

Estimation of Totals and Regression Parameters by Combining Information from Two Independent Surveys

J. N. K. Rao

Carleton University, Ottawa, Canada jrao@math.carleton.ca

Jae Kwang Kim

Iowa State University, Ames, Iowa, USA jkim@iastate.edu

Combining data from two or more independent surveys is frequently encountered in practice. We consider two scenarios of combining information from two independent surveys to make inferences on parameters of interest. In scenario 1, a large sample from survey 1 observes only auxiliary variables related to a variable of interest and a much smaller sample from survey 2 observes both the variables of interest and the auxiliary variables. We generate synthetic values of the variable of interest by fitting a working model to survey 2 data and then predicting the variable of interest associated with the auxiliary variables observed in survey 1. A projection estimator of a total or a domain total is simply obtained from survey 1 weights and associated synthetic values reported in survey 1 data file. For estimating distribution functions and quantiles, we generate synthetic values by stochastic imputation including fractional imputation. Replication variance estimators are obtained by augmenting the synthetic data file for survey 1. In scenario 2, regression analysis is studied when some of the predictor variables of interest are observed in a different survey. Instrumental variables and fractional imputation are used to make inference on the regression parameters.

Key Words: Fractional imputation, instrumental variables, synthetic values, regression analysis