# Towards Comparative Statistics & Multiple Inference for Small Areas

**Stefan Sperlich – GSEM**

joint work with Katarzyna Reluga & Maria José Lombardía

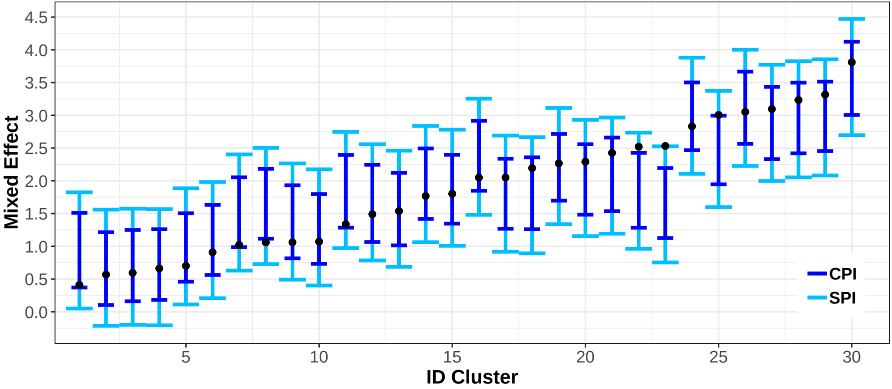
# Context

- classical SAE approach with (G)LMM
- Consider situations where mixed parameter is of interest
- not a new problem or type of data, nor a new model or estimator

# Problem

- want to allow for comparisons between areas
- but existing inference is *area-wise* (though <u>not</u> *area-specific*)
- When constructing 90% *classical* prediction intervals (CPI)
- by construction about 10% of area parameter are not in their CPI
- happens more likely to affect 'crucial' areas
- joint considerations and comparisons are invalid

# Bootstrap 95%CPI and SPI for $D = 30$ area means when $e_{dj} \sim N(0, 0.5)$, $u_d \sim N(0, 1)$, ... (REML)



# Modeling Framework - Notation

- Consider LMM with block diagonal covariance matrix (including NERM and FHM): $\boldsymbol{Y} = \boldsymbol{X}\beta + \boldsymbol{Z}\boldsymbol{u} + \boldsymbol{e}$
- Conditional covariance matrix of $\boldsymbol{Y}$: $\boldsymbol{V} = \boldsymbol{V}(\theta) = \boldsymbol{R} + \boldsymbol{Z}\boldsymbol{G}\boldsymbol{Z}^T$, $\boldsymbol{\theta}$ variance parameter, e.g. $\boldsymbol{\theta} = (\sigma_e^2, \sigma_u^2)$
- BLUP for $\beta$: $\tilde{\beta}(\boldsymbol{\theta}) = (\boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{Y}$
- BLUP for $\boldsymbol{u}$: $\tilde{\boldsymbol{u}}(\boldsymbol{\theta}) = \boldsymbol{G}\boldsymbol{Z}\boldsymbol{V}^{-1}(\boldsymbol{Y} - \boldsymbol{X}\tilde{\beta})$
- *Empirical* counterparts $\hat{\beta}$, $\hat{\boldsymbol{u}}$ with $\hat{\boldsymbol{\theta}}$ can be estimated by REML etc
- Consider mixed parameter,

  $\mu_d = \boldsymbol{k}_d^t\beta + \boldsymbol{m}_d^t\boldsymbol{u}_d$ with $\boldsymbol{k}_d \in \mathbb{R}^{p+1}$ and $\boldsymbol{m}_d \in \mathbb{R}^{q_d}$ known
- extensions possible for predictions like e.g. $\mu_d^{sm} = n_d^{-1}\left(\sum_{j=1}^{r_d} y_{dj} + \sum_{j=1}^{n_d-r_d} \boldsymbol{x}_{dj}^t\beta + \boldsymbol{z}_{dj}^t\boldsymbol{u}\right)$

# The MSE and its decompositions

- For constructing our PI use variance component of MSE

$$\mathrm{MSE}[\hat{\mu}_d] = \mathrm{MSE}[\tilde{\mu}_d] + \mathrm{E}[\hat{\mu}_d - \tilde{\mu}_d]^2 + 2 \cdot \mathbb{E}\left[(\tilde{\mu}_d - \mu_d)(\hat{\mu}_d - \tilde{\mu}_d)\right]$$

- First term accounts for variability of $\mu_d$ when $\boldsymbol{\theta}$ known. For $\boldsymbol{b}_d^t = \boldsymbol{k}_d^t - \boldsymbol{o}_d^t \boldsymbol{X}_d$ with $\boldsymbol{o}_d^t = \boldsymbol{m}_d^t \boldsymbol{G} \boldsymbol{Z}_d^t \boldsymbol{V}_d^{-1}$, it reduces to

$$\boldsymbol{m}_d^t(\boldsymbol{G}_d - \boldsymbol{G}_d \boldsymbol{Z}_d^t \boldsymbol{V}_d^{-1} \boldsymbol{Z}_d \boldsymbol{G}_d)\boldsymbol{m}_d + \boldsymbol{b}_d^t \left( \sum_{d=1}^{D} \boldsymbol{X}_d^t \boldsymbol{V}_d^{-1} \boldsymbol{X}_d \right)^{-1} \boldsymbol{b}_d$$
$$=: g_{1d}(\boldsymbol{\theta}) + g_{2d}(\boldsymbol{\theta})$$

- Second term is intractable. Third term disappears under normality of errors and random effects; it is rarely considered.
- Often PI construct using only $\boldsymbol{g}_1(\hat{\boldsymbol{\theta}}) = (g_{11}(\hat{\boldsymbol{\theta}}), \ldots, g_{1D}(\hat{\boldsymbol{\theta}}))^t$

# Constructing SPI with max-type statistics

Want $\mathcal{I}_{1-\alpha}$ such that $P(\mu_d \in \mathcal{I}_{1-\alpha}\ \forall d \in [D]) = 1-\alpha$, $[D] = \{1, \dots, D\}$.

Equivalent to finding critical value $c_{S_0}(1-\alpha)$ s.th.

$$\alpha = P\left(\max_{d=1,\dots,D}\left|\frac{\hat{\mu}_d - \mu_d}{\hat{\sigma}(\hat{\mu}_d)}\right| \geqslant c_{S_0}(1-\alpha)\right),$$

with $\hat{\sigma}^2(\hat{\mu}_d)$ estimated variability, e.g. $g_{1d}(\hat{\boldsymbol{\theta}})$

Clearly, $c_{S_0}(1-\alpha)$ is $(1-\alpha)^{th}$-quantile of

$$S_0 := \max_{d=1,\dots,D}|S_{0d}|, \text{ where } \quad S_{0d} = \frac{\hat{\mu}_d - \mu_d}{\hat{\sigma}(\hat{\mu}_d)}$$

Consequently,

$$\mathcal{I}^S_{1-\alpha} = \bigtimes_{d=1}^{D} \mathcal{I}^S_{d,1-\alpha}, \quad \text{where} \quad \mathcal{I}^S_{d,1-\alpha} = \{\hat{\mu}_d \pm c_{S_0}(1-\alpha)\hat{\sigma}(\hat{\mu}_d)\},$$

Pdf of $S_0$ is right-skewed; consider upper quantile for symmetric $\mathcal{I}^S_{d,1-\alpha}$

# Bootstrap approximation of quantile $c_{S_0}$

1. Obtain consistent estimators $\hat{\beta}$ and $\hat{\boldsymbol{\theta}} = (\hat{\sigma}_e^2, \hat{\sigma}_u^2)$.
2. Generate $D$ iid $W_1 \sim N(0,1)$ and set $u_d^* = \hat{\sigma}_u W_1$, $d = [D]$.
3. Generate $n$ iid $W_2 \sim N(0,1)$ and set $e_j^* = \hat{\sigma}_e W_2$, $j = [n]$.
4. Create sample $\boldsymbol{y}^* = \boldsymbol{X}\hat{\beta} + \boldsymbol{u}^* + \boldsymbol{e}^*$.
5. Fit model to obtain $\hat{\beta}^*$, $\hat{\boldsymbol{\theta}}^* = (\hat{\sigma}_e^{2*}, \hat{\sigma}_u^{2*})$, $\mu_{dj}^*$, $\hat{\mu}_{dj}^*$.
6. Repeat Steps 2-5 $B$ times.
7. Get quantile $c_{BS}(1-\alpha)$ and $\mathcal{I}_{1-\alpha}^{BS}$ of $S_B^{*(b)}$ using $\hat{\sigma}^{*(b)}(\hat{\mu}_d^{*(b)})$

Under FHM, $\hat{\boldsymbol{\theta}} = \hat{\sigma}_u^2$, $\hat{\sigma}^2(\hat{\mu}_d)$ accordingly

1. Generate $D$ independent copies of a variable $W_2 \sim N(0,1)$. Construct vector $\boldsymbol{e}^* = (e_1^*, e_2^*, \ldots, e_D^*)$ with elements $e_d^* = \sigma_{e_d} W_2$, $d = [D]$.

## Proposition

*Under CLL and some regularity conditions R.1-R.7 it holds*

$$\sup_{q\in\mathbb{R}} |P^*(S_B^* \leqslant q) - P(S_0 \leqslant q)| = o_P(1).$$

*and consequently it holds that for the bootstrap SPI, denoted* $\mathcal{I}_{1-\alpha}^{BS}$,

$$P\left(\mu_d \in \mathcal{I}_{1-\alpha}^{BS}\ \forall d \in [D]\right) \xrightarrow{D\to\infty} 1-\alpha.$$

## Corollary

*For* $D' < D$ *but* $O(D)$, $S_{B'}^{*(b)} := \max_{d=1,\dots,D'} \left|S_{Bd}^{*(b)}\right|$ *with quantile* $c_{B'S}(1-\alpha)$, *and* $\mathcal{I}_{1-\alpha}^{B'S} = \times_{d=1}^{D'} \{\hat{\mu}_d \pm c_{B'S}(1-\alpha)\hat{\sigma}(\hat{\mu}_d)\}$ *its SPI,*

$$P\left(\mu_d \in \mathcal{I}_{1-\alpha}^{B'S}\ \forall d \in [D']\right) \xrightarrow{D'\to\infty} 1-\alpha.$$

# Simultaneous Testing

For the multiple linear hypothesis

$$H_0 : \boldsymbol{A}\boldsymbol{\mu} = \boldsymbol{h} \quad vs. \quad H_1 : \boldsymbol{A}\boldsymbol{\mu} \neq \boldsymbol{h}, \qquad \boldsymbol{A} \in \mathbb{R}^{D' \times D}, \boldsymbol{h} \in \mathbb{R}^{D'} \text{ given}$$

Define for $\boldsymbol{\mu}^H = \boldsymbol{A}\boldsymbol{\mu}$ statistics

$$t_H := \max_{d=1,\dots,D'} |t_{H_d}|, \; S_{H_0} := \max_{d=1,\dots,D} |S_{H_0 d}|,$$

where

$$t_{H_d} = \frac{\hat{\mu}_d^H - h_d}{\hat{\sigma}(\hat{\mu}_d^H)} \qquad \text{and} \qquad S_{H_0 d} = \frac{\hat{\mu}_d^H - \mu_d^H}{\hat{\sigma}(\hat{\mu}_d^H)}$$

Test $t_H$ rejects $H_0$ at the $\alpha$-level if $t_H \geqslant c_{H_0}(1-\alpha)$, quantile of $S_{H_0}$

I.e., $\boldsymbol{h} \notin \mathcal{I}_{1-\alpha}^{H_0}$ with $\mathcal{I}_{1-\alpha}^{H_0} = \times_{d=1}^{D} \left\{ \hat{\mu}_d^H \pm c_{H_0}(1-\alpha)\hat{\sigma}(\hat{\mu}_d^H) \right\}$

Bootstrap approximation possible without generating samples from $H_0$ !

# Alternative Approaches

- **Volume of Tube Formula** following Sun and Loader (1994) AoS, but has oo many unknown terms
- **Beran's procedure** following Beran (1988) JASA, for *balanced simultaneous confidence sets*. But does not work with random effects.
- **Bonferroni procedure** as 'benchmark', but hinges on independence and quality of CPI
- **Monte-Carlo procedure**: popular in nonparametric spline regression; can be interpreted as Volume-of-Tube simplification

First is infeasible, others depend on MSE estimates, normality, normality approximations, etc.

# Monte-Carlo Procedure

Is based on approximation

$$\begin{bmatrix} \hat{\beta} - \beta \\ \hat{\boldsymbol{u}} - \boldsymbol{u} \end{bmatrix} \approx N\left\{0, \left(\boldsymbol{C}^t \hat{\boldsymbol{R}}^{-1} \boldsymbol{C} + \hat{\boldsymbol{G}}^+\right)^{-1}\right\}.$$

Recognize for $\bar{\boldsymbol{c}}_d = (\boldsymbol{k}_d^t, \boldsymbol{m}_d^t)^t$

$$S_0 = \max_{d=1,\dots,D} |S_{0d}| \approx \max_{d=1,\dots,D} \frac{\left| \bar{\boldsymbol{c}}_d^t \begin{bmatrix} \hat{\beta} - \beta \\ \hat{\boldsymbol{u}}_d - \boldsymbol{u}_d \end{bmatrix} \right|}{\hat{\sigma}(\hat{\mu}_d)} =: \max_{d=1,\dots,D} |S_{MCd}| = S_{MC}$$

Draw $K$ realisations from normal distribution above to
Estimate $S_{MC}$ quantile $c_{S_0}(1-\alpha)$, and construct

$$\mathcal{I}_{1-\alpha}^{MC} = \bigtimes_{d=1}^{D} \mathcal{I}_{d,1-\alpha}^{MC} \quad \text{where} \quad \mathcal{I}_{d,1-\alpha}^{MC} = \{\hat{\mu}_d \pm c_{MC}(1-\alpha)\hat{\sigma}(\hat{\mu}_d)\}$$

# The Simulation Design

- For the NERM: $X = U[0,1]$, $\beta = (1,1)$, $n_d = 5, 10, 15$, $D = 25, 50, 75$, different $(u, e)$
- For FHM: $\sigma_u^2 = 1$, $\sigma_{e_d}^2$ known, with 5 equally large groups with 0.7, 0.6, 0.5, 0.4, 0.3
- $B = 1000$ bootstrap samples with $I = 2500$ simulation runs
- Tried many other estimators for the variance and MSE of $\hat{\mu}$
- Construct 95% SPI for all (feasible) methods
- **Criteria**: empirical coverage and average width

$$\mathrm{EC} = \frac{1}{I}\sum_{k=1}^{I} 1\{\mu_d^{(k)} \in \mathcal{I}_{1-\alpha}^{P}\ \forall d \in [D]\}, \qquad \text{where P} = \text{BS, MC, BE, BO}$$

$$\mathrm{WS} = \frac{1}{DI}\sum_{d=1}^{D}\sum_{k=1}^{I} \rho_d^{(k)}, \qquad \rho_d^{(k)} = 2c_P^{(k)}(1-\alpha)\hat{\sigma}^{(k)}(\hat{\mu}_d),$$

Table: 95% SPI under the FHM for increasing $D$

| | EC (in %) | | | | WS | | | |
|---|---|---|---|---|---|---|---|---|
| D | BS | MC | BE | BO | BS | MC | BE | BO |
| 15 | 97.3 | 95.6 | 96.7 | 96.5 | 3.728 | 3.516 | 3.672 | 3.691 |
| 30 | 96.6 | 95.2 | 94.8 | 96.6 | 3.792 | 3.664 | 3.688 | 3.818 |
| 60 | 95.7 | 92.6 | 89.6 | 93.9 | 3.973 | 3.804 | 3.760 | 3.873 |
| 90 | 95.2 | 93.3 | 84.0 | 94.4 | 4.024 | 3.920 | 3.694 | 3.970 |

Table: 95% SPI under the NERM for different scenarios

| | | EC (in %) | | | | WS | | | |
|---|---|---|---|---|---|---|---|---|---|
| | $D : n_d$ | BS | MC | BE | BO | BS | MC | BE | BO |
| $\sigma_e^2 = 0.5$, $\sigma_u^2 = 1$ | 15:5 | 95.4 | 92.9 | 93.6 | 93.8 | 1.876 | 1.754 | 1.803 | 1.794 |
| | 30:5 | 95.2 | 93.9 | 92.5 | 94.4 | 1.947 | 1.890 | 1.871 | 1.910 |
| | 60:5 | 94.9 | 93.7 | 88.7 | 94.2 | 2.041 | 2.011 | 1.936 | 2.023 |
| | 90:5 | 95.2 | 94.4 | 84.4 | 94.9 | 2.101 | 2.079 | 1.926 | 2.088 |
| $\sigma_e^2 = 1$, $\sigma_u^2 = 1$ | 15:5 | 96.7 | 91.2 | 93.8 | 94.4 | 2.695 | 2.358 | 2.488 | 2.488 |
| | 30:5 | 95.5 | 92.8 | 92.5 | 94.4 | 2.671 | 2.552 | 2.567 | 2.608 |
| | 60:5 | 95.0 | 93.7 | 89.1 | 94.5 | 2.774 | 2.719 | 2.631 | 2.750 |
| | 90:5 | 95.2 | 94.2 | 83.2 | 94.8 | 2.850 | 2.811 | 2.614 | 2.833 |
| $\sigma_e^2 = 1$, $\sigma_u^2 = 0.5$ | 15:5 | 98.3 | 87.3 | 92.3 | 96.5 | 2.816 | 2.156 | 2.362 | 2.488 |
| | 30:5 | 97.3 | 90.6 | 94.2 | 94.8 | 2.641 | 2.346 | 2.469 | 2.485 |
| | 60:5 | 95.3 | 92.7 | 89.7 | 94.5 | 2.616 | 2.513 | 2.478 | 2.577 |
| | 90:5 | 95.0 | 93.0 | 83.9 | 94.6 | 2.663 | 2.597 | 2.441 | 2.643 |

Table: 95% SPI under the NERM for a subset of $D' < D$ Small Areas

| | $D : n_d$ | $D'$ | EC (in %) BS | EC (in %) MC | EC (in %) BE | EC (in %) BO | WS BS | WS MC | WS BE | WS BO |
|---|---|---|---|---|---|---|---|---|---|---|
| | 15:5 | 3 | 95.3 | 94.9 | 91.4 | 95.1 | 2.006 | 2.006 | 1.922 | 2.033 |
| $\sigma_e^2 = 1$ | 30:5 | 6 | 95.9 | 95.6 | 94.7 | 95.7 | 2.175 | 2.175 | 2.152 | 2.187 |
| $\sigma_u^2 = 1$ | 60:5 | 12 | 96.5 | 96.1 | 94.6 | 96.1 | 2.350 | 2.350 | 2.280 | 2.358 |
| | 90:5 | 18 | 95.0 | 94.6 | 90.6 | 94.6 | 2.455 | 2.455 | 2.317 | 2.457 |

Remarks:

- Simulations for our Test lead to same conclusions
- Notice that the scenarios represented optimal conditions for Benferroni's procedure (BO); we constructed optimal CPI and had almost independent estimators $\hat{\mu}_d$.

# Extensions, References

- consider SPI for GLMM under exponential distributions
  ( Reluga, Lombardia, Sperlich (2021) in JASA )
- consider deviations from normality in $u$ and $e$ respectively
  ( Reluga, Sperlich (2021), preliminary results in early working papers )
- study deviations from independence, in particular space- or time-dependence ( to be done )
- multiple inference in LMM via Quadratic forms with Chi-square stat's
  ( Kramlinger, Krivibokova, Sperlich (2021) 2nd round JASA )
- simultaneous / uniform inference for other SAE techniques than LMM
  ( FNS project 200021-192345, period 2021-2025 )