DISCUSSION

J. Sedransk, National Center for Health Statistics, Census Bureau and University of Iowa
Bureau of the Census, Statistical Research Division, Washington, D.C. 20233

This study is an exceptionally thorough comparison of many imputation methods. The results will be useful not only to the Bureau of Labor Statistics, but also to anyone selecting an imputation method for use in an establishment survey where there is a covariate closely related to the variable of interest. The alternative imputation methods that are studied include standard methods such as simple random imputation (within adjustment cells), "nearest neighbor" hot deck and numerous regression-based procedures. There is also a Bayesian approach. In the latter one samples from the predictive distribution which leads naturally to multiple imputations.

There are several concerns about this research. First, nonresponding establishments are defined to be those establishments that do not respond sufficiently quickly. Are such establishments typical of the actual nonrespondents? To have unequivocal results, a methodological study is needed. In such a study the nonrespondents would be identified and the values of the variables of interest obtained from follow-up interviews or from administrative records.

In this paper the emphasis is on the quality of predictions, $(\hat{Y}_{t,i} - Y_{t,i})$ or $|\hat{Y}_{t,i} - Y_{t,i}|$, usually averaged over months (t), establishments (i) and size classes. This is a useful, yet incomplete, comparison: An imputation method having a small mean or mean absolute error may not complete a data set in a manner useful for secondary data analysis. For example, one might wish to know how various imputation methods would fare if the objective were to provide a confidence interval for the population mean using both the observed and imputed values. While the authors do not address this issue directly, they do recognize its importance in special cases (e.g., adding random residuals to regression-based imputation).

In my reading of the complete paper (available from the authors), I wondered about the thoroughness of the modelling of the regression lines. For example, other than quoting values of R^2, I did not see any evidence about how closely the models fit the observed data. A lack of thoroughness in modelling might explain the following somewhat surprising results: (a) The authors claim that using a very small number of adjustment cells is preferable to using a large number of adjustment cells, the adjustment cells being formed using the variable "number of employees," and (b) the much larger (mean absolute) errors corresponding to the larger size classes (Figure 4.4.1 in the complete paper). Here, only one adjustment cell was used. It is possible that more careful modelling is needed, especially for the larger size classes. Would a link among the regression coefficients corresponding to the eight size classes be helpful? One possibility would be to assume that

$$\beta_1,\ldots,\beta_8 \mid \beta^* \overset{iid}{\sim} N(\beta^*,\sigma_\beta^2)$$

and

$$\beta^* \sim \text{locally uniform},$$

or variants of this for (a) all eight regression coefficients, or (b) subsets of the eight regression coefficients.

The authors take as the variable of interest "number of employees," but also use the same variable to define the adjustment cells. Doing this might overstate the value of imputation (relative to surveys where adjustment cells cannot be formed in this way).

Is modelling the nonresponse process important? For the regression-based methods, the principal issue is whether or not there is the <u>same</u> regression for (a) respondents and (b) nonrespondents over the relevant range of $Y_{t-1,i}$ for <u>nonrespondents</u>. From Figure 2.2.1 in the complete version of the paper, it is clear that $f(Y|R)$ and $f(Y|NR)$ are substantially different and, thus, the nonresponse process may be important for <u>some</u> imputation methods.

In any study as large as this one there are bound to be anomalies whose resolution may clarify the roles of the alternative methods. Here is one example. Consider mean absolute error and regression (R) imputation (Table 4.2.1 in the complete paper) vs. regression plus random residual (RR) imputation (Table 5.1 with the residuals being distributed as $N(0,s^2)$). Going from R to RR there are large increments in mean absolute error for models 2, 4 and 8, but not for model 6. Why?

In summary, this is a careful, informative study. However, additional work is needed to resolve some of the questions that could not be addressed because of inadequacies in the data set or a lack of time to investigate all possible aspects of the problem.