Small area model selection

Snigdhansu Chatterjee

School of Statistics, University of Minnesota

joint work with several colleagues and students, including:
Megan Heyman, Lindsey Dietz, Taps Maiti, Hao Ren . . ., also
very much work in progress!

June 19, 2016

1. **What is a small area model?**
2. **The Fay-Herriot model**
3. **FH model selection: problems and conventional approach**
4. **Small area model selection:The Wild Scale Enhanced Bootstrap**
5. **Some preliminary results**

Outline

A Simple Two-level Model

Efron and Morris (JASA, 1975)

For $i = 1, \ldots, n$,

$$\text{Level 1: (Sampling Distribution)} \quad Y_i | \theta_i \overset{ind}{\sim} N(\theta_i, 1),$$
$$\text{Level 2: (Prior Distribution)} \quad \theta_i \overset{ind}{\sim} N(\mu, \psi).$$

The above model can be also viewed as a simple linear mixed model:

$$Y_i = \mu + U_i + E_i,$$

where $\{U_i\}$ and $\{E_i\}$ are independent with $U_i \overset{iid}{\sim} N(0, \psi)$ and $E_i \overset{iid}{\sim} N(0, 1)$ $i = 1, \cdots, n$.

A bit more useful version

For $i = 1, \ldots, n$,

$$\text{Level 1:} \quad Y_i = h_{1i}^{-1}(\tilde{Y}_i)|\theta_i \overset{ind}{\sim} N(\theta_i, D_i),$$

$$\text{Level 2:} \quad \theta_i \overset{ind}{\sim} N(x_i'\beta, \psi),$$

where

- $\tilde{Y}_i$: *direct estimator* of small area parameter of interest $h_{1i}(\theta_i)$ (mean, total, proportion),
- $h_{1i}(Y_i)$ is a one-to-one measurable function of Y_i (we assume this is known in this talk),
- D_i is the sampling variance of Y_i usually approximated or/and estimated,
- x_i^T: a $p \times 1$ column vector of known auxiliary variables,
- (β, ψ) are hyperparameters

1. **SAIPE State level model for poverty rate:** $Y_i = \tilde{Y}_i$; D_i are estimated by a replication-based method.
2. **Firm Alarm Probabilities (Carter and Rolph 1974):** $Y_i = \arcsin(\sqrt{\tilde{Y}_i})$, where n_i is the sample size for area i; $D_i = \frac{1}{4n_i}$. In Chilean poverty mapping, similar transformation is used with n_i representing effective sample size to incorporate complex sample design.
3. **Baseball Data Analysis (Efron and Morris 1975):** $Y_i = \sqrt{n_i} \arcsin(2\tilde{Y}_i - 1)$, where n_i is the sample size for area i; $D_i = 1$.
4. **Per-capita income (Fay and Herriot 1979):** $Y_{1i} = \log(\tilde{Y}_i)$; $D_i = 9/N_i$, where N_i is the population size.

A famous small area model

The Fay-Herriot (1979) model

Level I: (Sampling scheme) $Y_i|\theta_i \sim N(\theta_i, D_i), \; i = 1, \ldots, n.$

Level II: (Small area model) $\theta_i \sim N(\mathbf{x}_i^T\beta, \psi), \; i = 1, \ldots, n.$

- The D_i's known.
- The unknown hyperparameters are $\xi = (\beta, \psi)$.
- The covariates: $\mathbf{x}_i \in \mathbb{R}^p, \; i = 1, \ldots, n$ are fixed or random.
- *The objects of interest are:* $\theta_i \in \mathbb{R}, \; i = 1, \ldots, n.$

The general model:

$$\begin{aligned} &\text{Level I: (Sampling scheme) } \mathbf{Y}_{ij}|\boldsymbol{\theta}_i \sim f_{ij}(\cdot;\boldsymbol{\theta}_i,\xi_f),\ j=1,\ldots,n_i;\\ &\text{Level II: (Small area model) } \quad \boldsymbol{\theta}_i \sim g_i(\cdot;\xi_g),\ i=1,\ldots,n. \end{aligned}$$

- Statistical analysis based on the observed data gains in accuracy and precision by utilizing the commonality between the small areas, which is captured in the **Level II** model.
- The unknown parameters are $\xi = (\xi_f, \xi_g)$; these may be *regression coefficients*, *variance components*, and so on.
- The objects of interest are the $\boldsymbol{\theta}_i$'s which are $O_p(1)$ random variables in any paradigm.
- The model should be interpreted based on the philosophy: *All (statistical) models are wrong but some are useful.* (G. E. P. Box)

Example

US poverty data

American Community Survey (ACS) data obtained from large-sized counties from each state, along with several covariates, census and other figures.

```
> colnames(ACSData)
 [1] "ST"            "CTY"            "acspop"
 [4] "acspopse"      "acspov"         "acspovse"
 [7] "acspovrt"      "acspovrtse"     "acspop017"
[10] "acspop017se"   "acspov017"      "acspov017se"
[13] "acspovrt017"   "acspovrt017se"  "cenpop"
[16] "cenpov"        "cenpov017"      "cenpovrt"
[19] "cenpovrt017"   "cenpop017"      "foodstamps"
[22] "dempop"        "dempop017"      "foodstamprt"
[25] "......."            "....."           "......"
```

Which of these are useful covariates?

Outline

The Fay-Herriot model

$$[\mathbf{Y}|\boldsymbol{\theta}] \sim N_n(\boldsymbol{\theta}, \mathbf{D} = diag(D_1, \ldots, D_n)),$$
$$\boldsymbol{\theta} \sim N_n(\mathbf{X}\beta, \psi \mathbf{I}_n).$$

The D_i's known, unknown hyperparameters are $\xi = (\beta, \psi)$.

- The target conditional (posterior) distribution:

$$\pi(\theta_i|\mathbf{Y}_i, \xi) \propto f_i(\mathbf{Y}_i; \theta_i, \xi) g_i(\theta_i; \xi)$$

  in the general model. This depends on the unknown parameter ξ.
- In the Fay-Herriot model, the posterior distribution of $\boldsymbol{\theta}$ given $\mathbf{y}$ is

$$\begin{aligned} (\boldsymbol{\theta}|\mathbf{Y}) &\sim N_n\left((\mathbf{I}_n - B)\mathbf{Y} + B\mathbf{X}\beta, (\mathbf{I}_n - B)D\right) \text{ where} \\ B &= diag\left(B_i = D_i/(D_i + \psi), i = 1, \ldots, n\right). \end{aligned}$$

The Fay-Herriot model target distribution

$$[\theta|\mathbf{Y}] \sim N_n\left((\mathbf{I}_n - B)\mathbf{Y} + B\mathbf{X}\beta, (\mathbf{I}_n - B)D\right).$$

- The best predictor (BP) is the conditional mean $\int t\pi(t|Y_i, \xi)dt$. In the Fay-Herriot model, this is the *Best Linear Unbiased Predictor* (BLUP) $\tilde{\theta} = (\mathbf{I}_n - B)\mathbf{Y} + B\mathbf{X}\beta$.
- The empirical best predictor is the conditional mean $\int t\pi(t|Y_i, \hat{\xi})dt$, for a suitable estimator $\hat{\xi}$. In the Fay-Herriot model, this is the *Empirical Best Linear Unbiased Predictor* (EBLUP) $\hat{\theta} = (\mathbf{I}_n - \hat{B})\mathbf{Y} + \hat{B}\mathbf{X}\hat{\beta}$.

The Fay-Herriot model

The Fay-Herriot model target distribution

$$[\boldsymbol{\theta}|\mathbf{Y}] \sim N_n\left((\mathbf{I}_n - B)\mathbf{Y} + B\mathbf{X}\beta, (\mathbf{I}_n - B)D\right).$$

- The best predictor (BP) is the conditional mean $\int t\pi(t|Y_i, \xi)dt$. In the Fay-Herriot model, this is the *Best Linear Unbiased Predictor* (BLUP) $\tilde{\boldsymbol{\theta}} = (\mathbf{I}_n - B)\mathbf{Y} + B\mathbf{X}\beta$.
- The empirical best predictor is the conditional mean $\int t\pi(t|Y_i, \hat{\xi})dt$, for a suitable estimator $\hat{\xi}$. In the Fay-Herriot model, this is the *Empirical Best Linear Unbiased Predictor* (EBLUP) $\hat{\boldsymbol{\theta}} = (\mathbf{I}_n - \hat{B})\mathbf{Y} + \hat{B}\mathbf{X}\hat{\beta}$.
- Recall Box's statement: *All models are wrong, some are useful*! In order to be useful in practice, EBLUPs often have to be transformed, Winsorized or otherwise robustified, and benchmarked.

The Fay-Herriot model

$$[\mathbf{Y}|\theta] \sim N_n(\theta, \mathbf{D} = diag(D_1, \ldots, D_n)), \quad \theta_i \sim N_n(\mathbf{X}\beta, \psi \mathbf{I}_n).$$

The D_i's known, unknown hyperparameters are $\xi = (\beta, \psi)$.

- Parameters $\xi = (\beta, \psi)$ are estimated from the marginal $[\mathbf{Y}]$.
- The Prasad-Rao estimation scheme obtains the ordinary least squares estimator for β and a moment-based estimator for ψ. Thus $\hat{\beta}_{PR} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y}$ and $\hat{\psi}_{PR} = \max\left\{(n-p)^{-1}\left(||\mathbf{Y} - \mathbf{X}\hat{\beta}||^2 - \sum(1 - H_{ii})D_i\right), \varepsilon\right\}$.
- The "hat matrix" is denoted by H, identity matrix is $\mathbf{I}$.

The Fay-Herriot model

$$[\mathbf{Y}|\theta] \sim N_n(\theta, \mathbf{D} = diag(D_1, \ldots, D_n)), \quad \theta_i \sim N_n(\mathbf{X}\beta, \psi\mathbf{I}_n).$$

The D_i's known, unknown hyperparameters are $\xi = (\beta, \psi)$.

- The Fay-Herriot estimation scheme minimizes a contrast function for ψ, and obtains a weighted least squares estimator of β. This is often better than Prasad-Rao estimator.
- Maximum likelihood, reduced maximum likelihood, robust, Bayes estimates, and many other estimators have been studied.

The Fay-Herriot model target distribution

$$[\theta|\mathbf{Y}] \sim N_n\left((\mathbf{I}_n - B)\mathbf{Y} + B\mathbf{X}\beta, (\mathbf{I}_n - B)D\right).$$

$$\begin{aligned}
\hat{\beta}_{PR} &= (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y}, \\
\hat{\psi}_{PR} &\doteq (n-p)^{-1}\left(||\mathbf{Y} - \mathbf{X}\hat{\beta}||^2 - \sum(1 - H_{ii})D_i\right), \\
\hat{B}_i &= D_i/(D_i + \hat{\psi}), i = 1, \ldots, n, \\
\textbf{EBLUP}: \quad \hat{\theta}_i &= (1 - \hat{B}_i)Y_i + \hat{B}_i\mathbf{x}_i^T\hat{\beta} = Y_i - \hat{B}_i(Y_i - \mathbf{x}_i^T\hat{\beta}) \\
&= \left[D_i + \hat{\psi}\right]^{-1}\left\{\hat{\psi}Y_i + D_i\mathbf{x}_i^T\hat{\beta}\right\}, \\
\mathbb{V}[\theta_i|\mathbf{Y}] &= \left[D_i + \hat{\psi}\right]^{-1}\left\{D_i\,\hat{\psi}\right\}.
\end{aligned}$$

The Mean Squared Prediction Error (MSPE)

Target distribution and EBLUP

$$\theta_i|\mathbf{Y} \sim N\Big((1-B_i)Y_i + B_i\mathbf{x}_i^T\beta, (1-B_i)D_i\Big),$$

$$\textbf{EBLUP}: \quad \hat{\theta}_i = (1-\hat{B}_i)Y_i + \hat{B}_i\mathbf{x}_i^T\hat{\beta}.$$

- The Mean Squared Prediction Error (**MSPE**): For any predictor T_i of θ_i, the MSPE is $MSPE(T_i) = \mathbb{E}\left[\theta_i - T_i\right]^2$.
- $MSPE(T_i) = \mathbb{E}\left(\mathbb{V}(\theta_i|\mathbf{Y})\right) + \mathbb{E}\left(T_i - \mathbb{E}(\theta_i|\mathbf{Y})\right)^2 =$ $\mathbb{E}(\text{ Condl. Var}) + \mathbb{E}(\text{ Estimation gap}) = O(1) + O(n^{-1})$ typically, hence any estimator of *MSPE* should achieve $o(n^{-1})$ accuracy at least.
- For general predictor T_i, this can be extremely hard to compute (up to $o(n^{-1})$ accuracy).

Outline

Why bother with model selection in small areas?

- In most statistical model selection problems, over-fitting is less of an issue compared to under-fitting. Also, under-fitting is easier to detect. *This is not the case for mixed effects models like the Fay-Herriot model.*
- Under-fitting does not necessarily result in noticeable lack of goodness-of-fit in $[\theta|\mathbf{Y}]$.
- *Over-fitting is extremely dangerous.* It results in more weight on the "prior", *i.e.* on the auxiliary information rather than the observed data. The conditional variance is low, so we are very confident about our wrong predictor.
- The FH parameters do not have the same interpretation as classical parametric models (Bayesian or frequentist) like regression models.

The Fay-Herriot model targets

$$\hat{\beta}_{PR} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y},$$

$$\hat{\psi}_{PR} \doteq (n-p)^{-1}\left(||\mathbf{Y}-\mathbf{X}\hat{\beta}||^2 - \sum(1-H_{ii})D_i\right),$$

$$\hat{\theta}_i = \left(\hat{\psi}Y_i + D_i\mathbf{x}_i^T\hat{\beta}\right) / \left(D_i + \hat{\psi}\right), \quad \mathbb{V}[\theta_i|\mathbf{Y}] = \left\{D_i\,\hat{\psi}\right\} / \left(D_i + \hat{\psi}\right).$$

- *Not using all the required covariates*: $||\mathbf{Y}-\mathbf{X}\hat{\beta}||^2$ is too high, $\Rightarrow\ \psi$ is over-estimated.
- $\Rightarrow$ both conditional mean and variances can be wrong.
- $\Rightarrow$ unconditional mean is wrong.
- *But, does not necessarily result in noticeable lack of goodness-of-fit in the marginal of* **Y**, consequently underfitting is hard to detect.

The Fay-Herriot model targets

$$\hat{\beta}_{PR} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y},$$

$$\hat{\psi}_{PR} \doteq (n-p)^{-1}\left(||\mathbf{Y} - \mathbf{X}\hat{\beta}||^2 - \sum(1 - H_{ii})D_i\right),$$

$$\hat{\theta}_i = \left(\hat{\psi}Y_i + D_i\mathbf{x}_i^T\hat{\beta}\right) / \left(D_i + \hat{\psi}\right), \ \mathbb{V}[\theta_i|\mathbf{Y}] = \left\{D_i\,\hat{\psi}\right\} / \left(D_i + \hat{\psi}\right).$$

- *Over-fitting is extremely dangerous* : $||\mathbf{Y} - \mathbf{X}\hat{\beta}||^2$ is too low, $\Rightarrow$ ψ is under-estimated.
- Hence more weight on the “prior”, *i.e.* on the auxiliary information rather than on the directly observed data.
- The conditional variance is low, so we are very confident about our wrong predictor.
- Classical model selection tools do not work (parameters β and ψ have different roles to play, and have different interpretation from standard parametric (Bayesian or frequentist) models).

Example

Modified from Datta, Rao, Smith (2005, Biometrika)

- We have $n = 5n_0$ small areas.
- We specify D_i's as n_0 copies of $2, 0.6, 0.5, 0.4, 0.2$.
- We use Level-II variance $\psi = 1$.
- We have 5 covariates, only the first 2 of which have non-zero β coefficients associated with them.
- The goal is to see how some established, and some exploratory methods perform in selecting the correct covariates.

The modified DRS simulation framework

Example

n	AIC	BIC	MoM	T1	T2
20	46	51	82	36	81
50	48	58	83	61	82
200	63	64	85	66	81
500	56	63	86	69	84

Table: Estimated correct covariate selection proportions of various techniques.

- *BIC* performs very poorly is a slightly more complex model.
- *MoM* is a method of moment model selection criterion. Not very efficient, inadequate outside linear mixed models.
- T_1 is a likelihood-based criterion. Requires a tuning constant, which may be hard to obtain in real data problems.
- T_2 is something a few co-authors started on (I was called in to show that this works): It is based on m-out-of-n bootstrap and differential fitting to observations in the resample and outside it. Not consistent in general.
- We need to compare with the fence methods (Jiang, Ngyuen *et al.*).
- *We have not considered the variance component ψ at all!*

Outline

General issues with model selection

- Both regression coefficients and variance components should be considered when doing model selection.
- The variance component in the FH model (and in other mixed models) is based on residuals, so under or over-fitting of regression terms affects it, and this acts as a shock-absorber (dampens the effect of fitting bad models).
- *Post model selection inference:* In general, a model *selected* by some criterion does not have classical regularity properties: it's parameters are super-efficient.
- *Issues with parameter on the boundary of the parameter space:* Since $\psi \in [0, \infty)$, the distribution of $\hat{\psi}$ is non-trivial. Testing for $\psi = 0$ is not easy.

So what can be done?

- Use resampling!
- (on second thoughts: empirical likelihood may work also, needs to be explored).
- Our method (sketch below) can obtain a consistent estimator of the joint distribution of the scores of all models and parameters included in the models.
- Thus, we *can perform* post-model selection inference, can include all parameters, and variance components being zero or not does not matter.
- Our (preliminary) proof hold for all linear mixed models.
- *Another matter:* We cannot simultaneously have correct model chosen with probability tending to 1 (property of BIC), and bounded risk function (property of AIC).

FH model selection with resampling

- We need at least one model to estimate ψ consistently, so we assume that the maximal model (the one which includes every parameter) is adequate, ie, does not exclude any necessary covariates.
- Let $\hat{\psi}$ be the variance component estimate from the maximal model.
- In model *s*, instead of the EBLUP, we use fixed effects from the model itself, and the random effect using the variance component of the maximal model and the residuals from model *s*, thus:

$$\boldsymbol{\theta}_{sb} = P_s\mathbf{Y} + (\mathbb{I}_n - \hat{B})(\mathbb{I}_n - P_s)\mathbf{Y},$$

where
$P_s =$ projection on the column space of covariates included in model *s*, and

$$\hat{\mathbf{B}} = diag\left(\hat{B}_i = D_i/(D_i + \hat{\psi}), i = 1, \ldots, n\right)$$

FH model selection: WiSE errors

- Define a sequence $\{\tau_n\}$ such that $\tau_n \to \infty$ and $\tau_n/n \to 0$ as $n \to \infty$.
- For the b^{th} bootstrap Monte Carlo step, generate a vector U_b of i.i.d. random variables with mean zero and variance one.
- Define

$$E_{sb} = \sqrt{\tau_n} diag\big(U_b\big)(\mathbb{I}_n - \hat{P})\mathbf{Y}$$

  where $(\mathbb{I}_n - \hat{P})\mathbf{Y}$ are the residuals from the maximal model.
- (Slight variations of the above work are also OK.)

FH model selection: WiSE errors

- Now define $\mathbf{Y}_{sb} = \boldsymbol{\theta}_{sb} + E_{sb}$ as the **Wi**ld **S**cale **E**nhanced bootstrap version of $\mathbf{Y} = \theta + E$ of the Fay-Herriot model.
- Fit model s to this. Obtain:
- The score of model s, defined as

$$\hat{\Gamma}(s) = \mathbb{E}_{B} ||\mathbf{Y} - \hat{\theta}_{sb}||^2,$$

  where $\hat{\theta}_{sb}$ is the EBLUP under model s, and $\mathbb{E}_{B}$ is bootstrap expectation (ie, expectation conditional on the observed data).
- Obtain the WiSE boosttrap distribution of any parameter of interest under model s.

Outline

A simulation experiment

- We generate $n = 100$ sized datasets, from the model

$$Y_i = \mu + U_i + E_i, \text{ where}$$

- $E_i \overset{ind.}{\sim} N(0, D_i)$ with known $D_i's$ and $U_i \overset{i.i.d.}{\sim} N(0, \psi)$.
- Four models considered: (s1) $\mu \neq 0$, $\psi > 0$, (s2) $\mu \neq 0$, $\psi = 0$, (s3) $\mu = 0$, $\psi > 0$, (s4) $\mu = 0$, $\psi = 0$.
- We try out the above with bootstrap Monte Carlo size $B = 200$, and $R = 500$ independent replications.

WiSE-boot model selection simulations

Example

Correct Selected	s1	s2	s3	s4
s1 ($\mu \neq 0$, $\psi > 0$)	96.2	3.8	0	0
s2 ($\mu \neq 0$, $\psi = 0$)	0	100	0	0
s3 ($\mu = 0$, $\psi > 0$)	32.0	0.6	63.8	3.6
s4 ($\mu = 0$, $\psi = 0$)	0	2	0	98

Table: Estimated percentage of times correct model is selected.

Model scores under s1

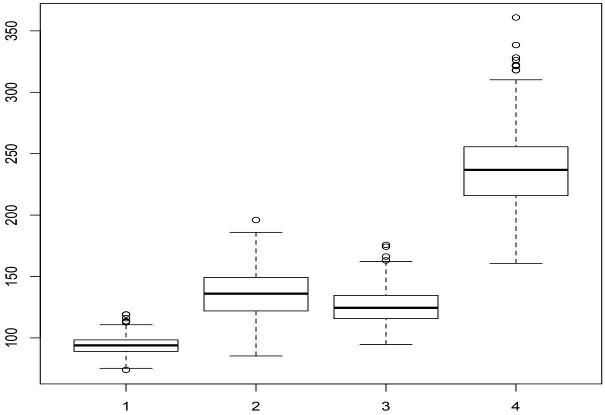

Figure: Model scores $\hat{\Gamma}(\cdot)$ under model s1 ($\mu \neq 0$, $\psi > 0$)

Model scores under s2

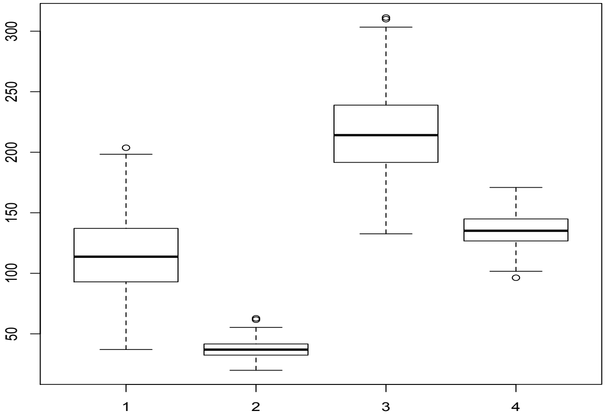

Figure: Model scores $\hat{\Gamma}(\cdot)$ under model s2 ($\mu \neq 0$, $\psi = 0$)

Model scores under s3

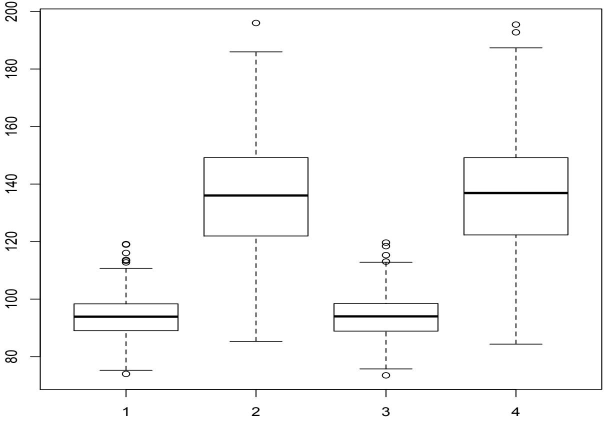

Figure: Model scores $\hat{\Gamma}(\cdot)$ under model s3 ($\mu = 0$, $\psi > 0$)

WiSE-Boot sampling distributions

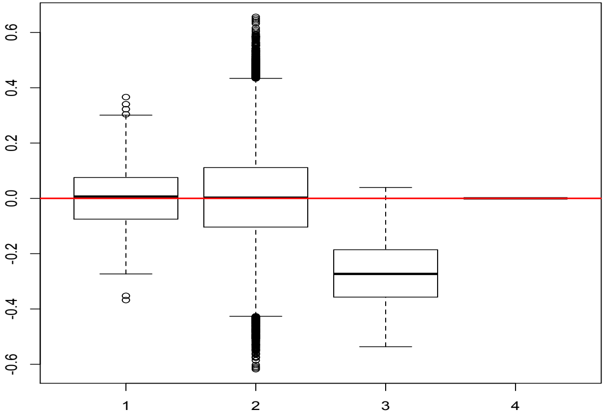

Figure: Sampling distribution of $\hat{\mu}$ and its WiSE-Boot approximation in s1 (all simulations), s1 (simulation # 425) and s3

Model scores under s4

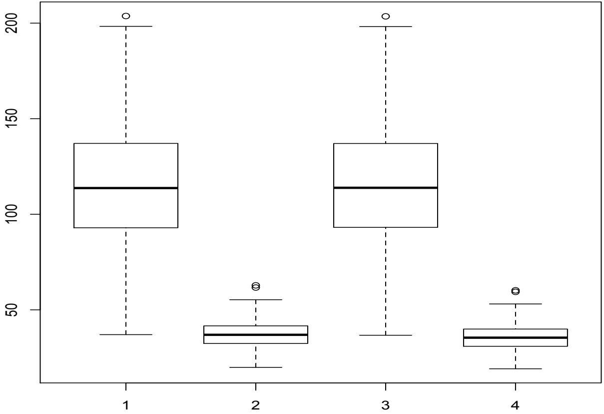

Figure: Model scores $\hat{\Gamma}(\cdot)$ under model s4 ($\mu = 0$, $\psi = 0$)

A different experiment with covariates

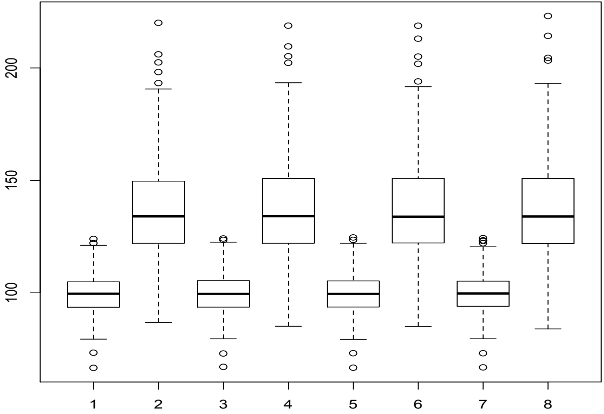

Figure: Model scores $\hat{\Gamma}(\cdot)$: Only random effect case in a different experiment with more models and covariates

WiSE-Boot sampling distributions

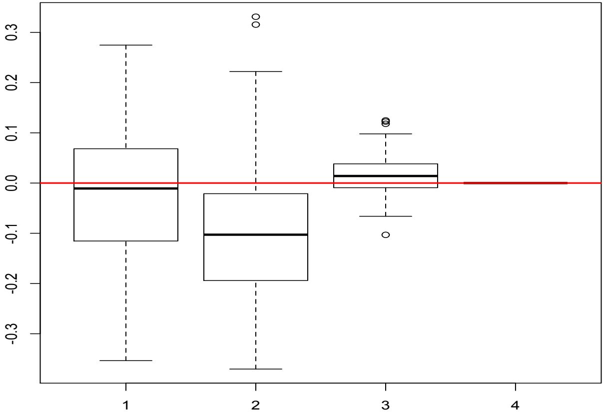

Figure: Same experiment: sampling distribution and its WiSE-Boot approximation in the three “good” models

Example

County	Data		Pred.
County	*D*	*Y*	*EBLUP*
1	0.517	16.76	15.63
6	0.029	1.96	2.29
11	7.581	158.03	158.57
23	1.793	28.53	26.82
40	0.030	0.84	1.20

Table: Partial results for State #6.

Example

County	Data		Pred.			
County	D	Y	$EBLUP$	$q_{0.05}$	$q_{0.5}$	$q_{0.95}$
1	0.517	16.76	15.63	14.52	15.73	18.11
6	0.029	1.96	2.29	1.90	2.34	2.88
11	7.581	158.03	158.57	156.50	166.14	192.31
23	1.793	28.53	26.82	25.32	27.26	30.37
40	0.030	0.84	1.20	0.84	1.22	1.52

Table: Partial results for State #6.

1

21

Poverty Mapping for the Chilean Comunas

Casas-Cordero, C. [1], Encina, J. [2] and Lahiri, P.[3]

21.1. Introduction

The eradication of poverty has been at the center of various public policies in Chile and has guided public policy efforts. The nationwide survey estimate of the poverty rate has declined since the early 90´s suggesting some progress towards this goal. While this result is encouraging, erratic time series patterns have emerged for small *comunas* - the smallest territorial entity in Chile. Moreover, for a handful of extremely small comunas, survey estimates of poverty rates are unavailable for some or all time points simply because the survey design, which traditionally focuses on precise estimates for the nation and large geographical areas, excludes these comunas for some or all of the time points. In any case, direct survey estimates of poverty rates typically do not meet the desired precision for small comunas and thus the assessment of implemented policies is not straightforward at the comuna level. In order to successfully monitor trends, identify influential factors, develop effective public policies and eradicate poverty at the comuna level, there is a growing need to improve on the methodology for estimating poverty rates at this level of geography.

Chile's official data source for poverty statistics is the National Socioeconomic Characterization Survey

Chilean poverty data example

- The data is from 334 comunas (administrative subdivision) from 30 groups of Chile, on proportion of households with income below a threshold, directly estimated using a trimmed comuna weight.
- Several auxiliary variables are available.
- Preliminary model selection chooses around 40 % of auxiliary variables.

Acnowledgement:

- This research is partially supported by the National Science Foundation (NSF) under grant # IIS-1029711, # SES-0851705, and by the National Aeronautics and Space Administration (NASA) under grant #-1502546.
- This research is partially supported by the Institute on the Environment (IonE), and College of Liberal Arts (CLA).

Thank you