DISCUSSION

Stephen M. Miller, Bureau of Labor Statistics
441 G St. N.W. Room 2126, Washington D.C. 20212

I would like to congratulate all the authors on fine and interesting papers. I will limit my comments to the papers of Hulting and Harville, and Dol and Steerneman.

Bayesian and Non-Bayesian Procedures for Small Area Estimation: Relationships and Computational Aspects

This paper by Hulting and Harville is an interesing discussion of the general problem of estimating, and making inferences about, small area means. They approach the problem from the point of view of mixed linear model theory, and apply some of the results from that theory to the small area problem. The application of mixed linear model theory for the purposes of constructing small area estimates is not a new one, but Hulting and Harville go the aditional step and show us how to apply mixed linear model theory to construct variances and confidence intervals. They also bring in the tools of Bayesian analysis and the use of *credible sets*, and in addition tackle the problem of computational efficiency. I have the following remarks.

Remark 1. The model they assume is the following

$$y_{ij} = \underset{\sim}{x}_{ij}'\underset{\sim}{\beta} + \underset{\sim}{k}_{ij}'\underset{\sim}{v} + e_{ij}\,, \tag{1}$$

where the subscripting refers to the *j-th* unit sampled from the *i-th* small area. Notice that a model is being applied at the *micro* or *unit level*. This implies a model at the *small area level* namely

$$\bar{y}_i = \bar{\underset{\sim}{x}}_i'\underset{\sim}{\beta} + \bar{\underset{\sim}{k}}_i'\underset{\sim}{v} + \bar{\underset{\sim}{e}}_i'\,, \tag{2}$$

where the means are calculated over the n_i units in the *i-th* small area. Many researchers feel more comfortable with the aggregate level model in (2) than the micro level model in (1). A question that should be addressed is to what extent do the results of Hulting and Harville require only the macro level model in (2) instead of the micro level model in (1)? The assumption of normality would seem to play a key role here.

Remark 2. The paper uses the basic idea of contructing *estimated* normal theory based confidence intervals. To what extent is it possible to use replication based inference, such as the Bootstrap, to construct confidence intervals for the small area problem? Such an approach would have to start with some kind of micro level model as in (1), but would presumably have the advantage of treating the underlying *small area distribution* of the units nonparametrically. Dr. Hulting has informed me that they have looked at *parametric Bootstrap* confidence intervals, but I believe the more general problem remains open.

Asymptotic Optimal Bilinear And QR Estimators For Small Area Parameters

This paper by Dol and Steerneman uses the tools of superpopulation models to construct *bilinear* and *QR estimators* for linear combinations of characteristics of units within small areas or subpopulations. Their basic working superpopulation model is

$$\underset{\sim}{y} = \underset{\sim}{x}\underset{\sim}{\beta} + \underset{\sim}{e} \tag{3}$$

where $\underset{\sim}{y}$ is an $N \times 1$ vector of the population characteristics, $\underset{\sim}{x}$ is an $N \times k$ matrix of auxillary variables, and $\underset{\sim}{e}$ is the model error which has mean $\underset{\sim}{0}$ and covariance matrix $\underset{\sim}{I}_N\sigma^2$. Their goal is to estimate (*or predict*) the linear combination $\underset{\sim}{c}'\underset{\sim}{y}$, where $\underset{\sim}{c}'$is a predefined vector of constants. They do a nice job of deriving results in terms of expected designed-based optimality as well as model-based optimality, and overall their results form a nice notational bridge between designed-based theory and *prediction theory*. I have the following remarks.

Remark 1. The authors use the *replica method* to prove their asymptotic results which implies that they are letting the number of units in each small area tend to infinity. Many people have proposed that asymptotics should be done in the small area problem by keeping the number units in each small area bounded, and let the number of small areas tend to infinity. It would be nice to see the authors go back and prove their results in this context.

Remark 2. The authors prove optimality results involving the ordinary least squares estimator by using the fact that $\hat{\underset{\sim}{\beta}}_{ols}$ is the best linear unbiased estimator of $\underset{\sim}{\beta}$ when the covariance matrix of the model error $\underset{\sim}{e}$ is $\underset{\sim}{I}_N\sigma^2$. They go on to say that for more general covariance structures generalized least squares would have to be used. But there exist situations where $\hat{\underset{\sim}{\beta}}_{ols}$ is still the best linear unbiased estimator under covariance structures other than $\underset{\sim}{I}_N\sigma^2$. The authors may want to consider whether this may have any useful consequences for their particular problem, since it implies that there exist situations where their estimators would still be optimal under more general covariance structures.

Remark 3. It was unclear to me how the authors would estimate the variance on their estimators in practice. It is true that expressions were given for the *expected mean square error*, but it is unclear how to estimate the components in those expressions. Part of my confusion comes from the mixing of design-based and model-based theory. On a related subject, have the authors considered developing optimal *biquadratic estimators* to estimate population quadratic forms such as the variance within a small area?

Summary

Both papers approach the problem of small area estimation by using the tools of linear model theory applied at the individual unit level. The Hulting and Harville paper allow for a more complex error structure than do Dol and Steerneman, but the second paper makes more of an attempt to account for the sample design by explicitly making use of the first and second order design inclusion probabilities. I hope all of the authors continue to work in this area, and continue to produce stimulating and thought provoking research. Thank you.