less other assumptions are made. Therefore, one often restricts attention to linear estimators, $c = a\overline{X} + b$. Within this class, the estimator which minimizes the mean squared error depends only upon the first two prior moments, both of which can often be estimated with $(\overline{X}_1, \ldots, \overline{X}_T)$. The optimal linear estimator is often the same as the unrestricted Bayes estimator derived under a conjugate prior (Rao, 1976). When the conditional distribution of $\overline{X}_i$ is binomial, the optimal linear estimator is a composite estimator,

$$c_i = W_i\overline{X}_i + (1 - W_i)\mu ,$$

where

$$W_i = \sigma^2\{(1 - 1/n_i)\sigma^2 + \mu(1 - \mu)/n_i\}^{-1}$$

and n_i denotes the number of observations from small area i (Spjøtvoll and Thomsen, 1987). With these weights we have that

$$(1)\ E\left\{(1/T)\sum_{i=1}^{T}(c_i - \mu)^2\right\} = \sigma^2(1/T)\sum_{i=1}^{T} W_i \le \sigma^2.$$

It follows that the variation between the small area estimators can be much smaller than the prior known variance. I have often observed this phenomenon in practice; a consequence is usually that the range of the small area estimators is much smaller than expected. (Expectations are based on information outside the sample.) In practice the parameter σ^2 is often of great importance in itself. As said in the introduction, "Increasing concern with issues of distribution, equity and disparity (Brackstone, 1987)." To me, this means that the disparity between the small area is important and should be easily read from a table presenting small-area estimators. As mentioned by Ghosh and Rao, there are composite estimators which have the same expectation and variance as the prior distribution, one of which is simply to use $\{W_i\}^{1/2}$ instead of W_i as weights in the composite estimator.

When area-specific auxiliary information is available and a model like (4.1) in the paper is used, I have often observed a similar "overshrinkage" as under the simpler model above. An inequality similar to (1) can be found under model (4.1), but now σ^2 denotes the variance of the residual in equation (4.1). Again $\{W_i\}^{1/2}$ can be used to avoid "overshrinkage".

Due to the often observed "overshrinkage" and the fact that our models seem too complicated to many of our users of small-area estimators, I have often found it very difficult to make them use the optimal estimators presented in the paper. On the other hand, a number of sample-size dependent estimators are more easily "sold" to the user and therefore more used up until now.

In Statistics Norway a number of administrative registers are available and used to construct small-area estimators. In many cases it is natural to use nested error regression models. However, progress in this area has been slow due to difficulties concerning model diagnostics for linear models involving random effects. I therefore find Section 7.1 particularly interesting and shall use this section intensively in our further hunt for feasible small area estimates.

Rejoinder

M. Ghosh and J. N. K. Rao

We thank the discussants for their insightful comments as well as for providing various extensions of the models and the methods reviewed in our paper. These expert commentaries have brought out many diverse issues and concerns related to small area estimation, particularly on the model-based methods.

Several discussants emphasised the importance of model diagnostics in the context of small area estimation. We agree wholeheartedly with the discussants on this issue. As noted in Section 7.1 of our article, the literature on this topic is not extensive, unlike standard regression diagnostics. We hope that future research on small area estimation will give greater emphasis to model validation issues.

A second concern expressed by some of the discussants is that the composite estimators typically used for small area estimation may "overshrink" towards a synthetic estimator. Thomsen, in his discussion, suggests that a larger weight should be given to the direct estimator. We agree with his suggestion but are hesitant to recommend blanket use of the weight $W_i^{1/2}$, instead of W_i, to the direct estimator $(0 < W_i < 1)$. We believe that the weight should be determined adaptively meeting certain optimality criteria as in Louis (1984) and Ghosh (1992). Cressie and Kaiser, in their discussion, address con-