A semiparametric inference to regression analysis with missing covariates in survey data

Shu Yang and Jae Kwang Kim

Department of Statistics, Iowa State University

Abstract: Parameter estimation in parametric regression models with missing covariates is considered under a survey sampling setup. Under missing at random, a semiparametric maximum likelihood approach is proposed which requires no parametric specification of the marginal covariate distribution. By drawing from the von Mises calculus and V-Statistics theory, we obtain an asymptotic linear representation of the semiparametric maximum likelihood estimator (SMLE) of the regression parameters, which allows for a consistent estimator of asymptotic variance. An EM algorithm for computation is then developed to implement the proposed method using fractional imputation. Simulation results suggest that the SMLE method is robust, whereas the fully parametric method is subject to severe bias under model misspecification. A rangeland study from the National Resources Inventory (NRI) is used to illustrate the practical use of the proposed methodology.

Keywords and phrases: Asymptotic linearization representation; Fractional imputation; Nonparametric maximum likelihood estimator; Nonresponse.