DISCUSSION

Ralph Folsom, Research Triangle Institute

Beginning with the Lorah, Singh, and Tegels paper, I would like to compliment the authors on their careful consideration of stratification and clustering approaches to improving the precision of Black and Spanish statistics for the Current Population Survey. They have presented a convincing case for implementing a within PSU minority housing unit stratification and oversampling scheme combined with a reduction in the average minority housing unit cluster size. Recognizing that such minority directed strategies reduce the design efficiency for total population statistics, the authors have been careful to set constraints on the level of such variance inflation that would be tolerated. It is in this direction that I feel the CPS redesign effort could be made more definitive in terms of a formal survey cost minimization subject to multiple variance constraints set for key minority and total population statistics. I would also recommend consideration of a composite PSU size measure that anticipates the planned oversampling of second stage minority strata. In this coordinated PSU selection and minority housing unit oversampling scheme, minority strata are defined for each primary frame unit, common sampling rates are specified for the minority strata across PSUs, and a composite PSU size measure is then used to select PSUs. The composite size measure consists of a weighted sum of the minority stratum housing unit (HU) cluster counts with weights equal to the desired stratum sampling rates. The planned assignment of m clusters per PSU is then allocated to minority strata within selected PSUs in proportion to the product of the stratums HU cluster count and the associated sampling rate. PSU selections and subsequent cluster allocations based on such a composite size measure yield self-weighting HU samples within the minority strata and equalize the assignment of sample clusters to the noncertainty PSUs.

Concerning the specification of optimum sampling rates for the minority strata, I believe that one should consider the implications of oversampling strategies on the total variance of minority and total populaton statistics across all PSUs. Explicit constraints can then be set for the variances or relative variances of selected Black, Spanish, and Total population statistics. The optimum minority stratum sampling rates are those that minimize survey costs subject to satisfying these explicit precision requirements. While the PSU contributions to these variance functions can be viewed as fixed if the primary sample design is a given, these contributions will have a strong impact on the HU sampling rates required to meet the precision requirements. As fixed quantities relative to the HU sampling rate optimization, these PSU variance contributions would be subtracted from the total variance function to specify the appropriate constraint for the within PSU variance contribution. The empirical investigations presented by Lorah, Singh, and Tegels suggest grouping HUs into Black and NonBlack strata across all PSUs and further stratifying NonBlack HUs by ED to isolate High Percent Spanish areas in selected PSUs with significant Spanish populations. With this design structure in mind, one could use CPS data to parameterize a variance model in terms of a fixed PSU contribution and a separate within PSU variance component divided by the associated global HU sampling rate for each of the three minority strata. The within PSU Spanish stratum variance components would estimate the sum of the separate PSU level quantities aggregated across those frame units designated for Spanish stratification.

The optimum cluster size consideration could also be built into the global variance model with the separate minority HU strata allowed to have different optimum cluster sizes. Armed with survey cost components associated with clusters and housing units within clusters, one could then perform a formal design optimization minimizing variable survey costs subject to specified precision constraints for selected minority and total population statistics. Such a multiple constraint survey cost minimization would seem to be the logical next step in the optimization of minority stratum sampling rates and cluster sizes.

Turning to the Huang and Ernst paper comparing an alternative estimator to the current CPS composite estimator, I would like to commend the authors on an excellent demonstration of how a plausible model can be used to simplify the comparison of two complex composite estimators. Relative to the plausibility of the model, I would have appreciated seeing some evidence of the assumed variance stability and covariance stationarity over time. Noting that the AK composite typically had a smaller variance than the simple composite estimator under a model assuming no rotation group bias, I wondered how efficient the AK composite would prove to be relative to the minimum variance linear unbiased estimator (MVLU) that could be formed by weighted least squares under the specified model. Kirk Wolter's September 1979 JASA paper explores such estimators based on the covariance stationary model for the Census Bureau's new Retail Trade Survey demonstrating that composite type estimators can be very efficient relative to the MVLU estimator. Concerning the comparison of the current 4-8-4 rotation pattern with a 3-9-3 pattern, it appeared that the efficiencies presented for the two estimators were relative to the variance of simple average estimators over the associated eight and six rotation group estimators respectively. I was curious as to what one could say regarding a direct comparison between the two rotation patterns with respect to variance and bias.

Considering the rotation group bias models explored by Huang and Ernst, the assumption that the estimate based on the first month in sample may be the least biased suggests that one consider a simple linear model of the form

$$E\{Y_{hi}\} = \begin{cases} \mu_h & \text{when } i=1 \\ \mu_h + a_{hi} & \text{when } i>1 \end{cases}$$

with $a_{hi} = (\mu_{hi} - \mu_{h1})$ denoting the rotation group bias effects for the second through eighth month in sample. Armed with a 27 month series of these eight rotation group estimates of level (y_{hi}) and consistent sample estimators for the associated variances and covariances, one could use weighted least squares to approximate the minimum variance unbiased estimators for the monthly level parameters. The monthly change and annual average statistics derived from the MVLU level estimates would also be MVLU estimators. With the Census Bureau's computing resources, this would not seem to be an altogether unreasonable solution. One could at least compare the results of such MVLU estimation with the optimum AK composite for a few selected items over a given historical series. If the effect of the two stages of CPS ratio adjustment is ignored in the variance-covariance matrix estimation other than to imbed the adjustments into the sample weights, then a simple variance covariance matrix approximation can be economically produced by computing the between PSU within stratum sums of squares and cross products matrix among the vectors of weighted PSU level month by rotation group totals.