Empirical Small Area Prediction of Sheet and Rill Erosion Using a Zero-inflated Lognormal Model

Introduction
- Small area estimation widely used when sample sizes too small for direct estimation.
- Skewed data w/ zeros: Conservation Emissions Assessment Project Sheet and rill erosion (RUSLE2) data in South Dakota contains about 15% zeros.
- Small area predictors and MSE estimators for a lognormal model have closed-form expressions (Berg and Chandra, 2014). Can we extend this to a zero-inflated model?
- How does empirical Bayes compare to the plug-in predictor (Chandra and Chambers, 2016)?

Zero-Inflated Lognormal Model
- Let $i = 1,...,D$ index areas, $j = 1,...,N_i$ index units in each area.
- Variable of interest: $y_{ij} = y^*_ijδ_j + ε_j$ where $δ_j = 0$ if observed value is positive, 0 otherwise.
- $y^*_ij$ and $ε_j$ follow lognormal distribution.
- Positive part: $y^*_ij \sim Bernoulli(p_i), ε_j \sim lognormal$, $g(·)$ is a parametric link function. 
  
- Binary part: $δ_j ∼ Bernoulli(p_j), ε_j \sim lognormal$
  
- Model assessment:
  - Backward variable selection applied to the fixed effects with a threshold of .1.
  - For the binary part, the Hosmer-Lemeshow test shows no significant lack of fit.

Small Area Prediction
- Use empirical Bayes method to predict population means at small area level.
- For MSE estimator, we propose:
  - Use empirical Bayes method to predict population means at small area level.
- Possible explanatory variables related to the USLE:
  - Positive part: logS, loglogR, logK, logR, logS, logF, landuse, lat, long
  - Binary part: logR, logK, logS, logF, landuse, lat, long

Conservation Effects Assessment Project (CEAP)
- Response variable $y^*$: sheet and rill erosion, as measured by the Revised Universal Soil Loss Equation (RUSLE2), an update of a model for sheet and rill erosion called USLE.
- Possible explanatory variables related to the USLE:

Simulation Results
- Relative MSE of PI/ZI predictor to EB predictor:
  - Model: EB predictor
  - Simulation study on the proposed one-step MSE estimator

Summary
- We developed EB predictors based on a zero-inflated lognormal for SAE:
  - EB and plug-in have similar efficiency, unless data extremely sparse.
  - For D = 60, the “one-step” MSE estimator is a reasonable approximation.
  - For D = 30, the bootstrap MSE estimator: RB 2%~3%, CP 94%~96%.
  - EB predictor is typically more efficient than direct estimators in terms of MSE in CEAP

References: