

**ESTIMATING DISTRIBUTION FUNCTIONS FROM SURVEY DATA
USING NONPARAMETRIC REGRESSION**

by

Alicia A. Johnson and F. Jay Breidt, Colorado State University
Jean D. Opsomer, Iowa University

March 1, 2004

ABSTRACT

Auxiliary information is often used to improve the precision of estimators of the finite population cumulative distribution function through the use of superpopulation models. A variety of approaches are available to construct such estimators, including design-based, model-based and model-assisted methods. The superpopulation modeling framework can be either parametric or nonparametric, and the estimators can be constructed as either linear or nonlinear functions of the observations. In this article, we argue that model-assisted estimators based on a nonparametric model are a good overall choice for distribution function estimators, because they have good efficiency properties and are robust against model misspecification. When such estimators are constructed as linear functions of the data, they are also easily incorporated into the existing survey estimation paradigm through the use of survey weights. Theoretical properties of nonparametric distribution function estimators based on local linear regression are derived, and their practical behavior is evaluated in a simulation study.