PREDICTING RANDOM EFFECTS IN GROUP RANDOMIZED TRIALS

Ed Stanek
Dept. of Biostatistics and Epidemiology
School of Public Health & Health Sciences
UMASS, Amherst, MA

Outline

- Description of the Problem
  - Example
- Basic Models & Ideas
- Approaches
- Comparison of Results
- Extensions
- Open Questions

Description of the Problem

Clustered Population (cluster=group)

Schools, Clinics, Hospitals, Cities, Neighborhoods, Physician Practices, Families, Litters, States, Studies…

Data is available only on some clusters

Observational studies, multi-stage samples, group-randomized trials

For selected clusters, response is observed on a subset of units (possibly with response error).

Example: Development of Negative Behaviors in Schools

Population:

Clusters = Schools in a city (N=11)

Units = All 6th grade students in the schools (M varies by school)

Response = Score on a set of 9 "bullying" questions.

"During this school year, other kids in school called me names or swore at me"

1=Not at all

2=Once

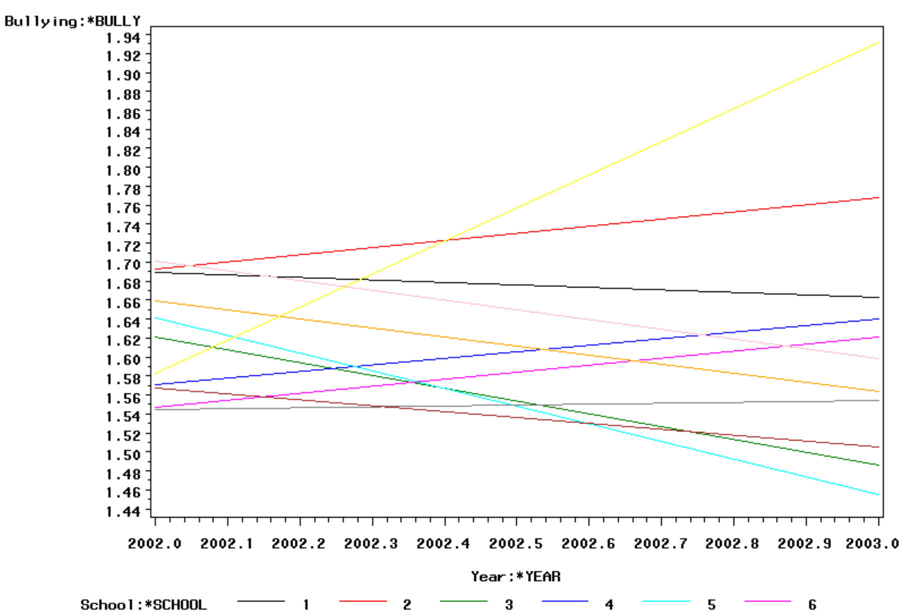
3=2-3 Times

4=4 or More times

Data Available

School	Total Students	Response Subset	Percent
1	180	67	37%
2	99	46	46%
3	122	69	57%
4	67	42	63%
5	328	134	41%
6	223	57	26%
7	114	61	54%
8	319	133	42%
9	202	117	58%
10	241	142	59%
11	63	29	46%

Figure 1. Bullying of 6th and 7th graders



Change in Bullying Scores By School

	Observed Change in Bullying Score By School			
School	n	Diff	Std	6^{th} Grade Ave
1	63.00	-0.03	0.58	1.69
2	45.00	0.08	0.63	1.69
3	65.00	-0.13	0.59	1.62
4	42.00	0.07	0.81	1.57
5	130.00	-0.19	0.65	1.64
6	41.00	0.07	0.65	1.55
7	61.00	0.01	0.73	1.55
8	112.00	-0.10	0.67	1.70
9	112.00	-0.10	0.65	1.66
10	130.00	-0.06	0.69	1.57
11	26.00	0.35	0.94	1.58

Change in Bullying Scores

School	Observed	BLUP	Sampling Fraction
1	-0.03	-0.03	37%
2	0.08	0.01	46%
3	-0.13	-0.08	57%
4	0.07	0.01	63%
5	-0.19	-0.13	41%
6	0.07	0.01	26%
7	0.01	-0.01	54%
8	-0.10	-0.07	42%
9	-0.10	-0.07	58%
10	-0.06	-0.05	59%
11	**0.35**	**0.07**	46%

Background

Early work -1950's (Scheffe, Wilk and Kempthorne)

- "We see that in formulating a model one must ask for each factor whether one is interested individually in the particular levels occurring in the experiment or primarily in a population from which the levels in the experiment can be regarded as a sample: the main effects are accordingly treated as fixed or as random." (Scheffe, p254 1956)

Background

- Early work -1950's (Scheffe, Wilk and Kempthorne)
  - "We see that in formulating a model one must ask for each factor whether **one is interested individually in the particular levels occurring in the experiment** or primarily in a population from which the levels in the experiment can be regarded as a sample: the main effects are accordingly treated as fixed or as random." (Scheffe, p254 1956)

Background

- "… the decision as to whether the main effects of any factor, say A, are to be treated as fixed or random obviously affects the meaning of the main effects of A…" (Scheffe, p255. 1956)

Basic Models & Ideas

Simple Response Error Model:

School: $s = 1,\ldots,N$
Student: $t = 1,\ldots,M$

$$Y_{stk} = y_{st} + W_{stk}$$

where $E\left(W_{stk}\right) = 0$

We define $\mu_s = \frac{1}{M}\sum_{t=1}^{M} y_{st}$ and $\mu = \frac{1}{N}\sum_{s=1}^{N} \mu_s$

$$\beta_s = (\mu_s - \mu) \qquad \varepsilon_{st} = (y_{st} - \mu_s)$$

Then

$$y_{st} = \mu + \beta_s + \varepsilon_{st}$$

or

$$\begin{aligned} Y_{stk} &= y_{st} + W_{stk} \\ &= \mu + \beta_s + \varepsilon_{st} + W_{stk} \end{aligned}$$

Suppose we select a 'sample' of Schools, and a 'sample' of students in the selected school.

Consider $\mu_s = \frac{1}{M}\sum_{t=1}^{M} y_{st}$

New Model for $i = 1, \ldots, n$

define random variables U_{is}

and school in ith position: $\sum_{s=1}^{N} U_{is}\mu_s$

Similarly, for a student in the jth position, $j = 1,...,m$

define the RVs: $U_{jt}^{(s)}$

Then in school "s", $\sum_{t=1}^{M} U_{jt}^{(s)} y_{st}$

2-Stage sample (in the ith school):

$$Y_{ij} = \sum_{s=1}^{N} \sum_{t=1}^{M} U_{is} U_{jt}^{(s)} y_{st}$$

Combining terms, the Response Error Model:

$$Y_{stk} = \mu + \beta_s + \varepsilon_{st} + W_{stk}$$

becomes

$$Y_{ijk} = \mu + B_i + E_{ij} + W_{ijk}^*$$

where

ith school effect $B_i = \sum_{s=1}^{N} U_{is} \beta_s$

jth student effect $E_{ij} = \sum_{s=1}^{N} \sum_{t=1}^{M} U_{is} U_{jt}^{(s)} \varepsilon_{st}$

kth response error $W_{ijk}^* = \sum_{s=1}^{N} \sum_{t=1}^{M} U_{is} U_{jt}^{(s)} W_{stk}$

Model Properties

$$Y_{ijk} = \mu + B_i + E_{ij} + W^*_{ijk}$$

$$E_{\xi_1}\left(B_i\right) = 0$$

$$E_{\xi_2}\left(E_{ij}\right) = 0$$

The Latent Value of the School in the ith postion:

$$\mu + B_i = \sum_{s=1}^{N} U_{is}\mu_s$$

Breaking Up the Parameter

Recall: $\mu_s = \frac{1}{M}\sum_{t=1}^{M} y_{st}$

then $\mu_s = \frac{1}{M}\left(\sum_{j=1}^{m} Y_{sj} + \sum_{j=m+1}^{M} Y_{sj}\right)$

Let $\overline{Y}_{sI} = \frac{1}{m}\sum_{j=1}^{m} Y_{sj}$ and $\overline{Y}_{sII} = \frac{1}{M-m}\sum_{j=m+1}^{M} Y_{sj}$

Then $\mu_s = f\overline{Y}_{sI} + \left(1-f\right)\overline{Y}_{sII}$ where $f = \frac{m}{M}$

How should we predict $\mu + B_i$?

If there is no response error, use

$$\hat{\mu}_s = f\overline{Y}_{sI} + (1-f)\hat{\overline{Y}}_{sII}$$

Suppose the sample includes 95% of the students:

$$\hat{\mu}_s = 0.95\overline{Y}_{sI} + 0.05\hat{\overline{Y}}_{sII}$$

Models and Approaches

1. Henderson's Mixed Model
2. Bayesian Models
3. Super-population Models
4. Random permutation Models

Solution (Henderson) $Y_{ijk} = \mu + B_i + \left(E_{ij} + W^*_{ijk}\right)$

$$\mathbf{Y} = \mathbf{X}\alpha + \mathbf{ZB} + \mathbf{E}$$

$$\text{var}\left(\mathbf{B}\right) = \mathbf{G} \quad \text{var}\left(\mathbf{E}\right) = \mathbf{R}$$

$$\mathbf{\Sigma} = \text{var}\left(\mathbf{ZB} + \mathbf{E}\right) = \mathbf{ZGZ}' + \mathbf{R}$$

WLS Equations

$$\left(\mathbf{X}'\mathbf{\Sigma}^{-1}\mathbf{X}'\right)\hat{\alpha} = \mathbf{X}'\mathbf{\Sigma}^{-1}\mathbf{Y}$$

Substituting $\mathbf{\Sigma}^{-1} = \mathbf{R}^{-1} - \mathbf{R}^{-1}\mathbf{Z}\left(\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1}\right)^{-1}\mathbf{Z}'\mathbf{R}^{-1}$

$$\mathbf{X}'\mathbf{R}^{-1}\mathbf{X}\hat{\alpha} - \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z}\left[\left(\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1}\right)^{-1}\mathbf{Z}'\mathbf{R}^{-1}\left(\mathbf{Y} - \mathbf{X}\hat{\alpha}\right)\right] = \mathbf{X}'\mathbf{R}^{-1}\mathbf{Y}$$

Suppose we define:

$$\hat{\mathbf{B}} = \left(\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1}\right)^{-1}\mathbf{Z}'\mathbf{R}^{-1}\left(\mathbf{Y} - \mathbf{X}\hat{\alpha}\right)$$

Henderson's Mixed Model Equations

$$\mathbf{X}'\mathbf{R}^{-1}\mathbf{X}\hat{\alpha} + \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z}\hat{\mathbf{B}} = \mathbf{X}'\mathbf{R}^{-1}\mathbf{Y}$$

$$\mathbf{Z}'\mathbf{R}^{-1}\mathbf{X}\hat{\alpha} + \left(\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1}\right)\hat{\mathbf{B}} = \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Y}$$

Now $\hat{\mathbf{B}} = \left(\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1}\right)^{-1}\mathbf{Z}'\mathbf{R}^{-1}\left(\mathbf{Y} - \mathbf{X}\hat{\alpha}\right)$

or equivalently since $\left(\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1}\right)^{-1}\mathbf{Z}'\mathbf{R}^{-1} = \mathbf{GZ}'\mathbf{\Sigma}^{-1}$

$$\hat{\mathbf{B}} = \mathbf{GZ}'\mathbf{\Sigma}^{-1}\left(\mathbf{Y} - \mathbf{X}\hat{\alpha}\right)$$

Alternative Rational for Mixed Model Equations

Start with: $\mathbf{Y} = \mathbf{X}\alpha + \mathbf{ZB} + \mathbf{E}$

$$\mathbf{\Sigma} = \mathrm{var}\left(\mathbf{ZB} + \mathbf{E}\right) = \mathbf{ZGZ}' + \mathbf{R}$$

Express Joint Distribution of $\begin{pmatrix} \mathbf{Y} \\ \mathbf{B} \end{pmatrix}$

$$E\begin{pmatrix} \mathbf{Y} \\ \mathbf{B} \end{pmatrix} = \begin{pmatrix} \mathbf{X}\alpha \\ \mathbf{0}_n \end{pmatrix} \qquad \mathrm{var}\begin{pmatrix} \mathbf{Y} \\ \mathbf{B} \end{pmatrix} = \begin{pmatrix} \mathbf{\Sigma} & \mathbf{ZG} \\ \mathbf{GZ}' & \mathbf{G} \end{pmatrix}$$

The BLUP of $\alpha + B_i$ is $\hat{\mathbf{B}} = \mathbf{GZ}'\mathbf{\Sigma}^{-1}\left(\mathbf{Y} - \mathbf{X}\hat{\alpha}\right)$

Example: Sample: (n schools, m students/school)

$$\mathbf{Y} = \mathbf{X}\alpha + \mathbf{ZB} + \mathbf{E} \qquad \mathbf{X} = \mathbf{1}_{nm} \qquad \mathbf{Z} = \mathbf{I}_n \otimes \mathbf{1}_m$$

$$\mathbf{G} = \sigma^2 \mathbf{I}_n \qquad \mathbf{R} = \sigma_e^2 \mathbf{I}_{nm}$$

$$\mathrm{var}\left(B_i\right) = \sigma^2 \quad \mathrm{var}\left(E_{ij}\right) = \sigma_e^2$$

$$\hat{\alpha} + \hat{B}_i = \bar{\bar{Y}} + k\left(\bar{Y}_i - \bar{\bar{Y}}\right)$$

where $\bar{Y}_i = \frac{1}{m}\sum_{j=1}^{m} Y_{ij}$ and $\bar{\bar{Y}} = \frac{1}{n}\sum_{i=1}^{n} \bar{Y}_i$

and $k = \frac{\sigma^2}{\sigma^2 + \sigma_e^2 / m}$

Bayesian Estimation

Hierarchical Model

Student (1): $Y_{stk} = \mu_{st} + E_{stk}$

School (2): $Y_{sj} = \mu_{s} + E_{sj}$

Population (3): $Y_{i} = \mu + E_{i}$

Joint Model:

Level 1: $Y_{ijk} = A_{ij} + E_{ijk}$ $\quad A_{ij} = A_{i} + E_{ij}$ $\quad \mathbf{E} \sim N\left(\mathbf{0}_{nm}, \mathbf{R}\right)$

Level 2: $Y_{ij} = A_{i} + E_{ij}$ $\quad A_{i} = A + B_{i}$ $\quad \mathbf{B} \sim N\left(\mathbf{0}_{n}, \mathbf{G}\right)$

Level 3: $Y_{i} = A + E_{i}$ $\quad A \sim N\left(\alpha, \tau^{2}\right)$

$$\mathbf{Y} = \mathbf{X}A + \mathbf{Z}\mathbf{B} + \mathbf{E}$$

Bayesian Estimation $\quad \mathbf{Y} = \mathbf{X}A + \mathbf{Z}\mathbf{B} + \mathbf{E}$

Suppose: $\mathbf{G} = \sigma^{2}\mathbf{I}_{n}$ and $\mathbf{R} = \sigma_{e}^{2}\mathbf{I}_{nm}$

Assume σ^{2} and σ_{e}^{2} are constant, and $\tau^{2} = \infty$

Then $\quad E\left(\mathbf{B} \mid \mathbf{Y} = \mathbf{y}\right) = \hat{\mathbf{B}}$

where $\quad \hat{\mathbf{B}} = \mathbf{G}\mathbf{Z}'\mathbf{\Sigma}^{-1}\left(\mathbf{Y} - \mathbf{X}\hat{\alpha}\right)$

$$\hat{A} + \hat{B}_{i} = \bar{\bar{y}} + k\left(\bar{y}_{i} - \bar{\bar{y}}\right) \qquad \text{where} \quad k = \frac{\sigma^{2}}{\sigma^{2} + \sigma_{e}^{2}/m}$$

Super-population Model Predictor

Super-population

(collection of random variables that follow a model)

$$Y_{ij}$$

$$E\left(Y_{ij}\right)=\mu$$

$$\operatorname{var}\left(Y_{ij}\right)=\sigma^2+\sigma_e^2 \qquad \operatorname{cov}\left(Y_{ij},Y_{ij*}\right)=\sigma^2$$

Realization (i.e. Population):

PSU Mean: $\bar{Y}_i=\frac{1}{M}\sum_{j=1}^{M}Y_{ij}$

Divide PSU mean into:

Sample: $\bar{Y}_{i,I}=\frac{1}{n}\sum_{j=1}^{n}Y_{ij}$

Remainder: $\bar{Y}_{i,II}=\frac{1}{N-n}\sum_{j=n+1}^{N}Y_{ij}$

$$\bar{Y}_i=f\bar{Y}_{i,I}+\left(1-f\right)\bar{Y}_{i,II}$$

BLUP: $f\bar{y}_i+\left(1-f\right)\hat{\bar{Y}}_{i,II}$

$$\hat{\bar{Y}}_{i,II}=\bar{\bar{y}}+k\left(\bar{y}_i-\bar{\bar{y}}\right)$$

$$k=\frac{\sigma^2}{\sigma^2+\sigma_e^2/m}$$

Random Permutation Model Predictor

Population (Schools and Students): y_{st}

Random Variables
(from permutations): $Y_{ij} = \sum_{s=1}^{N} \sum_{t=1}^{M} U_{is} U_{jt}^{(s)} y_{st}$

$$Y_{ijk} = \mu + B_i + \left(E_{ij} + W_{ijk}^* \right)$$

Predict $\mu + \mathrm{B}_i = f\overline{Y}_{iI} + \left(1 - f\right)\overline{Y}_{iII}$ by

$f\overline{y}_i + \left(1 - f\right)\hat{\overline{Y}}_{i,II}^*$ where $\hat{\overline{Y}}_{i,II}^* = \overline{\overline{y}} + k^* \left(\overline{y}_i - \overline{\overline{y}} \right)$

Random Permutation Model Predictor

$$f\overline{y}_i + \left(1 - f\right)\hat{\overline{Y}}_{i,II}^*$$

$$\hat{\overline{Y}}_{i,II}^* = \overline{\overline{y}} + k^* \left(\overline{y}_i - \overline{\overline{y}} \right)$$

where $k^* = \dfrac{\sigma^{*2}}{\sigma^{*2} + \sigma_e^2 / m}$

and $\sigma^{*2} = \sigma^2 - \dfrac{\sigma_e^2}{M}$

Figure 1a. Percent Increase in Expected MSE for Mixed Model (—) and Scott and Smith Model (- - -) Predictors Relative to the Random Permutation Model Predictors of the Latent Value of a Realized Sample PSU

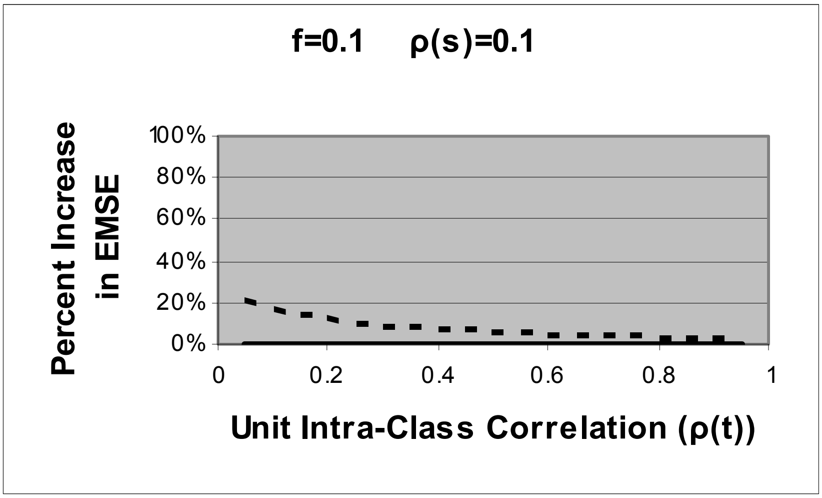


Figure 1b. Percent Increase in Expected MSE for Mixed Model (—) and Scott and Smith Model (- - -) Predictors Relative to the Random Permutation Model Predictors of the Latent Value of a Realized Sample PSU

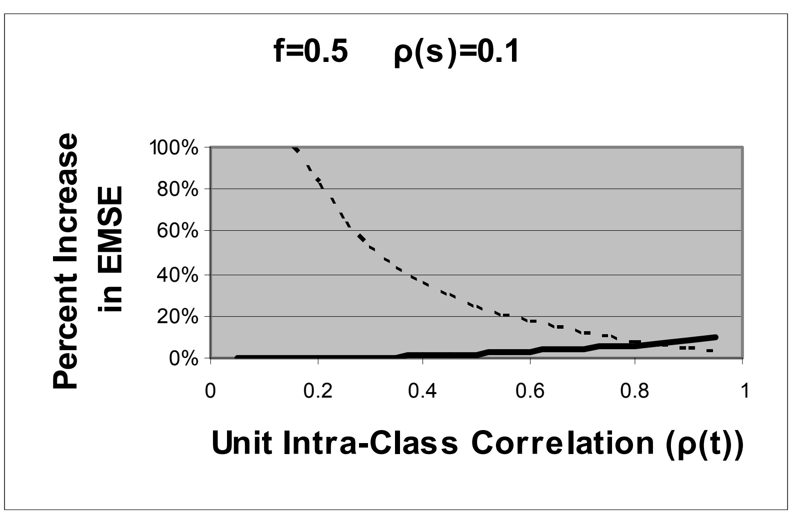


Figure 1c. Percent Increase in Expected MSE for Mixed Model (—) and Scott and Smith Model (- - -) Predictors Relative to the Random Permutation Model Predictors of the Latent Value of a Realized Sample PSU

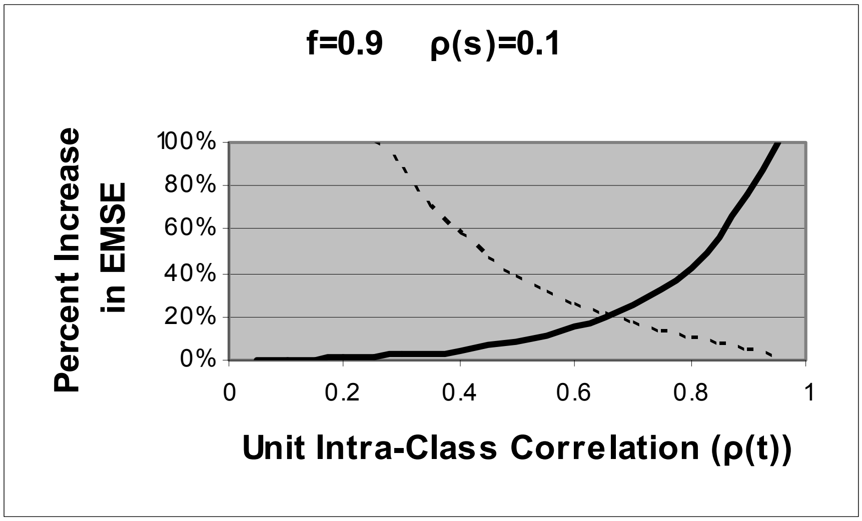


Figure 1d. Percent Increase in Expected MSE for Mixed Model (—) and Scott and Smith Model (- - -) Predictors Relative to the Random Permutation Model Predictors of the Latent Value of a Realized Sample PSU

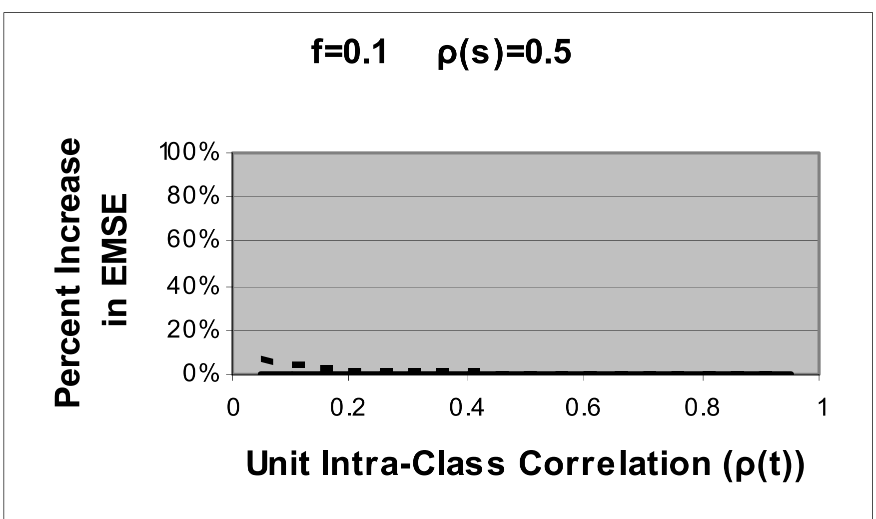


Figure 1e. Percent Increase in Expected MSE for Mixed Model (—) and Scott and Smith Model (- - -) Predictors Relative to the Random Permutation Model Predictors of the Latent Value of a Realized Sample PSU

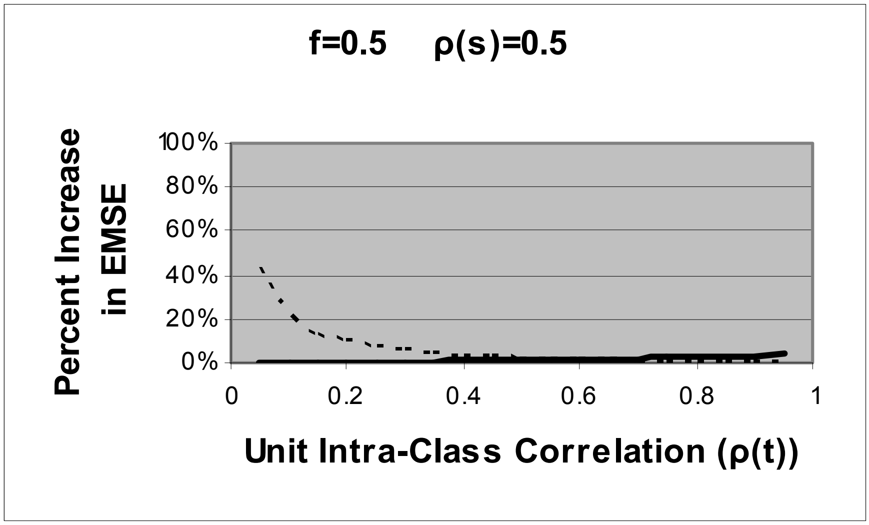


Figure 1f. Percent Increase in Expected MSE for Mixed Model (—) and Scott and Smith Model (- - -) Predictors Relative to the Random Permutation Model Predictors of the Latent Value of a Realized Sample PSU

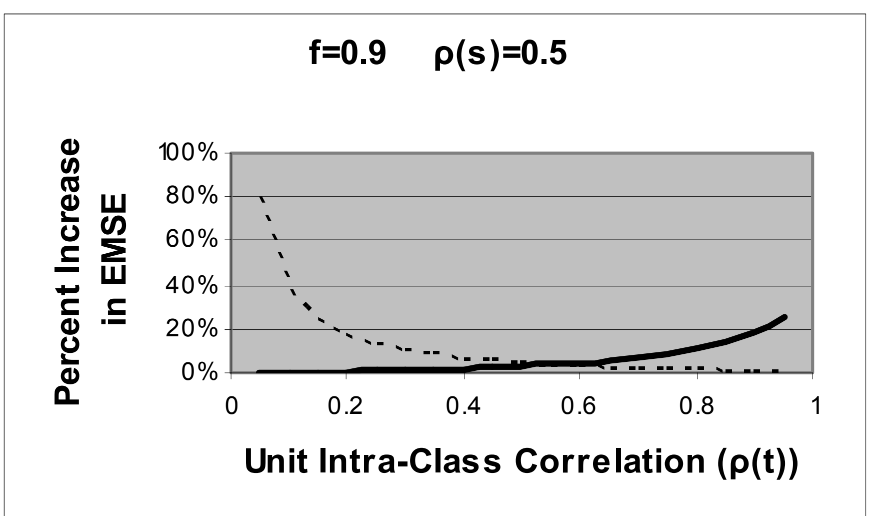


Figure 1g. Percent Increase in Expected MSE for Mixed Model (—) and Scott and Smith Model (- - -) Predictors Relative to the Random Permutation Model Predictors of the Latent Value of a Realized Sample PSU

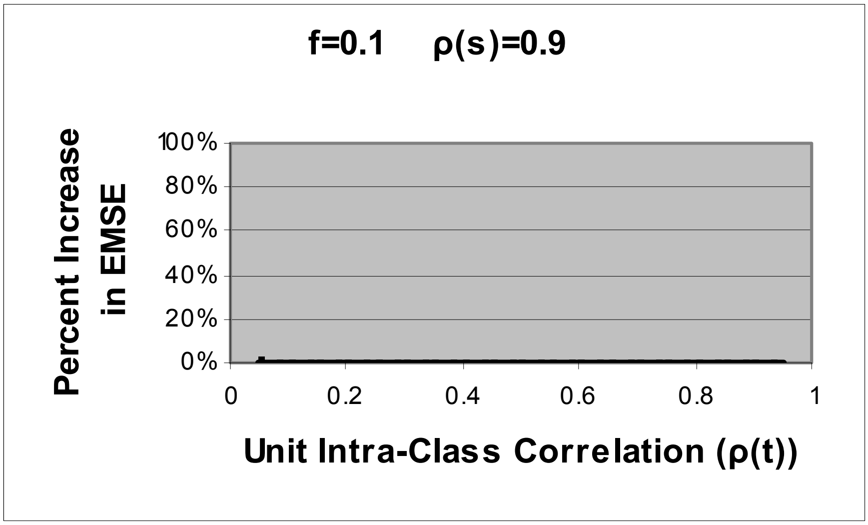


Figure 1h. Percent Increase in Expected MSE for Mixed Model (—) and Scott and Smith Model (- - -) Predictors Relative to the Random Permutation Model Predictors of the Latent Value of a Realized Sample PSU

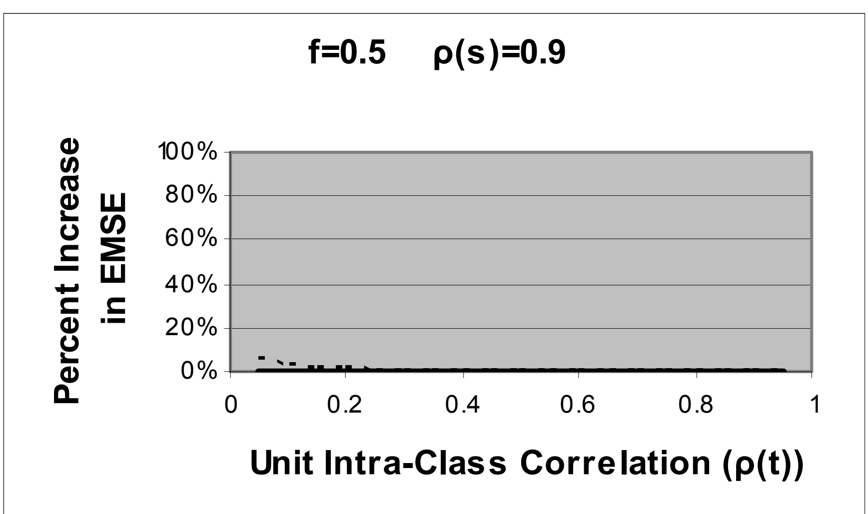
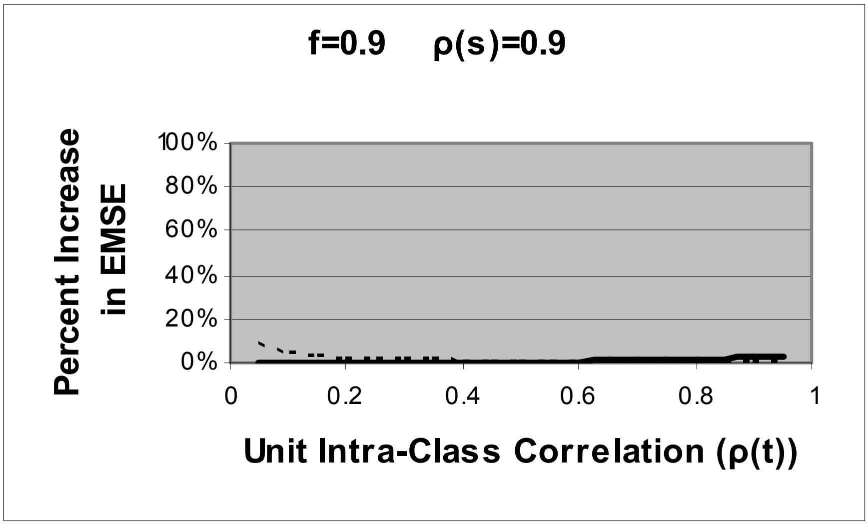


Figure 1i. Percent Increase in Expected MSE for Mixed Model (—) and Scott and Smith Model (- - -) Predictors Relative to the Random Permutation Model Predictors of the Latent Value of a Realized Sample PSU



Other Issues

"Improvement" in predictor is based on lower Average MSE.

Estimates needed of shrinkage constant

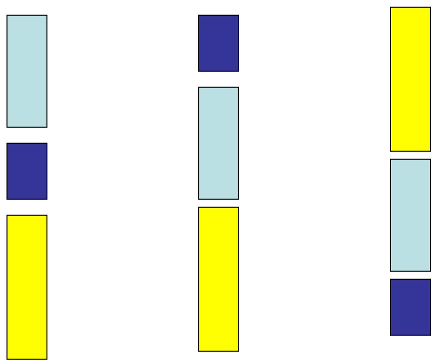
Partial solution: Minimize problems by using other 'covariates' to predict PSUs.

What happens when cluster sizes differ?

Are we predicting the right thing?

Unequal Size Clusters

Permutations



SRS: Unit versus Random Effect

$$Y_i = \sum_{s=1}^{N} U_{is} y_s \quad \text{versus} \quad \mathbf{Y}_i = \begin{pmatrix} U_{i1} y_1 \\ U_{i2} y_2 \\ \vdots \\ U_{iN} y_N \end{pmatrix}$$

Different linear combinations define:

The parameter for unit s: y_s

The random variable Y_i $\qquad Y_i = \mu + \mathrm{B}_i$

Uncertainty in interpretation- more work needed
Problem too simple - need extensions to other settings (Predictors of Y_i)

Thanks

Summary

Mixed Model

$$E(\mathbf{Y}) = \mathbf{X}\mu$$
$$\operatorname{var}(\mathbf{Y}) = \bigoplus_{i=1}^{N}\left(\sigma_i^2\mathbf{I}_M + \sigma^2\mathbf{J}_M\right)$$

$$\hat{p} = \mathbf{e}_i'\left(\mathbf{1}_n\hat{\mu} + k_i\left(\overline{\mathbf{Y}}_I - \mathbf{1}_n\hat{\mu}\right)\right)$$

Scott and Smith Model

$$E_{\xi_1\xi_2}(\mathbf{Y}) = \mathbf{X}\mu$$
$$\operatorname{var}_{\xi_1\xi_2}(\mathbf{Y}) = \bigoplus_{i=1}^{N}\left(\sigma_i^2\mathbf{I}_M + \delta^2\mathbf{J}_M\right)$$

$$\hat{P}_i = f\mathbf{e}_i'\overline{\mathbf{Y}}_I + (1-f)\mathbf{e}_i'\left(\mathbf{1}_n\hat{\mu}^* + k_i^*\left(\overline{\mathbf{Y}}_I - \mathbf{1}_n\hat{\mu}^*\right)\right)$$

where $f = \frac{m}{M}$ for $i \le n$

2-Stage Random Permutation Model

$$E_{\xi_1\xi_2}(\mathbf{Y}) = \mathbf{X}\mu$$
$$\operatorname{var}_{\xi_1\xi_2}(\mathbf{Y}) = \bigoplus_{i=1}^{N}\left(\sigma_e^2\mathbf{I}_M + \sigma^{*2}\mathbf{J}_M\right) - \sigma^{*2}\frac{\mathbf{J}_{NM}}{N}$$

$$\hat{T}_i = f\mathbf{e}_i'\overline{\mathbf{Y}}_I + (1-f)\mathbf{e}_i'\left(\mathbf{1}_n\overline{Y} + k_i\left(\overline{\mathbf{Y}}_I - \mathbf{1}_n\overline{Y}\right)\right)$$
$$\overline{Y} = \frac{\sum_{i=1}^{n}\overline{Y}_i}{n}$$

2-Stage Random Permutation Model

$$E_{\xi_1\xi_2}(\mathbf{Y}) = \mathbf{X}\mu$$
$$\operatorname{var}_{\xi_1\xi_2}(\mathbf{Y}) = \bigoplus_{i=1}^{N}\left(\sigma_e^2\mathbf{I}_M + \sigma^{*2}\mathbf{J}_M\right) - \sigma^{*2}\frac{\mathbf{J}_{NM}}{N}$$

$$\hat{T}_i = f\mathbf{e}_i'\overline{\mathbf{Y}}_I + (1-f)\mathbf{e}_i'\left(\mathbf{1}_n\overline{Y} + k_i\left(\overline{\mathbf{Y}}_I - \mathbf{1}_n\overline{Y}\right)\right)$$
$$\overline{Y} = \frac{\sum_{i=1}^{n}\overline{Y}_i}{n}$$

2-Stage Random Permutation Model With Response Error

$$E_{\xi_1\xi_2}(\mathbf{Y}) = \mathbf{X}\mu$$
$$\operatorname{var}_{\xi_1\xi_2}(\mathbf{Y}) = \bigoplus_{i=1}^{N}\left(\left(\sigma_e^2 + \overline{\overline{\sigma}}^2\right)\mathbf{I}_M + \sigma^{*2}\mathbf{J}_M\right) - \sigma^{*2}\frac{\mathbf{J}_{NM}}{N}$$

$$\hat{T}_i = f\mathbf{e}_i'\left(\mathbf{1}_n\overline{Y} + k_r^*\left(\overline{\mathbf{Y}}_I - \mathbf{1}_n\overline{Y}\right)\right) + (1-f)\mathbf{e}_i'\left(\mathbf{1}_n\overline{Y} + k^*\left(\overline{\mathbf{Y}}_I - \mathbf{1}_n\overline{Y}\right)\right)$$

Mixed $\hat{p} = \mathbf{e}_i'\left(\mathbf{1}_n\hat{\mu} + k_i\left(\overline{\mathbf{Y}}_I - \mathbf{1}_n\hat{\mu}\right)\right)$

S&S $\hat{P}_i = f\mathbf{e}_i'\overline{\mathbf{Y}}_I + (1-f)\mathbf{e}_i'\left(\mathbf{1}_n\hat{\mu}^* + k_i^*\left(\overline{\mathbf{Y}}_I - \mathbf{1}_n\hat{\mu}^*\right)\right)$

RP $\hat{T}_i = f\mathbf{e}_i'\overline{\mathbf{Y}}_I + (1-f)\mathbf{e}_i'\left(\mathbf{1}_n\overline{Y} + k\left(\overline{\mathbf{Y}}_I - \mathbf{1}_n\overline{Y}\right)\right)$

RP&E $\hat{T}_i = f\mathbf{e}_i'\left(\mathbf{1}_n\overline{Y} + k_r^*\left(\overline{\mathbf{Y}}_I - \mathbf{1}_n\overline{Y}\right)\right) + (1-f)\mathbf{e}_i'\left(\mathbf{1}_n\overline{Y} + k^*\left(\overline{\mathbf{Y}}_I - \mathbf{1}_n\overline{Y}\right)\right)$

	Between	Within		
Mixed	σ^2	σ_i^2	$k_i = \frac{\sigma^2}{v_i}$	$v_i = \sigma^2 + \frac{\sigma_i^2}{m_i}$
S&S	δ^2	σ_i^2	$k_i^* = \frac{\delta^2}{v_i^*}$	$v_i^* = \delta^2 + \frac{\sigma_i^2}{m_i}$

RP	σ^2	$\sigma_s^2 \;\; \sigma_e^2$	$k = \frac{\sigma^{*2}}{v}$	$v = \sigma^{*2} + \frac{\sigma_e^2}{m}$	$\sigma^{*2} = \sigma^2 - \frac{\sigma_e^2}{M}$
		$\sigma_s^2 \;\; \sigma_e^2$	$k^* = \frac{\sigma^{*2}}{v^*}$	$v^* = \sigma^{*2} + \frac{\sigma_e^2 + \bar{\bar{\sigma}}^2}{m}$	
RP&E	σ^2	Response			
		$\sigma_{st}^2 \;\; \bar{\bar{\sigma}}^2$	$k_r^* = \frac{v}{v_r}$	$v_r = v + \frac{\bar{\bar{\sigma}}^2}{m}$	

Other Issues

"Improvement" in predictor is based on lower Average MSE.

Estimates needed of shrinkage constant

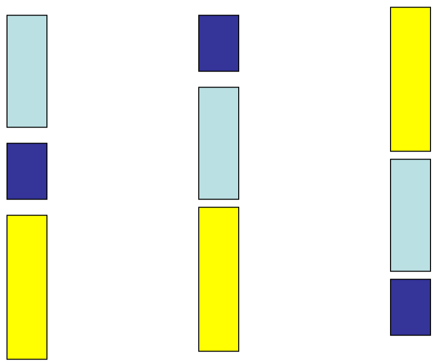
Partial solution: Minimize problems by using other 'covariates' to predict PSUs.

What happens when cluster sizes differ?

Are we predicting the right thing?

Unequal Size Clusters

Permutations



Pop. Values $y_{st} = \mu + \beta_s + \varepsilon_{st}$

Usual RVs $Y_{ij} = \sum_{s=1}^{N} \sum_{t=1}^{M_s} U_{is} U_{jt}^{(s)} y_{st}$

Expanded RVs $R_{isjt}^{*} = U_{is} U_{jt}^{(s)} y_{st}$

$\mathbf{R}_{st}^{*} = \left(\mathbf{U}_s \otimes \mathbf{U}_t^{(s)}\right) y_{st}$ where $\mathbf{U}_s = \left(U_{1s} \quad U_{2s} \quad \cdots \quad U_{Ns}\right)'$

$\mathbf{U}_t^{(s)} = \left(U_{1t}^{(s)} \quad U_{2t}^{(s)} \quad \cdots \quad U_{M_s t}^{(s)}\right)'$

$\mathbf{R}^{*} = \left(\mathbf{R}_{1+}^{*}{}' \quad \mathbf{R}_{2+}^{*}{}' \quad \cdots \quad \mathbf{R}_{N+}^{*}{}'\right)'$ where $\mathbf{R}_{s+}^{*} = \left(\mathbf{R}_{s1}^{*}{}' \quad \mathbf{R}_{s2}^{*}{}' \quad \cdots \quad \mathbf{R}_{sM_s}^{*}{}'\right)'$

$$\mathbf{R}^* = \begin{pmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \\ \vdots \\ \mathbf{X}_N \end{pmatrix} \mu + \begin{pmatrix} \mathbf{X}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{X}_N \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_N \end{pmatrix} + \begin{pmatrix} \mathbf{X}_{1+} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_{2+} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{X}_{N+} \end{pmatrix} \begin{pmatrix} \boldsymbol{\varepsilon}_1 \\ \boldsymbol{\varepsilon}_2 \\ \vdots \\ \boldsymbol{\varepsilon}_N \end{pmatrix} + \mathbf{E}$$

where $\mathbf{X}_s = \dfrac{\mathbf{1}_{NM_s^2}}{NM_s}$ and $\mathbf{X}_{s+} = \mathbf{I}_{M_s} \otimes \dfrac{\mathbf{1}_{NM_s}}{NM_s}$

$$E_{\xi_1\xi_2}\left(\mathbf{R}^*\right) = \begin{pmatrix} \mathbf{X}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{X}_N \end{pmatrix} \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_N \end{pmatrix} + \begin{pmatrix} \mathbf{X}_{1+} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_{2+} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{X}_{N+} \end{pmatrix} \boldsymbol{\varepsilon}$$

$$\operatorname{var}_{\xi_1\xi_2}\left(\mathbf{R}^*\right) = \bigoplus_{s=1}^{N} \left[\left(\frac{1}{M_s - 1} \right) \left(\bigoplus_{t=1}^{M_1} y_{st}^2 - \frac{\mathbf{y}_s \mathbf{y}_s{}'}{M_s} \right) \otimes \frac{\mathbf{I}_N}{N} \otimes \mathbf{P}_{M_s} + \left(\frac{1}{N-1} \right) \mathbf{y}_s \mathbf{y}_s{}' \otimes \mathbf{P}_N \otimes \frac{\mathbf{J}_{M_s}}{M_s^2} \right]$$

$$- \frac{1}{N(N-1)} \begin{pmatrix} \mathbf{y}_1 \otimes \mathbf{P}_N \otimes \dfrac{\mathbf{1}_{M_1}}{M_1} \\ \mathbf{y}_2 \otimes \mathbf{P}_N \otimes \dfrac{\mathbf{1}_{M_2}}{M_2} \\ \vdots \\ \mathbf{y}_N \otimes \mathbf{P}_N \otimes \dfrac{\mathbf{1}_{M_N}}{M_N} \end{pmatrix} \begin{pmatrix} \mathbf{y}_1 \otimes \mathbf{P}_N \otimes \dfrac{\mathbf{1}_{M_1}}{M_1} \\ \mathbf{y}_2 \otimes \mathbf{P}_N \otimes \dfrac{\mathbf{1}_{M_2}}{M_2} \\ \vdots \\ \mathbf{y}_N \otimes \mathbf{P}_N \otimes \dfrac{\mathbf{1}_{M_N}}{M_N} \end{pmatrix}'$$

Parameters of Interest

Cluster Total $P_i^{\circ}=\sum_{s=1}^{N}U_{is}M_s\mu_s$ or $P_i^{\circ}=\mathbf{g}_i'^{\circ}\mathbf{R}^*$

where $\mathbf{g}_i'^{\circ}=\mathbf{1}_N'\left(\bigoplus_{s=1}^{N}\left[\mathbf{1}_{M_s}'\otimes\left(\mathbf{e}_i'\otimes\mathbf{1}_{M_s}'\right)\right]\right)$

Cluster Mean $P_i^{*}=\sum_{s=1}^{N}U_{is}\mu_s$ or $P_i^{*}=\mathbf{g}_i'^{*}\mathbf{R}^*$

where $\mathbf{g}_i'^{*}=\mathbf{1}_N'\left(\bigoplus_{s=1}^{N}\left[\mathbf{1}_{M_s}'\otimes\left(\mathbf{e}_i'\otimes\frac{\mathbf{1}_{M_s}'}{M_s}\right)\right]\right)$

Collapsing Random Variables

$$\mathbf{R}^*=\mathbf{A}'\mathbf{Y}+\mathbf{B}'\mathbf{R}^*$$

where $P_i=\mathbf{g}_i'\ \mathbf{A}'\mathbf{Y}$ and $\mathbf{B}'\mathbf{R}^*=\mathbf{0}$

For unbiasedness, it is necessary to have PPS sampling.

Unequal Sampling Fractions, change target parameter.

Predict

$P_i^{\bullet}=\mathbf{g}_i'^{\bullet}\mathbf{Y}^{\bullet}$ where $P_i^{*}=E_{\xi_2}\left(P_i^{\bullet}\right)$

How to Do Collapsing?

Collapse to sample/remainder Cluster totals

$$\begin{pmatrix} \sum_{s=1}^{N} U_{is} f_s M_s \bar{Y}_{sI} \\ \hline \sum_{s=1}^{N} U_{is} \left(1-f_s\right) M_s \bar{Y}_{sII} \end{pmatrix}$$

Collapse to scaled sample/remainder Cluster totals

$$\begin{pmatrix} \sum_{s=1}^{N} U_{is} f_s \bar{Y}_{sI} \\ \hline \sum_{s=1}^{N} U_{is} \left(1-f_s\right) \bar{Y}_{sII} \end{pmatrix}$$

Collapse to sample mean/remainder mean for Clusters

$$\begin{pmatrix} \sum_{s=1}^{N} U_{is} \bar{Y}_{sI} \\ \hline \sum_{s=1}^{N} U_{is} \bar{Y}_{sII} \end{pmatrix}$$

RP-Balance $\hat{T}_i = f\mathbf{e}_i'\bar{\mathbf{Y}}_I + \left(1-f\right)\mathbf{e}_i'\left(\mathbf{1}_n\bar{Y} + k\left(\bar{\mathbf{Y}}_I - \mathbf{1}_n\bar{Y}\right)\right)$

Predict Totals (use totals)

$$\hat{P}_i^{\circ} = \mathbf{e}_{iI}'\mathbf{Y}_I^{\circ} + \left(\frac{1-f}{f}\right)\mathbf{e}_{iI}'\left(\mathbf{1}_n\left(\frac{\mathbf{1}_n'\mathbf{Y}_I^{\circ}}{n}\right) + fk^{\circ}\left(\mathbf{Y}_I^{\circ} - \mathbf{1}_n\left(\frac{\mathbf{1}_n'\mathbf{Y}_I^{\circ}}{n}\right)\right)\right)$$

Predict Means (use weighted totals)(with PPS)

$$\hat{P}_i^{*} = \mathbf{e}_{iI}'\mathbf{Y}_I^{*} + \left(\frac{1-f}{f}\right)\mathbf{e}_{iI}'\left(\mathbf{1}_n\left(\frac{\mathbf{1}_n'\mathbf{Y}_I^{*}}{n}\right) + fk^{*}\left(\mathbf{Y}_I^{*} - \mathbf{1}_n\left(\frac{\mathbf{1}_n'\mathbf{Y}_I^{*}}{n}\right)\right)\right)$$

Predict Means (use means)(for non-PPS sampling)

$$\hat{P}^{\bullet} = c\mathbf{e}_{iI}'\mathbf{Y}_I^{\bullet} + \mathbf{e}_{iI}'\left[\left(1-c\right)\mathbf{1}_n\left(\frac{\mathbf{1}_n'\mathbf{Y}_I^{\bullet}}{n}\right) + ck^{\bullet}\left(\mathbf{Y}_I^{\bullet} - \mathbf{1}_n\left(\frac{\mathbf{1}_n'\mathbf{Y}_I^{*}}{n}\right)\right)\right]$$

RP σ^2 σ_s^2 σ_e^2 $k=\frac{\sigma^{*2}}{v}$ $v=\sigma^{*2}+\frac{\sigma_e^2}{m}$ $\sigma^{*2}=\sigma^2-\frac{\sigma_e^2}{M}$

Totals $k^\circ=\frac{\sigma_{\tau+}^2}{v^\circ}$ $v^\circ=f\sigma_{\tau+}^2+\sigma_{e+}^2$ $\sigma_{\tau+}^2=\sigma_\tau^2-\sigma_{e+}^2$

$\tau_s=M_s\mu_s$ $\sigma_\tau^2=\frac{\sum_{s=1}^{N}\left(\tau_s-\bar{\tau}\right)^2}{N-1}$ $\sigma_{e+}^2=\frac{\sum_{s=1}^{N}M_s\sigma_s^2}{N}$

Means (PPS) $k^*=\frac{\sigma^{*2}}{v^*}$ $v^*=f\sigma^{*2}+\sigma_e^{*2}$ $\sigma^{*2}=\sigma^2-\sigma_e^{*2}$

$\sigma_e^{*2}=\frac{1}{N}\sum_{s=1}^{N}\frac{\sigma_s^2}{M_s}$

Means (non-PPS) $k^\bullet=\frac{\sigma^{*2}}{v^\bullet}$ $v^\bullet=\sigma_{el}^{\bullet 2}+\sigma^{*2}$ $\sigma_{el}^{\bullet 2}=\frac{1}{N}\sum_{s=1}^{N}\frac{\sigma_s^2}{m_s}$

Accounting for a Covariate

Basic Idea: (Wenjun Li's Dissertation)

Use Seemingly Unrelated Regression Models to represent the ''Usual' N Random Variables for y and x

Simplest Example:

SRS w/o rep.

Focus on estimating population total

x= 0 or 1

Pop total for x could be known, or unknown

With Known X Totals

Model $$\begin{pmatrix}\mathbf{Y}\\ \mathbf{X}\end{pmatrix} = \left(\mathbf{I}_2 \otimes \mathbf{1}_N\right)\boldsymbol{\mu} + \mathbf{E}$$

Transform X $$\mathbf{X} = \mathbf{P}_N\mathbf{X} + \mathbf{1}_N \mu_x$$

$$\mathbf{P}_N\mathbf{X} = \left(\mathbf{X} - \mathbf{1}_N \mu_x\right)$$
$$= \mathbf{X}^*$$

$$\begin{pmatrix}\mathbf{Y}\\ \mathbf{X}^*\end{pmatrix} = \left(\mathbf{I}_2 \otimes \mathbf{1}_N\right)\begin{pmatrix}\mu_y\\ 0\end{pmatrix} + \mathbf{E}^*$$

Predictor of Total $$\hat{P} = n\bar{Y} + \left(N-n\right)\left[\bar{Y} - \left(\frac{N}{N-n}\right)k\left(\bar{X} - \mu_x\right)\right]$$

where $$k = \frac{\sigma_{xy}}{\sigma_x^2}$$

SRS- Further Extensions

Question: Can we use the predictor of a random variable at a position as an estimator of a unit?

Godambe's Result: $(N-1)^n$ random variables

Royall (1969): N random variables

Expanded RVs: $(N-1)^2$ random variables

Example: N=3, n=2. Use a predictor of a position to predict unit=1

			Permutation					
Position (i)	Unit (j)	Response	abc	acb	bac	bca	cab	cba
1	1	Y11	1	1	0	0	0	0
2	1	Y21	0	0	1	0	1	0
3	1	Y31	0	0	0	0	0	0
1	2	Y12	0	0	0	0.5	0	0
2	2	Y22	0	0	0	0	0	0.5
3	2	Y32	0	0	0	0	0	0
1	3	Y13	0	0	0	0	0	0.5
2	3	Y23	0	0	0	0.5	0	0
3	3	Y33	0	0	0	0	0	0

Source: c02ed53.xls

Example: N=3, n=2. Use a predictor of a position to predict unit=1

			Permutation					
Position (i)	Unit (j)	Response	abc	acb	bac	bca	cab	cba
1	1	Y11	1	1	0	0	0	0
2	1	Y21	0	0	1	0	1	0
3	1	Y31	0	0	0	0	0	0
1	2	Y12	0	0	0	0.5	0	0
2	2	Y22	0	0	0	0	0	0.5
3	2	Y32	0	0	0	0	0	0
1	3	Y13	0	0	0	0	0	0.5
2	3	Y23	0	0	0	0.5	0	0
3	3	Y33	0	0	0	0	0	0

Source: c02ed53.xls

Interpreted as the predictor of a unit, this isn't a linear combination of the expanded random variables.

We need to allow for the coefficient for other units to change depending upon whether or not the unit of interest is in the sample.

g	Unit (j) not in Sample	Unit (j) in Sample	Position (i)	Unit (j)	Response (Z)	abc	acb	bac	bca	cab	cba
1	0	1	1	1	Vs(1)U(11)y1	1	1	0	0	0	0
1	0	1	2	1	Vs(1)U(21)y1	0	0	1	0	1	0
0	0	1	3	1	Vs(1)U(31)y1	0	0	0	0	0	0
0	0	1	1	2	Vs(1)U(12)y2	0	0	0	0	0	0
0	0	1	2	2	Vs(1)U(22)y2	0	0	0	0	0	0
0	0	1	3	2	Vs(1)U(32)y2	0	0	0	0	0	0
0	0	1	1	3	Vs(1)U(13)y3	0	0	0	0	0	0
0	0	1	2	3	Vs(1)U(23)y3	0	0	0	0	0	0
0	0	1	3	3	Vs(1)U(33)y3	0	0	0	0	0	0
0	1	0	1	1	Vr(1)U(11)y1	0	0	0	0	0	0
0	1	0	2	1	Vr(1)U(21)y1	0	0	0	0	0	0
1	1	0	3	1	Vr(1)U(31)y1	0	0	0	0	0	0
0	1	0	1	2	Vr(1)U(12)y2	0	0	0	0.5	0	0
0	1	0	2	2	Vr(1)U(22)y2	0	0	0	0	0	0.5
0	1	0	3	2	Vr(1)U(32)y2	0	0	0	0	0	0
0	1	0	1	3	Vr(1)U(13)y3	0	0	0	0	0	0.5
0	1	0	2	3	Vr(1)U(23)y3	0	0	0	0.5	0	0
0	1	0	3	3	Vr(1)U(33)y3	0	0	0	0	0	0

Partition into a sample and remainder

Collapse random variables similarly to unequal cluster size problems

Results with N=3 and n=2 lead to a non-unique solution. ???

Maybe larger N and n are needed?

Maybe solutions will always be non-unique?

Summary of Ideas

- If first stage is exchangeable, second stages can not easily retain nesting of units in the first stage unit. Identifiability of 1st stage unit matters.
- By expanding the random variables, units can be tracked. High dimension singular variance structures result, and the problem needs to be ‘reduced’.
- Projections to lower dimensional spaces appear ‘arbitrary’, but seem to ‘work’.
- Extending the expanded random variables may enable a unit to be identified, and estimated. Our hope is that we obtain an estimator that equals the predictor of a position.
- A variety of variance components need to be estimated to implement the results.