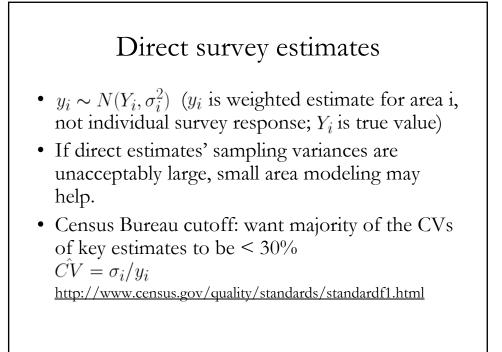
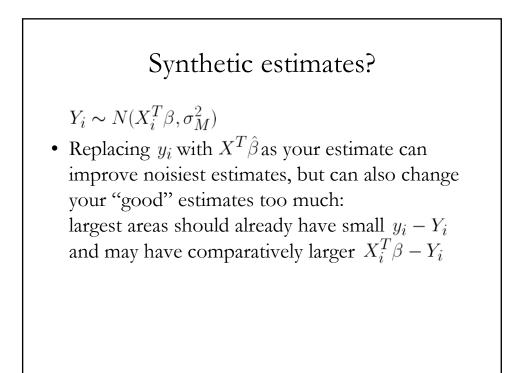
Computer Applications for Small Area Estimation, Part 1

Jerzy Wieczorek Small Area Estimation Research Group, CSRM 1/17/2013

Outline

- Basic approach for continuous area-level data
- The estimates we need and how to get them
- SAS example "by hand" with **PROC REG** and **DATA** steps
- SAS example "automated" with PROC MIXED
- Further resources
- Plan for next time





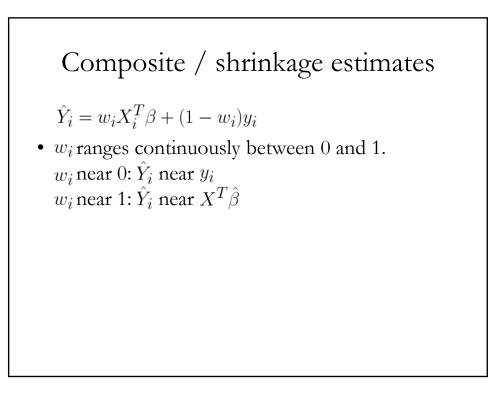
Direct for largest, else synthetic?

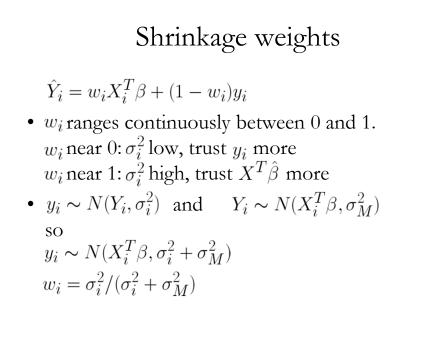
 $\hat{Y}_i = w_i X_i^T \beta + (1-w_i) y_i$

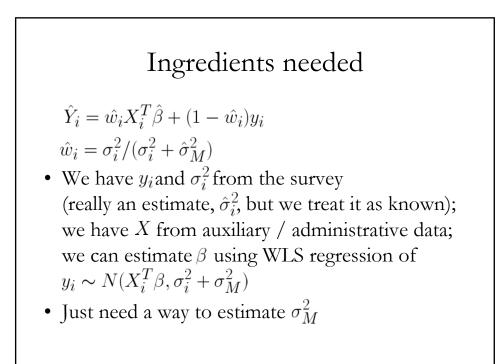
• Each w_i is 0 or 1.

 $w_i = 0$: use direct estimate, ignore regression

 $w_i = 1$: use regression, ignore direct estimate

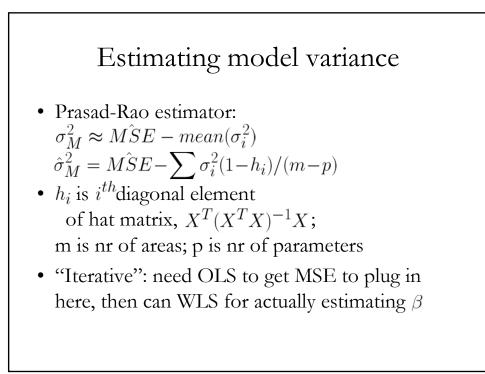


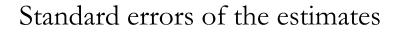




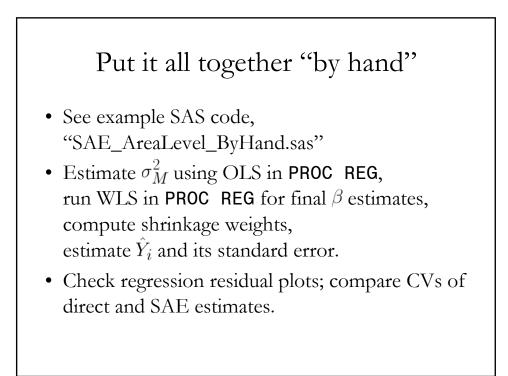
Estimating model variance

- Several good estimators; "REML" usually best
- "Prasad-Rao" simpler (for illustration only!)
- Under model $y_i \sim N(X_i^T \beta, \sigma_i^2 + \sigma_M^2)$ regression MSE estimates average of $\sigma_i^2 + \sigma_M^2$ $\hat{MSE} \approx mean(\sigma_i^2 + \sigma_M^2) \approx mean(\sigma_i^2) + \sigma_M^2$ $\sigma_M^2 \approx \hat{MSE} - mean(\sigma_i^2)$
- Adjust for estimation of β too



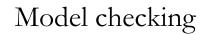


- Standard errors of the new estimates should account for σ_i^2 , σ_M^2 , and estimation of β $g_1 = w_i \sigma_M^2 = \sigma_M^2 \sigma_i^2 / (\sigma_i^2 + \sigma_M^2)$ $g_2 = w_i^2 Var(X^T \hat{\beta})$ $\widehat{MSE}(\hat{Y}_i) = g_1 + g_2$
- More advanced: can also add g₃, a term to account for estimation of σ²_M; see Rao (2003)



"Automate it" with **PROC** MIXED

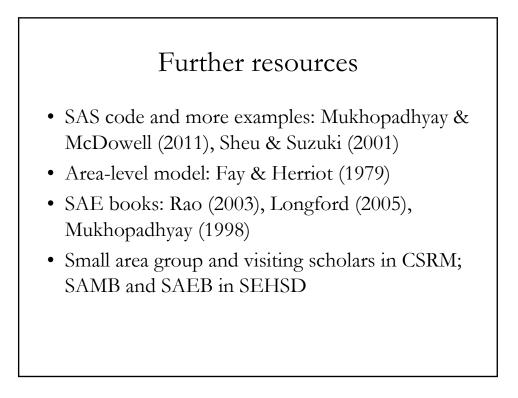
- See example SAS code, "SAE_AreaLevel_ProcMixed.sas"
- Put in an initial guess for σ_M^2 , then let **PROC** MIXED estimate it for you using REML
- **PROC** MIXED also produces the shrinkage estimates and their standard errors for you
- Check plots of Marginal Studentized Residuals



- Shrinkage weights: \hat{w}_i are not all near 0 or 1?
- Model variance: if $\hat{\sigma}_M^2$ too close to 0, σ_i^2 may be overestimates
- Raking factors: is sum of county-level SAE estimates close to state-level direct estimate?
- Compare to a "truth deck" (full census or admin data): check if point estimates and MSE are good, CI coverage is nominal, etc.

Complications

- Your data are not normal as given, but are approximately normal on a transformed scale? log(y_i) ~ N(X_i^Tβ, σ_i² + σ_M²) Then need to correct for bias when transforming estimates back to original scale.
- Your data are not normal at all, but rather binomial, Poisson, etc.? Hierarchical Bayes modeling is more flexible (next time!)



Bibliography

- Fay, R.E., and Herriot, R.A. (1979). Estimates of income for small places: an application of James-Stein procedures to census data. Journal of the American Statistical Association, 74, 269-277. http://www.jstor.org/stable/10.2307/2286322
- Longford, N.T. (2005). Missing Data And Small-area Estimation: Modern Analytical Equipment for the Survey Statistician. New York: Springer.
- Mukhopadhyay, P. (1998). Small Area Estimation in Survey Sampling. New Delhi: Narosa Pub House.
- Mukhopadhyay, P.K., and McDowell, A. (2011). Small Area Estimation for Survey Data Analysis Using SAS Software. Proceedings of the SAS Global Forum 2011 Conference. <u>http://support.sas.com/resources/papers/proceedings11/336-2011.pdf</u>
- Rao, J.N.K. (2003). Small Area Estimation. New York: Wiley.
- Sheu, C., and Suzuki, S. (2001). Meta-Analysis Using Linear Mixed Models. Behavior Research Methods, vol. 33, issue 2, 102-107. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.28.9389&rep=rep1&type=pdf