USE OF GENERALIZED VARIANCE FUNCTION MODELS IN INFERENCE FROM SOCIAL AND ECONOMIC SURVEY DATA

J. L. Eltinge[1], D. S. Jang[2], M.J. Cho1

ABSTRACT

Analyses of data from social and economic surveys sometimes use generalized variance function models to approximate the design variance of point estimators of population means and proportions. Analysts may use the resulting standard error estimates to compute associated confidence intervals or test statistics for the means and proportions of interest. In comparison with design-based variance estimators computed directly from survey microdata, generalized variance function models have several potential advantages, including operational simplicity; increased stability of standard errors; and, for cases involving public-use datasets, reduction of disclosure limitation problems arising from the public release of stratum and cluster indicators.

These potential advantages, however, may be offset in part by several inferential issues. First, the properties of inferential statistics based on generalized variance functions (e.g., confidence interval coverage rates and widths) depend heavily on the relative empirical magnitudes of the components of variability associated, respectively, with:

(a) the random selection of a subset of items used in estimation of the generalized variance function model;
(b) the selection of sample units under a complex sample design;
(c) the lack of fit of the generalized variance function model; and
(d) the generation of a finite population under a superpopulation model.

Second, under conditions, one may link each of components (a) through (d) with different empirical measures of the predictive adequacy of a generalized variance function model. Consequently, these measures of predictive adequacy can offer us some insight into the extent to which a given generalized variance function model may be appropriate for inferential use in specific applications.

Some of the proposed diagnostics are applied to data from the U.S. Survey of Doctoral Recipients and the U.S. Current Employment Survey. For the Survey of Doctoral Recipients, components (a), (c) and (d) are of principal concern. For the Current Employment Survey, components (b), (c) and (d) receive principal attention, and the availability of population microdata allow the development of especially detailed models for components (b) and (c).

KEY WORDS: Current Employment Survey; Design-based inference; Misspecification effect; Public-use dataset; Superpopulation; Survey of Doctoral Recipients.

---

[1] U.S. Bureau of Labor Statistics, Eltinge_J@bls.gov
[2] Mathematica Policy Research, U.S.A.