Statistical Science
2011, Vol. 26, No. 2, 266–270
DOI: 10.1214/11-STS346REJ
Main article DOI: 10.1214/10-STS346

Rejoinder

J. N. K. Rao

First, I would like to thank the three discussants (Glen Meeden, Joe Sedransk and Eric Slud) for constructive comments on my paper and for providing additional relevant references, particularly on frequentist model diagnostics (Slud) and Bayesian model checking (Sedransk). I totally agree with Sedransk that studying alternative methods of making inference for finite populations is an "underserved field of research." I will first address the constructive comments of the discussants on the comparison of methods for handling sampling errors in the context of estimation with fairly large domain samples. Subsequently, I will respond to the discussions on small area estimation.

HANSEN ET AL. EXAMPLE

In Section 3.2, I cited the well-known Hansen, Madow and Tepping (HMT) example illustrating the dangers of using model-dependent methods with fairly large samples even under minor model misspecifications. Sedransk argues in his discussion that new advances in model diagnostics, such as model averaging, might remedy the difficulty noted by HMT and provide improvements over the "straw man, the usual ratio estimator." I agree with Sedransk that it would be worthwhile analyzing this example and other examples to show how one can make valid model-dependent inferences routinely with fairly large domain samples that can provide significant improvements over the design-based (possibly model-assisted) methods, particularly in the context of official statistics with many variables of interest. If this goal can be achieved, then I believe model-dependent methods (frequentist or Bayesian) will have significant impact on practice, similar to their current use in small area estimation with small domain samples. The HMT example showed the importance of using design weights under their design with deep stratification by size and disproportional sample allocation. The usual design unbiased weighted estimator is almost as efficient as the usual combined weighted ratio estimator under the HMT design because of deep stratification by size, so I do not agree with Sedransk's comment on the importance of ratio estimator in the HMT example. It is interesting to note that under proportional sample allocation, the BLUP estimator (unweighted ratio estimator) under the incorrectly specified ratio model is identical to the combined weighted ratio estimator and hence it performs well because it is design consistent, unlike under disproportional sample allocation. The HMT example demonstrated the importance of design consistency, and in fact as noted in Section 3.2, Little (1983) proposed restricting attention to models that hold for the sample and for which the corresponding BLUP estimator is design consistent. I have noted some limitations of this proposal in Section 3.2. It should be noted that the HMT illustration of the poor performance of the BLUP estimator used the repeated sampling design-based approach to evaluate confidence interval coverage. On the other hand, model-based inference is based on the distribution induced by the model conditional on the particular sample that has been drawn. However, Rao (1997) showed that the HMT conclusions still hold in the conditional framework because of the effective use of size information through size stratification.

ROLE OF DESIGN WEIGHTS

I will now turn to Meeden's useful comments on the role of design weights and the use of Polya posterior (PP) for making inferences after the sample is observed. As noted in Section 4.2, the PP approach when applicable permits routine interval estimation for any finite population parameter of interest through simulation of many finite populations from PP and this general interval estimation feature of PP is indeed attractive. Meeden notes in his discussion that an R package is also available for simulating many complete populations. However, so far the PP methodology considered only simple designs that may satisfy the assumption that the un-sampled units are like the sampled units (exchangeability) which limits its applicability in practice. Meeden agrees with my comment that the PP approach needs extension to more complex designs before it becomes attractive to users. Even for the simple designs where it is applicable, it would be useful

J. N. K. Rao is Distinguished Research Professor, School of Mathematics and Statistics, Carleton University, Ottawa, Ontario K1S 5B6, Canada (e-mail: jrao@math.carleton.ca).