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

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

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 constrained estimation at some length, emphasising the multivariate aspects of the problem but not invoking any optimality conditions.

Cressie and Kaiser as well as Holt suggest possible extensions of the two basic small area models (4.4) and (4.6) given in Section 4. Their general hierarchical modeling ensures that the variability among observation vectors for the different small areas is attributable not only to sampling variability but also to variability among the associated regression coefficients, β_i. Holt's model looks promising since it allows the β_i to depend on area level auxiliary variables, Z_i, thus effectively integrating the use of unit level and area level covariates into a single model.

A slightly less general version of Cressie and Kaiser's hierarchical model (3) appears in Datta and Ghosh (1991) where a full hierarchical Bayes analysis is presented. In an earlier version of our paper (Ghosh and Rao, 1991) we have in fact considered the general model of Datta and Ghosh but decided to abandon it in the revision in favour of the simpler, but widely used, models (4.4) and (4.5) in order to keep the discussion more accessible to a general readership and the notation simple.

We now turn to some of the specific points raised by the discussants.

CRESSIE AND KAISER

Cressie and Kaiser stress the importance of nonlinear modelling which is especially needed for binary and count data. Our Section 7.3 gives a brief account of logistic regression and log-linear models suitable for such data. These can be viewed as special cases of generalized linear models (McCullagh and Nelder, 1989). Zerger and Karim (1991) have studied generalized linear models with random effects using a Gibbs sampling approach. Their results may be applicable to small area estimation. In a 1993 Ph.D. thesis at the University of Florida, Kannan Natarajan implemented an extensive hierarchical Bayes analysis under generalized linear models in the context of two-stage sampling within small areas. He used the Metropolis within Gibbs sampling algorithm (cf. Müller, 1991). His method is easier to implement than the procedure of Zeger and Karim (1991) due to logconcavity of certain posterior distributions which permits the use of adaptive rejection sampling of Gilks and Wild (1992).

We agree with Cressie and Kaiser regarding the multivariate aspects of small area estimation. Our analysis can be extended to produce approximately unbiased estimators of the off-diagonal elements of the mean-square error matrix as well as to obtain exact posterior covariances of small area means. Reporting these quantities in tables, however, is usually cumbersome since there will be $\binom{m}{2}$ such quantities when the number of small areas is m. Nevertheless, these estimates should be available, as they are needed in calculating measures of uncertainty at higher levels of aggregation.

HOLT

The example in Section 6 of our paper, based on a simple random sample drawn from a synthetic population, was introduced mainly to illustrate the proposed methods. We agree with Holt that a simulation study based on repeated samples from the population is better for comparing the relative performances of estimators. Such a simulation study was, in fact, conducted by Choudhry and Rao (1993) using both real and synthetic populations. Comparisons were made under customary repeated sampling (approach (c) of Holt) as well as under a conditional framework by fixing the values of samples sizes, n_i (approach (b) of Holt).

We also agree with Holt that one should be cautious in comparing relative performances based on summary measures, obtained by averaging across all small areas, without paying some attention to the distribution. Such summary measures, however, may be quite useful in an overall comparison of competing estimators, especially when there is no clear-cut winner when the small areas are judged individually.

SCHAIBLE AND CASADY

Despite many success stories of model-based indirect estimators, there are some practical problems associated with their use. We are grateful to Schaible and Casady for providing a comprehensive list of such problems.

We agree with them that models based on expediency "instill little confidence in either the producers or consumers of the estimators." Model diagnostics should be an integral part of any model-based procedure in order to alleviate this problem.

SINGH

We are glad that Singh has investigated under a simplified model some frequentist properties of the Kass-Steffey first-order approximation (KS-I) to the posterior variance and Hamilton's (1986) Monte Carlo integration method (H) of evaluating the posterior variance. He also suggests modifications, KS-II* and MH, to improve their accuracy; in particular his formula (13) which is a simplified version of the second-order approximation of Kass and Steffey. It would be useful to provide similar improved approximations for more complex models and to study their frequentist properties.

We agree with Singh that the Bayesian approximations KS-II* and MH have the advantage of dual interpretation in both frequentist and Bayesian contexts, although Prasad Rao's estimator of MSE performed better with respect to frequentist properties.

In the case of known variance components, Singh has demonstrated that the linear Bayes estimate (LBE) and its Bayes risk coincide with the BLUP estimator and its MSE. A similar result appears in Datta et al. (1992).

STASNY

Stasny provides an excellent account of USDA's program of country-level estimation of crop and livestock production. She also raises the important issue that the current small area estimation methods need to be modified in the presence of nonresponse. In this regard, Stasny's (1991) important work on hierarchical models for the probabilities of a survey classification and nonresponse might be relevant. We might add that the role of measurement error in small domain estimation is also important. Eltinge and Harter (1990) have studied the effect of measurement errors and propose some modified small area estimators.

We agree with Stasny that in some small area estimation problems historical data can be used to construct informative priors and obtain the resulting hierarchical Bayes estimates.

We are also delighted to learn about the success story that Arkansas is currently using satellite data, in conjunction with USDA survey data, for the production of county estimates based on small area models.

THOMSEN

We agree with Thomsen's observation that many users find the small area models too complicated and are bothered by the overshrinkage problem associated with the optimal estimators. Further work on model diagnostics and constrained estimation and the development of suitable packages to implement both model selection and estimation should alleviate this problem.

Thomsen also remarks that sample-size dependent estimators, such as those based on the weights (3.6), are more easily "sold" to the user. Such estimators are clearly useful and computationally attractive, but their limitations should also be noted. As mentioned in Section 3 of our paper, sample-size dependent estimators can fail to borrow strength from related domains even when the expected domain sample size, $E(n_i)$, is not large enough to make the direct estimators reliable. These estimators were originally designed to handle domains for which $E(n_i)$ is large enough to make the direct estimators satisfy reliability requirements (Drew, Singh and Choudhry, 1982). Another disadvantage of sample-size dependent estimators, noted in Section 3, is that the weights do not take account of the size of between area variation relative to within area variation for the characteristic of interest, unlike model-based estimators. Choudhry and Rao (1993) demonstrate that large efficiency gains can be achieved by using the EBLUP estimators when the between area variation is small relative to within area variation.

ADDITIONAL REFERENCES

ALBERT, J. H. (1988). Computational methods using a Bayesian hierarchical generalized linear model. *J. Amer. Statist. Assoc.* **83** 1037–1044.

ALBERT, J. H. and PEPPLE, P. A. (1989). A Bayesian approach to some overdispersion models. *Canad. J. Statist.* **17** 333–344.

BASS, J., GUINN, B., KLUGH, B., RUCKMAN, C., THORSON, J., and WALDROP, J. (1989). Report of the Task Group for Review and Recommendations on County Estimates, USDA National Agricultural Statistics Service, Washington, DC.

CHOUDHRY, G. H. and RAO, J. N. K. (1993). Evaluation of small area estimators: an empirical study. In *Small Area Statistic and Survey Designs* **1** 271–290. Central Statistical Office, Warsaw.

CRESSIE, N. (1986). Empirical Bayes estimation of undercount in the decennial census. Statistical Laboratory Preprint 86–58. Iowa State Univ., Ames.

CRESSIE, N. (1988). When are census counts improved by adjustment? *Survey Methodology* **14** 191–208.

CRESSIE, N. (1990b). Weighted smoothing of estimated undercount (with discussion). In *Proceedings of Bureau of the Census 1990 Annual Research Conference* 301–325, 362–366. U.S. Bureau of the Census, Washington, DC.

CRESSIE, N. (1992). Smoothing regional maps using empirical Bayes predictors. *Geographical Analysis* **24** 75–95.

ELTINGE, J. L. and HARTER, R. L. (1990). Small domain estimation in the presence of measurement and sampling errors. Technical Report, Dept. Statistics, Texas A&M Univ.

GHOSH, M. and RAO, J. N. K. (1991). Small area estimation: an appraisal. Technical Report 390, Dept. Statistics, Univ. Florida, Gainesville.

GILKS, W. R. and WILD, P. (1992). Adaptive rejection sampling for Gibbs sampling. *J. Roy. Statist. Soc. Ser. C* **41** 337–348.

GOEL, P. K. and DEGROOT, M. H. (1981). Information about hyperparameters in hierarchical models. *J. Amer. Statist. Assoc.* **76** 140–147.

HAMILTON, J. D. (1986). A standard error for the estimated state vector of a state-space model. *J. Econometrics* **33** 387–397.

HOLT, D. and MOURA, F. (1993). Mixed models for making small area estimates. In *Small Area Statistics and Surveys Designs* (G. Kalton, J. Kordos and R. Platek, eds.) **1** 221–231. Central Statistical Office, Warsaw.

IWIG, W. C. (1993). "The National Agricultural Statistics Service County Estimates Program", in "Indirect Estimators in Federal Programs", Statistical Policy Working Paper 21, Report of the Federal Committee on statistical Methodology, Subcommittee on Small Area Estimation, Washington, DC, 7.1-7.15.

LAAKE, P (1978). An evaluation of synthetic estimates of employment. *Scand. J. Statist.* **5** 57–60.

LINDLEY, D. V. and SMITH, A. F. M. (1972). Bayes estimates for the linear model (with discussion). *J. Roy. Statist. Soc. Ser. B* **34** 1–41.

MCCULLAGH, P. and NELDER, J. A. (1989). *Generalized Linear Models*, 2nd ed. Chapman and Hall, New York.

MÜLLER, P. (1991). A generic approach to posterior integration and Gibbs sampling. Technical Report 91–09, Dept. Statistics, Purdue Univ.

PAWEL, D. and FESCO, R. (1988). On the use of correlations in crop yields. *Proceedings of the Section on Survey Research Methodology* 391–396. Amer. Statist. Assoc., Washington, DC.

PFEFFERMANN, D. (1993). The role of sampling weights when modeling survey data. *Internat. Statist. Rev.* **61** 317–337.

RAO, C. R. (1976). Characterization for prior distributions and solutions to a compound decision problem. *Ann. Statist.* **4** 823–835.

SALLAS, W. M. and HARVILLE, D. A. (1981). Best linear recursive estimation for mixed linear models. *J. Amer. Statist. Assoc.* **76** 860–869.

SINGH, A. C., STUKEL, D. M. and PFEFFERMANN, D. (1993). Bayesian versus frequentist measures of uncertainty for small area estimators. *Proceedings of the Section on Survey Research Methods*, Amer. Statist. Assoc. To appear.

SMITH, A. F. M. and ROBERTS, G. O. (1993). Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods (with discussion). *J. Roy. Statist. Soc. Ser. B* **55** 3–23, 53–102.

STASNY, E. A. (1991). Hierarchical models for the probabilities of a survey classification and nonresponse: an example from the National Crime Survey. *J. Amer. Statist. Assoc.* **86** 296–303.

STASNY, E. A., GOEL, P. K. and RUMSEY, D. J. (1991). County estimates of wheat production. *Survey Methodology* **17** 211–225.

TIERNEY, L., KASS, R. E. and KADANE, J. B. (1989). Fully exponential Laplace approximations to expectations and variances of nonpositive functions. *J. Amer. Statist. Assoc.* **84** 710–716.

TUKEY, J. W. (1974). Named and faceless values: An initial exploration in memory of Prasanta C. Mahalanobis. *Sankhyā Ser. A* **36** 125–176.

TUKEY, J. W. (1983). Affidavit presented to District Court, Southern District of New York. *Cuomo et al. versus Baldridge.* 80 Civ. 4550 (JES).

WOLTER, K. M. and CAUSEY, B. D. (1991). Evaluation of procedures for improving population estimates for small areas. *J. Amer. Statist. Assoc.* **86** 278–284.

ZEGER, S. L. and KARIM, M. R. (1991). Generalized linear models with random effects; a Gibbs sampling approach. *J. Amer. Statist. Assoc.* **86** 79–86.

ZEHNWIRTH, B. (1988). A generalization of the Kalman filter for models with state-dependent observation variance. *J. Amer. Statist. Assoc.* **83** 164–167.