# Using Hierarchical Models to Estimate a Weighted Average of Stratum-specific Parameters

Babette Brumback
*in collaboration with*
Larry Winner, George Casella, Malay Ghosh, Allyson Hall, and Paul Duncan
University of Florida

# Outline:

- Motivating Example: Medicaid Reform
- Standard Estimators
- Using Hierarchical Models to Generate Adaptive Estimators
- The DerSimonian and Laird (1986) Model versus a Novel Alternative
- "Optimal" Estimation with Two Strata
- A Brief Comparison with Three Strata
- Application to the Medicaid Data

# Motivating Example:

Many applications of statistics involve estimating a weighted average of stratum-specific parameters:

$$\theta^s = \sum_{i=1}^{n} w_i \theta_i$$

$n:$ Number of strata

$w_i:$ Weights, positive, sum to one

$\theta_i:$ Stratum-specific parameters

# Motivating Example, continued

Our interest is in **Medicaid Reform Evaluation**: see ahca.myflorida.com/Medicaid.

A stratified sample, by health plan, of Medicaid participants in two Florida counties, Duval and Broward, will be surveyed for satisfaction with health care

# Motivating Example, continued

We are interested in plan-to-plan comparisons and also in aggregated summaries.

Estimating the aggregated summaries is the focus of this talk.

Equal sample sizes were selected for each health plan – the plans vary in size.

# Motivating Example, continued

For illustration, we focus on estimating the average age of Medicaid participants in Duval County:

| Plan | Size | PopMean | SampMean | SE |
|---|---|---|---|---|
| 1 | 2364 | 17.61 | 16.46 | 0.84 |
| 2 | 22126 | 17.16 | 17.33 | 0.97 |
| 4 | 655 | 9.49 | 9.75 | 0.30 |
| 6 | 5005 | 14.14 | 14.05 | 0.69 |
| 10 | 26812 | 15.54 | 15.48 | 0.81 |

# Standard Estimators

Probably the most common estimator of $\theta^s$ is the weighted sum of stratum means $Y_i$ , i=1,…,n:

$$\hat{\theta}_a^s = \sum_{i=1}^{n} w_i Y_i$$

# Standard Estimators

But if the $\theta_i$ are relatively homogeneous, then an estimator with lower mean-squared error weights the sample means proportionally to their inverse-variances:

$$\hat{\theta}_b^s = \frac{\sum_{i=1}^{n} \frac{Y_i}{\sigma_i^2}}{\sum_{i=1}^{n} \frac{1}{\sigma_i^2}}$$

# Test then Estimate

In practice, homogeneity is not known *a priori*, and thus one strategy is to test the null hypothesis of homogeneity and then to select either $\hat{\theta}_a^s$ or $\hat{\theta}_b^s$ depending on whether the test rejects or not.

# Test then Estimate

For the Medicaid Data, the test of homogeneity rejects with a p-value of less than 0.001. However, when just plans 1, 2, and 10 are evaluated in Duval County, the test does not reject, based on a p-value of 0.33.

Thus we choose $\hat{\theta}_a^s$ for the complete summary, and $\hat{\theta}_b^s$ for the 3 plan summary.

# Adaptive Estimators

We focus on a general class of adaptive estimators based on hierarchical models. These encompass $\hat{\theta}_a^s$ and $\hat{\theta}_b^s$ as limits. We model the $\theta_i$ as random and select the variance parameters to minimize mean squared error.

# The Hierarchical Model

$$Y \mid \theta, \mu, \delta \sim N(\theta, \Sigma)$$

$$\theta \mid \mu, \delta = Z_W \mu + \delta$$

$$\delta \sim N(h_W, T_W)$$

$$\mu \sim U^p(-\infty, \infty)$$

$$\mu \text{ X } \delta$$

$$Y = (Y_1, \ldots, Y_n)^T$$

$$\theta = (\theta_1, \ldots, \theta_n)^T$$

$$W = (w_1, \ldots, w_n)^T$$

# Posterior means

$$\hat{\delta} = (A + T_W^{-1})^{-1}(AY + T_W^{-1}h)$$

$$A = \Sigma^{-1}(I - P_Z^{\Sigma})$$

$$P_Z^{\Sigma} = Z(Z^T\Sigma^{-1}Z)^{-1}Z^T\Sigma^{-1}$$

$$\hat{\mu} = (Z^T\Sigma^{-1}Z)^{-1}Z^T\Sigma^{-1}(Y - \hat{\delta})$$

$$\hat{\theta}^s = W^T(Z\hat{\mu} + \hat{\delta}) \quad \text{(admissible)}$$

# Posterior means

Note :

$$Z^{T} T_{W}^{-1} (\hat{\delta} - h) = 0$$

This constraint influences our interpretation for μ, analogous to how constraints in models with nonrandom μ and δ would do.

# Model 1

Model 1 corresponds to the DerSimonian and Laird (1986) model:

$$Z = 1_n$$

$$\Sigma = \text{diagonal} \ (\sigma_1^2, \ldots, \sigma_n^2)$$

$$h = 0$$

$$T_W = \tau^2 I_n$$

# Model 2

Model 2 was briefly touched on by Ghosh and Maiti (1998) and Brumback and Brumback (2005):

$$Z = 1_n$$

$$\Sigma = \text{diagonal}\ (\sigma_1^2, \ldots, \sigma_n^2)$$

$$h = 0$$

$$T_W = \text{diagonal}\ \left(\frac{\tau^2}{w_1}, \ldots, \frac{\tau^2}{w_n}\right)$$

# Model 1 Estimator

$$\hat{\mu} = \left( \sum_{i=1}^{n} \gamma_i Y_i \right) / \left( \sum_{i=1}^{n} \gamma_i \right)$$

$$\gamma_i = \frac{1}{\tau^2 + \sigma_i^2}$$

$$\hat{\delta}_i = \tau^2 \gamma_i (Y_i - \hat{\mu}); \quad \sum_{i=1}^{n} \hat{\delta}_i = 0$$

$$\hat{\theta}_1^s = \sum_{i=1}^{n} \eta_i Y_i$$

$$\eta_i = \gamma_i \left( \tau^2 w_i + \frac{\sum_{j=1}^{n} \gamma_j \sigma_j^2 w_j}{\sum_{j=1}^{n} \gamma_j} \right)$$

# Model 2 Estimator

$$\hat{\mu} = \left( \sum_{i=1}^{n} \gamma_i Y_i \right) / \left( \sum_{i=1}^{n} \gamma_i \right)$$

$$\gamma_i = \frac{1}{\frac{\tau^2}{w_i} + \sigma_i^2}$$

$$\hat{\delta}_i = \frac{\tau^2}{w_i} \gamma_i (Y_i - \hat{\mu}); \quad \sum_{i=1}^{n} w_i \hat{\delta}_i = 0$$

$$\hat{\theta}_2^s = \hat{\mu}$$

# Limits as $\tau^2$ tends to 0 or $\infty$

Both $\hat{\theta}_1^s$ and $\hat{\theta}_2^s \rightarrow \hat{\theta}_a^s$ or $\hat{\theta}_b^s$ as $\tau^2 \rightarrow \infty$ or 0.

By selecting $\tau^2$ to minimize MSE, we induce both estimators to adapt towards $\hat{\theta}_a^s$ or $\hat{\theta}_b^s$ depending on the relative heterogeneity or homogeneity of the stratum-specific means.

# The Role of μ in Models 1 and 2

In Model 1, we can interpret $\mu$ as $\bar{\theta}$ and $\delta_i$ as $\theta_i - \bar{\theta}$.

In Model 2, we can interpret $\mu$ as $\theta^s$ and $\delta_i$ as $\theta_i - \theta^s$.

On first thought, this makes Model 2 ideal for our goal of estimating $\theta^s$.

However, one might use the data to negotiate between Model 1 and Model 2.

# "Optimal" Estimation with n=2

A natural question is whether we can use our adaptive estimators to approximately minimize MSE over all estimators of the form $\sum_{i=1}^{n} v_i Y_i$.

We consider the question for n = 2.

The MSE is minimized when

$$v_1 = v_1^* = \frac{2w_1 s_\theta^2 + \sigma_2^2}{2s_\theta^2 + (\sigma_1^2 + \sigma_2^2)}; \quad s_\theta^2 = (\theta_1 - \theta_2)^2 / 2.$$

# “Optimal” Estimation with n=2

$\hat{\theta}_1^s$ minimizes MSE when $\tau^2 = s_\theta^2$

$\hat{\theta}_2^s$ minimizes MSE when $\tau^2 = 2s_\theta^2 w_1(1 - w_1)$

Therefore, the two adaptive estimators coincide when n = 2.

In practice, $s_\theta^2$ must be estimated, and the resulting "risk - minimizing" estimator loses its optimality.

# A Brief Comparison with n=3

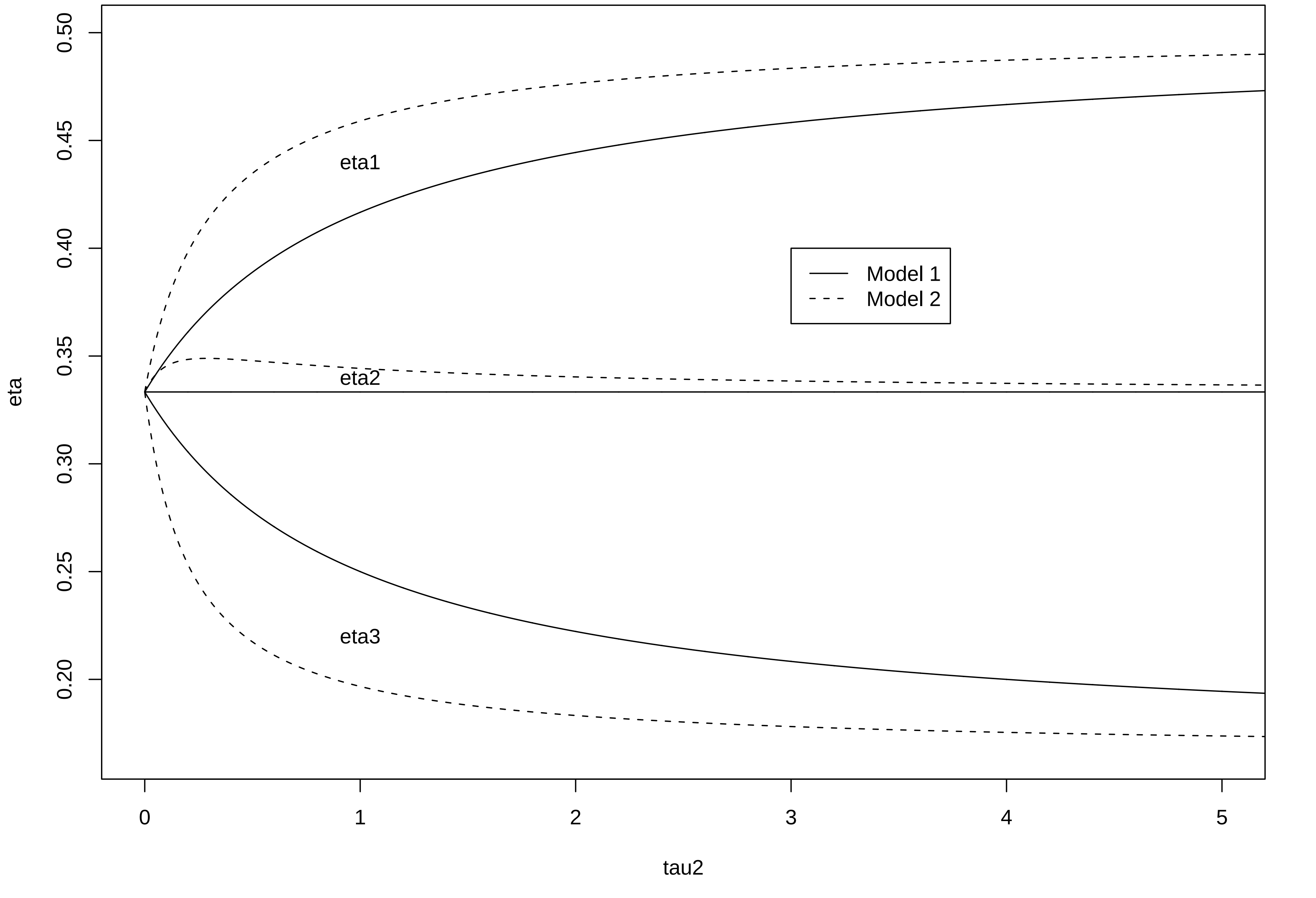
$\hat{\theta}_1^s$ and $\hat{\theta}_2^s$ will not coincide in general.

For n = 3, we provide a counterexample to the claim that one can always make a data - based choice of $\tau_1^2$ for Model 1 and $\tau_2^2$ for Model 2 such that the resulting estimators are identical.

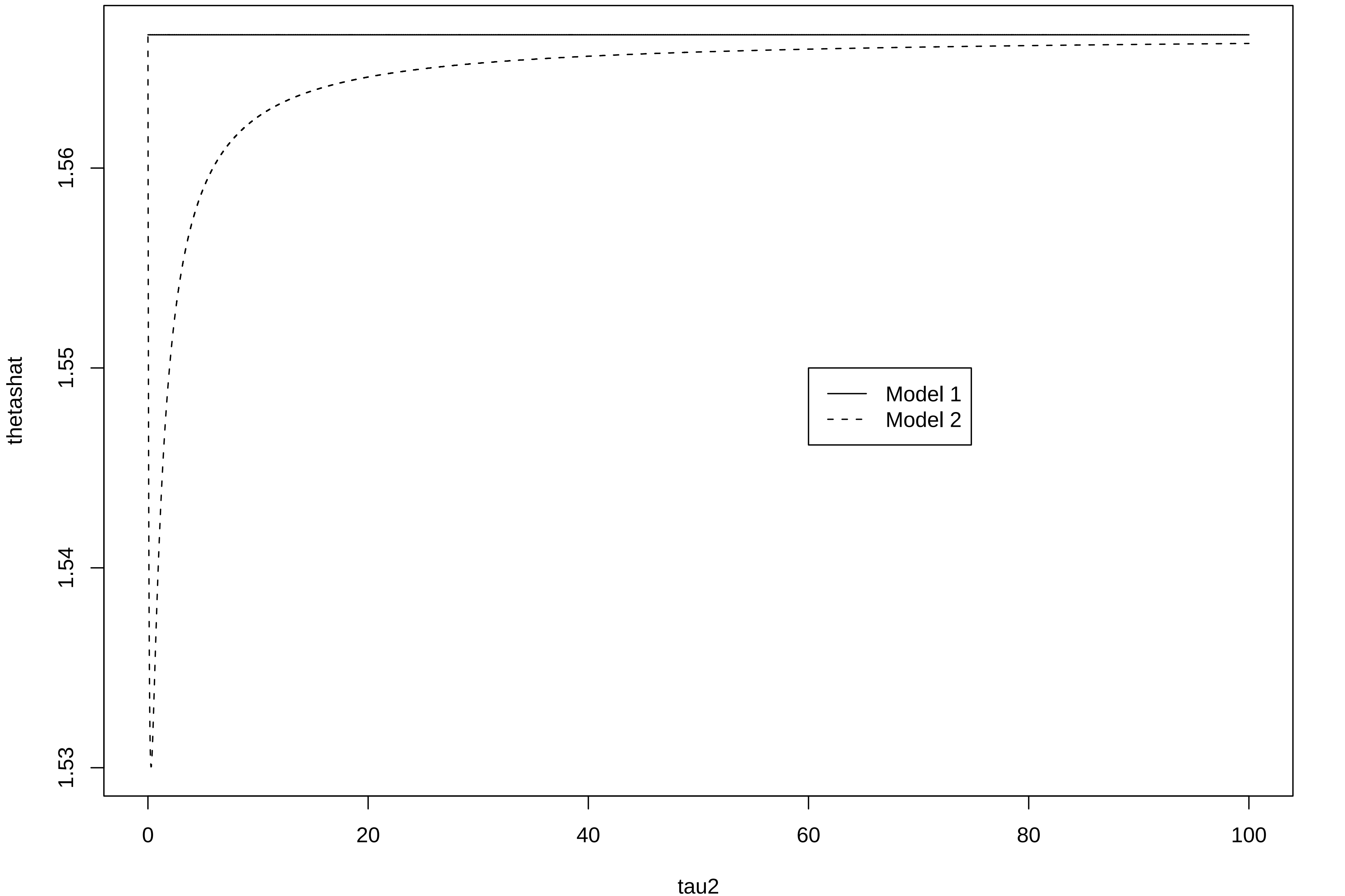
**Counterexample Dataset :**

$(w_1, w_2, w_3) = (1/2, 1/3, 1/6)$, all $\sigma_i^2 = 1$, and

$(Y_1, Y_2, Y_3) = (2.35, 0, 2.35)$



0.50
0.45
0.40
0.35
0.30
0.25
0.20
eta
eta1
eta2
eta3
Model 1
Model 2
0
1
2
3
4
5
tau2



1.56
1.55
1.54
1.53
thetashat
0
20
40
60
80
100
tau2
Model 1
Model 2

# Application to Medicaid Data

Recall, for Duval County:

| Plan | Size | PopMean | SampMean | SE |
|---|---|---|---|---|
| 1 | 2364 | 17.61 | 16.46 | 0.84 |
| 2 | 22126 | 17.16 | 17.33 | 0.97 |
| 4 | 655 | 9.49 | 9.75 | 0.30 |
| 6 | 5005 | 14.14 | 14.05 | 0.69 |
| 10 | 26812 | 15.54 | 15.48 | 0.81 |

# Application to Medicaid Data

| Summary | Estimator | Estimate | Simulated MSE |
|---|---|---|---|
| Duval | $\theta^s$ (truth) | 16.0634 | NA |
| | $\hat{\theta}_a^s$ | 16.0453 | 0.2909 |
| | $\hat{\theta}_b^s$ | 11.8086 | 18.7206 |
| | $\hat{\theta}_1^s$ | 15.8641 | 0.3542 |
| | $\hat{\theta}_2^s$ | 15.9527 | 0.3168 |

# Application to Medicaid Data

| Summary | Estimator | Estimate | Simulated MSE |
|---|---|---|---|
| Duval | $\theta^s$ (truth) | 17.2049 | NA |
| plans 1 & 2 | $\hat{\theta}_a^s$ | 17.2420 | 0.7690 |
| | $\hat{\theta}_b^s$ | 16.8354 | 0.4445 |
| | $\hat{\theta}_1^s$ | 16.8354 | 0.5045 |
| | $\hat{\theta}_2^s$ | 16.8354 | 0.5045 |

# Summary

The adaptive estimators are especially useful when prior knowledge indicates that the parameters might be relatively homogeneous.

In our investigation, we uncovered the importance that the constraints satisfied by posterior means have for interpretability of the direct parameters (i.e. $\mu$) of hierarchical models.

# Summary

The estimators based on Model 1 and Model 2 performed similarly in our application, even though they are theoretically distinct when $n \geq 3$. Will they usually be similar in practice?

We considered $\theta_i$ that are stratum means, but one could generalize to stratum odds ratios, etc.

We considered known $w_i$, but one could generalize to estimated $w_i$ – useful for adjustment for nonresponse bias, missing data, or confounding.

# Acknowledgements

We thank the Florida Agency for Health Care Administration for providing the Medicaid Data.

# References


Brumback, B.A. and Brumback, L.C. (2005). Comment on "Semiparametric Estimation of a Treatment Effect in a Pretest-Posttest Study with Missing Data," by Davidian M, Tsiatis A., and Leon S, *Statistical Science*, 20, 284-289.

Brumback, B.A., Winner, L.H., Casella, G, Ghosh, M, Hall, A., and Duncan, P. (2007). "Using Hierarchical Models to Estimate a Weighted Average of Stratum-specific Parameters." *submitted manuscript*.

DerSimonian, R. and Laird, N. (1986). Meta-Analysis in Clinical Trials. *Controlled Clinical Trials*, 7, 177-188.

Ghosh, M. and Maiti, T (1998). Discussion on the Papers by Firth and Bennett and Pfeffermann et al. *Journal of the Royal Statistical Society, Series B*, 60, 41-56.