i]h_2c}^{\text{HJ}} - \widehat{p}_i^{\text{HJ}}) - \frac{1}{n_{ih_1[i]h_2}} v_{ih_1[i]h_2\cdot\cdot} (\widehat{p}_{ih_1[i]h_2}^{\text{HJ}} - \widehat{p}_i^{\text{HJ}}) \right.$$

$$\widehat{p}_{ih_1[i]h_2}^{\text{HJ}} = \frac{\sum_{c \in S_{ih_1[i]h_2}} \sum_{k \in S_{ih_1[i]h_2c}} v_{ih_1[i]h_2ck} z_{ih_1[i]h_2ck}}{v_{ih_1[i]h_2\cdot\cdot}}, \qquad \widehat{p}_{ih_1[i]h_2c}^{\text{HJ}} = \frac{\sum_{k \in S_{ih_1[i]h_2c}} v_{ih_1[i]h_2ck} z_{ih_1[i]h_2ck}}{v_{ih_1[i]h_2c\cdot}}. \tag{1}$$

- When the data are sparse in an area, then the estimator will have uncertainty that is high, or not possible to estimate (because the variance formula breaks down).
- Formula (**??**) breaks down if all outcomes in an area are all 0 or are all 1, or there is a single cluster in the area.
- We have a variance fix based on a weighted beta distribution.
- **?** jointly model Hajék and variance estimators.

In the Beginning...

Since the dawn of time, Bayesians have been fixing binomials with $y = 0/n$, e.g., **?**, combining binomial likelihood with Be(2,2) prior.

Be(2,2) = Phantom data with $z^{\text{PH}} = 2$ successes out of $m^{\text{PH}} = 4$ which gives the augmented estimator,

$$\widehat{p}^{\text{AUG}} = \frac{z + z^{\text{PH}}}{m + m^{\text{PH}}}.$$

How to transfer over to the survey setting?

Consider the pseudo-likelihood,

$$L^{\text{PL}}(p) = \prod_{k \in S} \left[p^{y_k} (1-p)^{1-y_k} \right]^{w_k}.$$

Now define a pseudo prior, based on phantom observations and weights, y_k, w_k, $k \in S^{\text{PH}}$,

$$g^{\text{PH}}(p) = \prod_{k \in S^{\text{PH}}} \left[p^{y_k} (1-p)^{1-y_k} \right]^{w_k} [p(1-p)]^{-1}.$$

The Fix

This gives augmented posterior:

$$\pi^{\text{AUG}}(p) \propto \prod_{k \in S} \left[p^{y_k}(1-p)^{1-y_k} \right]^{w_k} \times \prod_{k \in S^{\text{PH}}} \left[p^{y_k}(1-p)^{1-y_k} \right]^{w_k}$$

Maximizing this augmented posterior gives estimator:

$$\widehat{p}^{\text{AUG}} = \frac{\sum_{k \in S} w_k y_k + \sum_{k \in S^{\text{PH}}} w_k y_k}{\sum_{k \in S} w_k + \sum_{k \in S^{\text{PH}}} w_k}.$$

The augmented set of observations, $k \in S^{\text{AUG}}$, can be used in estimators and variance formulas.

This is implemented as default for pre-modeled estimates at

https://sae4lmic.stat.uw.edu/

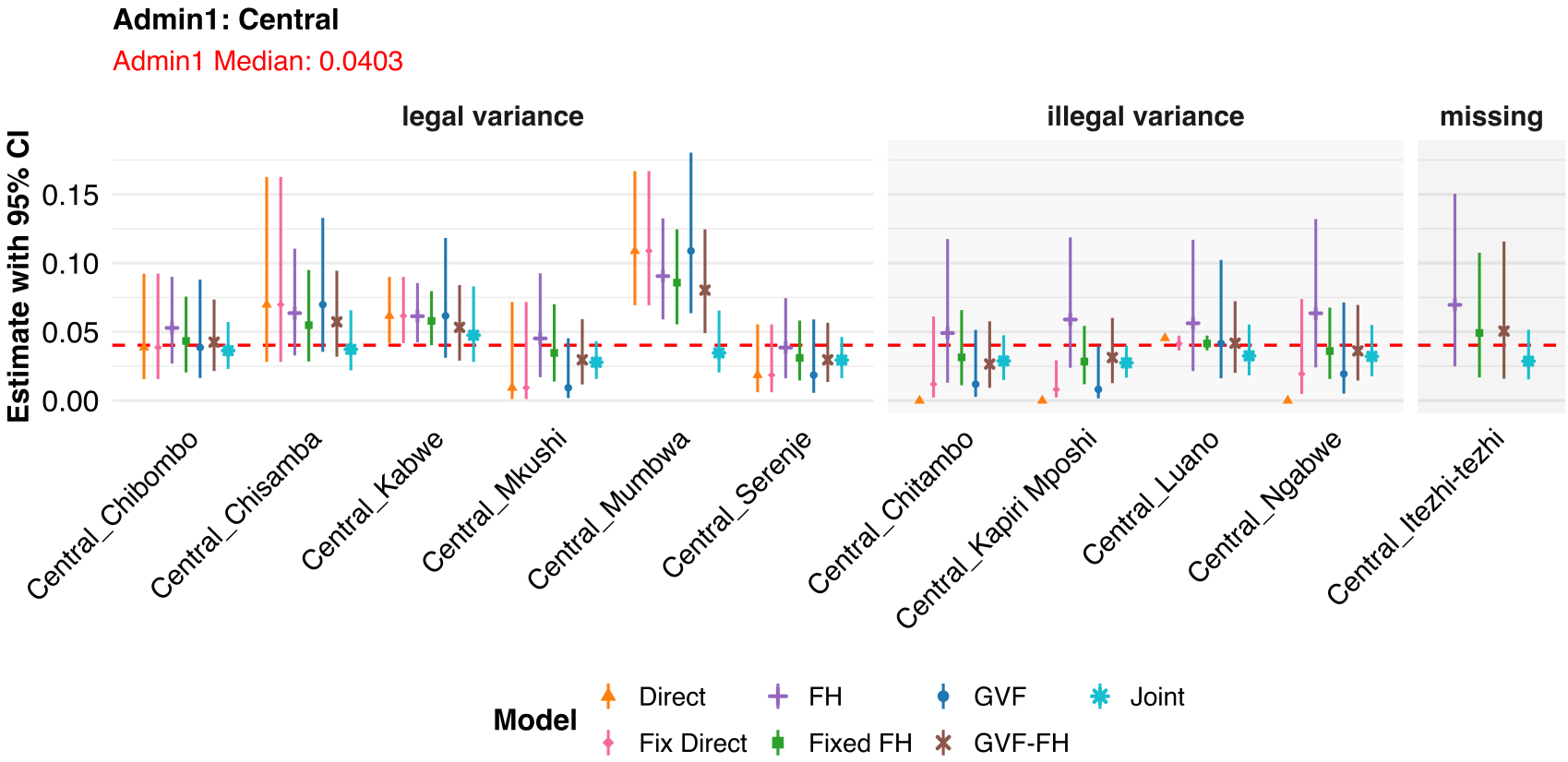


FIGURE 1: Empirical simulation coverage, mimicking DHS survey settings

Nested Spatial Models

Nested Unit-Level Model

Recall the area-level Fay-Herriot model:

$$\begin{aligned} Y_i &= \theta_i + \epsilon_i, \qquad \epsilon_i \sim_{ind} \mathsf{N}(0, \widehat{V}_i) \\ \theta_i &= \alpha + u_i, \end{aligned}$$

An Example with Nested Models

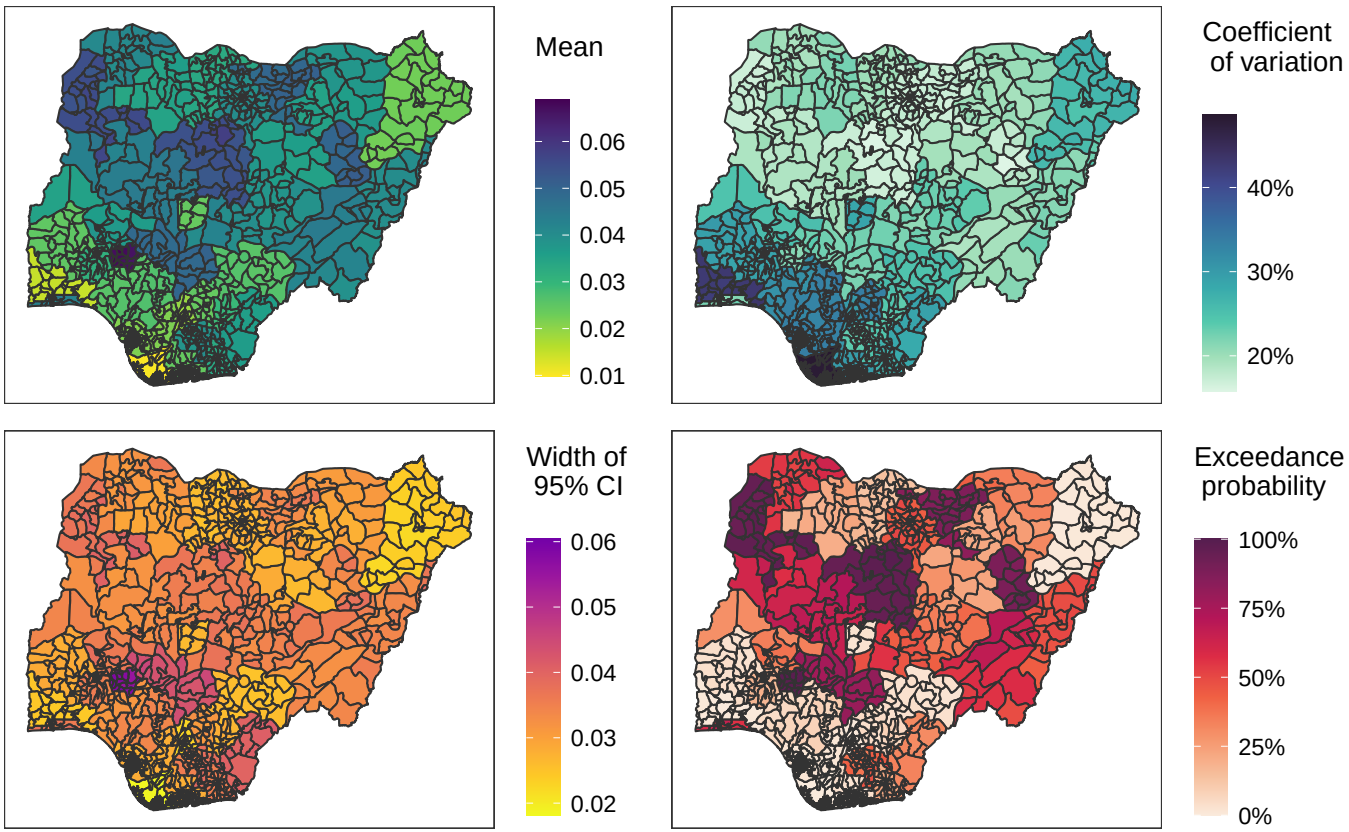


Figure 2: Nested discrete spatial model for Admin-1

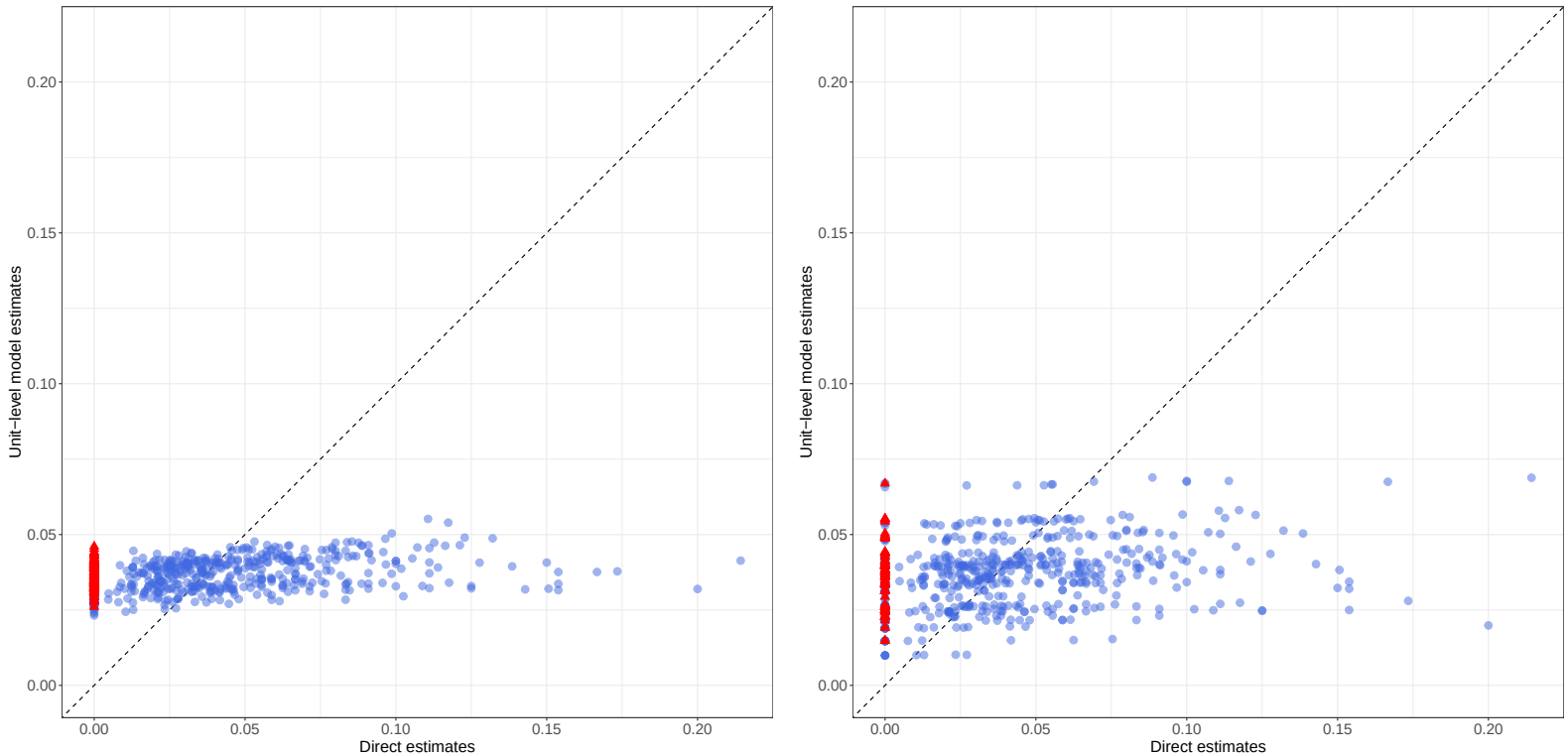


FIGURE 3: Estimates from non-nested (left) and nested (right) models, against weighted estimates. Notice the wider range in the right hand plot.