DISCUSSION

J.N.K. Rao, Carleton University
Dept. of Math. & Stat., Carleton University, Ottawa, ON K1S 5H6, Canada

The three papers presented in this session "composite type weighting and estimation in repeated surveys" study important estimation problems in repeated surveys. The paper by Lent, Miller and Cantwell investigates the effect of composite weights on CPS estimates. Singh's paper provides a unified approach of combining information through modified regression(MR). Examples of MR given in this paper cover composite estimation, dual frame estimation, small area estimation and some other estimation problems. The focus of Thompson's presentation is on estimating parameters of change, such as gross flows and level changes, in the presence of nonresponse.

The current CPS methods first adjusts the basic design weights for post-stratification and nonresponse, and then performs A-K composite estimation at a macro level to improve the estimates of level and change. The same (K, A)-coefficient are used for all characteristics to ensure additivity of estimates. But this method can cause difficulty to micro-data users because each month previous months' CPS data are needed to compute composite estimators. In a 1994 Proceedings paper, Lent et al. studied Fuller's method of incorporating composite estimation into micro-data weights to avoid the difficulty with the CPS method. Under this method, "composite weights" are computed for all sample persons, and the composite estimate of level for a particular labor force category is simply obtained as the sum of these weights over that category. The composite weights for each state are obtained through a raking process, using optimal composite estimates and demographic counts as marginal totals. The raking is done separately for each labour force category.

In the present paper, Lent et al. study the effect of CPS month-in-sample bias by calculating the MSEs for selected (K, A) pairs. The variances were calculated using correlations between rotation group estimates (Table 1) and biases from the bias indices given in Table 2. In Tabel 1 the correlation ρ_r for r months apart is taken as zero for months $r = 4$ through $r = 8$. This may not be realistic if rotation is done only within the sample psu's (as in the CPS or the Canadian Labour Force Survey) because the use of same psu's over months induces correlations for all r. It is interesting that the optimal K-values decreased significantly from those computed in their 1994 paper using variance as the criterion.

Lent et al. make another important contribution by looking at the effects of composite weighting on a wide variety of other variables not used in computing the composite weights. The comparisons between the methods are based on average CVs (over all the 2480 characteristics studied and all months). However, such overall measures can be misleading, and it is better to look at the distribution of CVs (such as box plots) as well as other measures such as quartiles. The calculation of CVs is based on a replication method due to Bob Fay, but the authors provide no details on how the pseudo-replicates are formed (jackknife or BRR?) and on other aspects of Fay's method. It would be useful to provide such details because Fay's method and compute programs are not widely known outside the Bureau of the census and the BLS.

As noted before, Singh's paper gives several applications of Modified Regression (MR) including composite estimation. Since the present session is on composite estimation, I will confine myself to commenting on MR-composite which is obtained by regressing the H-T zero function, $\hat{\theta}_{yt}^{HT} - \theta_{yt}$, for the current occasion t on a set of predictor zero functions based on the matched part of the sample as well as auxiliaty data, using a working covariance matrix of all zero functions. But details on the choice of working covariance matrix have not been spelled out. Moreover, the weights associated with MR-composite can take negative values, unlike in the raking method of Lent et al, especially with many predictor zero functions. The paper provides no empirical results on the efficiency of MR-composite relative to AK-composite and other composite estimates. For example, it would be useful to compare MR-composite with the estimator of Fuller and Lent et al.

Thompson's presentation at the session introduced interesting ideas on estimating change parameters in the presence of nonresponse. She discussed methods based on likelihoods and estimating functions and tried to adapt them to survey situations by introducing survey weights. Similar pseudo-likelihood approaches have been used before in the sample survey literature. She has also proposed an extension of bootstrap idea to survey situation for assessing precision in the presence of nonresponse, but its properties remain to be investigated. Thompson assumed no measurement error, but it should be noted that inferences on change parameters are very sensitive to measurement errors. For example, considerable attention has been given to methods of adjusting gross flow estimates for classification error (see e.g., Singh and Rao (1994), JASA, Vol. 90, 478-488).

Finally, my congratulations to all the authors for making excellent contrubutions to problems in repeated surveys.