Estimation of the distribution function using auxiliary information

by

Juan José Goyeneche

A Dissertation submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Major: Statistics

Major Professor: Dr. Wayne A. Fuller

Iowa State University

Ames, Iowa

1999

Graduate College
Iowa State University

This is to certify that the Doctoral Dissertation of

Juan José Goyeneche

has met the Dissertation requirements of Iowa State University

Committee Member

Committee Member

Committee Member

Committee Member

Committee Member

Major Professor

For the Major Program

For the Graduate College

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Estimation of the distribution function using auxiliary information

Juan José Goyeneche

Major Professor: Dr. Wayne A. Fuller
Iowa State University

The problem of estimating the finite population distribution function of a variable y is studied. The framework is one in which auxiliary information is available for each element in the population, and is similar to the framework used by Chambers and Dunstan (1986). In this study we introduce a new estimator, called the local-residuals estimator, of the finite population distribution function with auxiliary information. The local-residuals estimator is based on the distribution of the residuals from the regression of the variable of interest, y, on the vector of auxiliary variables, $\mathbf{x}$. One criticism of the estimator proposed by Chambers and Dunstan (1986) is that the performance of the estimator is poor when the superpopulation model is incorrectly specified. The local-residuals estimator is designed to be robust against model misspecification. The asymptotic properties of the local-residuals estimator are studied under different superpopulation models and the estimator is shown to be model consistent for the finite population distribution function. The conditions for asymptotic normality of the estimator are established and model consistent estimators of the variance of the local-residuals estimator are proposed. We also suggest an estimator of the superpopulation distribution function based on the local-residuals estimator. A Monte Carlo study compares the performance of the proposed estimator with alternative estimators presented in the literature.

1 INTRODUCTION

The problem of distribution function estimation appears when we are interested in knowing the population proportion of values of a variable that are less than or equal to a certain value, or set of values. Soil scientists may be interested in estimating the distribution of clay percent in the soil. Nutritionists may want to know the proportion of the population that consumes 30% or more of their calorie intake from saturated fat. Certain functions of the distribution function are also of interest, such as quantiles and functions of quantiles. A method of distribution function estimation similar to the one presented in this work is being applied to the estimation of Soil Components in the "Major Land Resource Area 107 Soil Survey Pilot Project" as described in Abbitt, Goyeneche and Schumi (1998).

In many situations auxiliary information is available. There are different types of auxiliary information. The values of auxiliary variables may be known for each element in the population, or for a large sample of the population. In other cases, only the population means of the auxiliary variables are known.

In this work, an estimator of the cumulative distribution function is presented that uses auxiliary information and local smoothing of the conditional distribution function, conditional on the auxiliary information. The notation and models used, and a review of the literature are presented in Chapter 2. The properties of the estimator are studied in Chapter 3. Monte Carlo results under different superpopulation models are presented in Chapter 4. Conclusions are presented in Chapter 5.

2 PREVIOUS WORK

2.1 Framework and Models

2.1.1 Notation

A *finite population* is a finite collection of elements or units. The number of elements in the population is denoted by N and is called the *population size*. Assume that the units of the finite population are identifiable and that a label is assigned to each unit. The set containing the N labels for the population elements is called the *sampling frame* and denoted by $\mathbb{U}$. Without loss of generality, assume that $\mathbb{U} = \{1, 2, \ldots, N\}$. When referring to the "unit of the population with label j", the shorter expression "unit j" is normally used.

Associated with each unit j in the population, there is a vector $\mathbf{y}_j$ of characteristics. Let $\mathcal{F}_y = \{\mathbf{y}_1, \ldots, \mathbf{y}_N\}$ be the entire set of N vectors. Sometimes, there is another vector $\mathbf{x}_j$ of auxiliary information associated with unit j. Let $\mathcal{F}_x = \{\mathbf{x}_1, \ldots, \mathbf{x}_N\}$ be the set containing the auxiliary information for the N units in the finite population.

A *sample* is a subset of units of the finite population. Let $\mathbb{A}$ denote the subset of labels from $\mathbb{U}$ that are in the sample. The values for the vectors $\mathbf{y}_j, j \in \mathbb{A}$, are observed. Often we refer to the set $\mathbb{A}$ as the sample, with the understanding that $\mathbb{A}$ is the set of labels of the units in the sample. The complement of the sample with respect to the finite population, the set of units that are not selected, is denoted by $\mathbb{A}^c = \mathbb{U} - \mathbb{A}$ (formally, $\mathbb{A}^c$ is the set of labels of the nonselected units). The number of elements in $\mathbb{A}$, denoted by n, is called the *sample size.* Let $\mathcal{A}$ be the set of all possible samples from $\mathbb{U}$. The *sample design* is a function $p(\cdot) : \mathcal{A} \longrightarrow [0,1]$ such that $p(a) = \mathrm{P}(\mathbb{A} = a)$ for any $a \in \mathcal{A}$, where $p(a)$ is the probability that the sample with labels in the set a is selected. The probability that unit j is selected in the sample is called the *first order inclusion probability*, or just *inclusion probability*, and is denoted by π_j, where

$$\pi_j = \mathrm{P}(j \in \mathbb{A}) = \sum_{a \in \mathcal{A}: j \in a} p(a),$$

and the sum is taken over all samples that contain unit j. Similarly, higher order inclusion probabilities can be defined. For instance, the *second order inclusion probability* is

$$\pi_{jk} = \mathrm{P}(j \in \mathbb{A}\ \cap\ k \in \mathbb{A}) = \sum_{a \in \mathcal{A}: j,k \in a} p(a),$$

where the sum includes all samples that contain both j and k.

The *indicator function* $I(\cdot)$ is widely used in sampling, and is defined as

$$I(l) = \begin{cases} 1 & \text{if } l \text{ is true} \\ 0 & \text{if } l \text{ is false} \end{cases}$$

where l is some logical expression. A particular use of the indicator function is in defining the *indicator variable* I_j as: $I_j = I(j \in \mathbb{A})$. That is, $I_j = 1$ when unit j has been selected and $I_j = 0$ otherwise.

A *finite population parameter* θ is some function of $\mathcal{F}_y$ and $\mathcal{F}_x$, $\theta = \theta(\mathcal{F}_y, \mathcal{F}_x)$. An *estimator* $\widehat{\theta}$ of θ is some function of the observed information. If $\mathcal{F}_x$ is known, that is, $\mathbf{x}_j$ is known for all $j \in \mathbb{U}$, then

$$\widehat{\theta} = \widehat{\theta}(\{\mathbf{y}_j, j \in \mathbb{A}\}, \mathcal{F}_x). \tag{2.1.1}$$

More details about different types of auxiliary information are given in Section 2.1.3.

Let us assume that $\mathbf{y}$ is scalar, or that we concentrate our attention on just one characteristic, y. The *finite population distribution function* for the variable y is

$$F_N(\dot{y}) = N^{-1} \sum_{i \in \mathbb{U}} I(y_i \leqslant \dot{y}) \tag{2.1.2}$$

for $\dot{y} \in \Re$. Unless otherwise noted, the terms *distribution function* and *finite population distribution function* will be used as synonyms. Assume $\mathcal{F}_x$ is known. An estimator of the distribution function expressed as a function of the observed y information and of $\mathcal{F}_x$ is

$$\widehat{F}_N(\dot{y}) = \widehat{F}_N(\dot{y}, \{y_j, j \in \mathbb{A}\}, \mathcal{F}_x). \tag{2.1.3}$$

2.1.2 Asymptotic considerations

Since the population under study is intrinsically finite (of size N), asymptotic calculations are based on a sequence of populations and samples with increasing sample size, n, and increasing population size, N (see Isaki and Fuller, 1982). A sequence of finite populations, which implies sequences of $\mathbb{U}_N$, $\mathcal{F}_{yN}$, $\mathcal{F}_{xN}$, $\mathbb{A}_N$, $\mathcal{A}_N$ and n_N are defined for $N = 1, 2, \ldots$. The asymptotic properties of an estimator $\widehat{\theta}_N$ are then defined in terms of this sequence.

We may treat $\mathcal{F}_{yN}$ as a set of fixed quantities, or consider it to be a particular realization of N random vectors $\mathbf{Y}_j$. The set of conditions that determines the joint distribution of $\mathbf{Y}_1, \ldots, \mathbf{Y}_N$ is called the *superpopulation model.* See Cassel et al. (1993, page 80). The set $\mathcal{F}_{xN}$ of auxiliary information is considered fixed in our discussion. It may be the case that $\mathcal{F}_{xN}$ is a part of the superpopulation model, but unless otherwise stated, the auxiliary information will be considered fixed.

Note that in the superpopulation context,

- when referring to a particular finite population, the population includes a set of units with their labels $\mathbb{U}_N$, the auxiliary information $\mathcal{F}_{xN}$, and a particular outcome for $\mathbf{Y}_1, \dots, \mathbf{Y}_N$. We will sometimes use $\boldsymbol{\mathcal{F}}_N$ to denote a particular finite population, where $\boldsymbol{\mathcal{F}}_N = \{\mathbb{U}_N, \mathcal{F}_{xN}, \mathcal{F}_{yN}\}$.
- treating $\mathcal{F}_{yN}$ as a set of fixed quantities is equivalent to considering a superpopulation model, but restricting our interest to a particular realization of the model, the realization that produced $\mathcal{F}_{yN}$. We may, then, use the superpopulation model approach in general, and condition on $\boldsymbol{\mathcal{F}}_N$ when interested in a particular finite population.

When we assume that the finite population is a realization from the superpopulation and assume the sample is selected according to a sampling design, the probability structure contains both the sampling probability and the superpopulation model.

The properties of an estimator can be considered with respect to a particular finite population. The estimator $\widehat{\theta}_N$ is *design unbiased* for the finite population parameter θ_N if

$$E(\widehat{\theta}_N \mid \boldsymbol{\mathcal{F}}_N) = \theta_N,$$

where the notation denotes conditioning on the particular finite population. Hence, the expectation is taken with respect to the sample design. For each sample $a \in \mathcal{A}_N$, $\widehat{\theta}_{Na}$ is based on the information from those units contained in the sample a, and the expectation is

$$E(\widehat{\theta}_N \mid \boldsymbol{\mathcal{F}}_N) = \sum_{a \in \mathcal{A}_N} \widehat{\theta}_{Na}\ p(a).$$

The estimator $\widehat{\theta}_N$ is *asymptotically design unbiased* for the finite population parameter θ_N if

$$\lim_{N\to\infty} E(\widehat{\theta}_N - \theta_N \mid \boldsymbol{\mathcal{F}}_N) = 0$$

The estimator $\widehat{\theta}_N$ is *design consistent* for the finite population parameter θ_N if

$$\lim_{N\to\infty} \mathrm{P}(|\widehat{\theta}_N - \theta_N| > \epsilon \mid \boldsymbol{\mathcal{F}}_N) = 0$$

for every $\epsilon > 0$.

The properties of an estimator can also be considered under the superpopulation model, for a particular sample $\mathbb{A}_N$. The estimator $\widehat{\theta}_N$ is *model unbiased* for the superpopulation parameter θ if

$$E(\widehat{\theta}_N \mid \mathbb{A}_N) = \theta$$

where the expectation is with respect to the model that generates $\mathbf{Y}_j$, and $\widehat{\theta}_N$ is a function of $\mathbf{Y}_j$, for $j \in \mathbb{A}_N$. The estimator $\widehat{\theta}_N$ is *asymptotically model unbiased* for the superpopulation parameter θ if

$$\lim_{N\to\infty} E(\widehat{\theta}_N - \theta \mid \mathbb{A}_N) = 0.$$

The estimator $\widehat{\theta}_N$ is *model consistent* for the superpopulation parameter θ if

$$\lim_{N\to\infty} \mathrm{P}(|\widehat{\theta}_N - \theta| > \epsilon \mid \mathbb{A}_N) = 0$$

for every $\epsilon > 0$.

2.1.3 Auxiliary information

Use of auxiliary information, either at the design stage or at the estimation stage, to improve the precision of estimates is very common in survey sampling. Auxiliary information can take different forms. The following is a nonexhaustive list of possible situations where the auxiliary information in the form of an auxiliary vector $\mathbf{x}$ is available:

- for each unit in the population the value of the auxiliary vector $\mathbf{x}$ is known.
- for multiphase samples, the value of $\mathbf{x}$ is observed for a large sample.
- summary information is known for $\mathbf{x}$ in the form of a histogram or frequency count for the finite population.
- only the population mean or total of $\mathbf{x}$ is known.

Usually, multiple sources of information are available for the auxiliary variables that form the vector $\mathbf{x}$. For instance, the values for a subgroup of the auxiliary variables are available for all units in the population, while the values for the rest of the auxiliary variables are only known for units in a sample larger than $\mathbb{A}$.

2.1.4 Superpopulation model

Assume that the finite population ($\mathcal{F}_{yN}$) is generated by a superpopulation model of the form

$$Y_k = \mathbf{x}_k'\boldsymbol{\beta} + h(\mathbf{x}_k)U_k, \; k = 1, 2, \dots , N, \tag{2.1.4}$$

where $\boldsymbol{\beta}$ is an unknown parameter, $h(\cdot) > 0$ is a known function that accounts for heteroescedasticity, and the U_k are independent identically distributed random variables with zero mean and distribution function $G(u)$. We will use the shortcut notation $h_k = h(x_k)$ sometimes. The set $\mathcal{F}_{xN}$ is assumed fixed. A realization of the superpopulation model random variables, denoted by $Y_1, Y_2, \ldots, Y_N$, corresponds to a particular finite population $\mathcal{F}_{yN} = \{y_1, y_2, \ldots, y_N\}$.

The *superpopulation distribution function* of Y is

$$\begin{aligned} F(\dot{y}) = \mathrm{P}(Y \leqslant \dot{y}) &= N^{-1} \sum_{i \in \mathbb{U}} \mathrm{P}(Y_i \leqslant \dot{y} \,\big|\, \mathbf{x} = \mathbf{x}_i) \\ &= N^{-1} \sum_{i \in \mathbb{U}} E\{I(Y_i \leqslant \dot{y}) \,\big|\, \mathbf{x} = \mathbf{x}_i\}. \end{aligned} \tag{2.1.5}$$

Note that

$$\mathrm{P}(Y \leqslant \dot{y}) = \mathrm{P}(Y \leqslant \dot{y} \,\big|\, \mathcal{F}_{xN}),$$

since the set $\mathcal{F}_{xN}$ is assumed fixed. The superpopulation distribution function (2.1.5) can be seen as the model expectation of the finite population distribution function defined in (2.1.2).

2.2 Distribution Function Estimation without Auxiliary Information

The Horvitz-Thompson estimator of $F_N(\dot{y})$ is

$$\widehat{F}_{HT}(\dot{y}) = \Big\{ \sum_{j \in \mathbb{A}} \pi_j^{-1} \Big\}^{-1} \Big\{ \sum_{j \in \mathbb{A}} \pi_j^{-1} I(y_j \leqslant \dot{y}) \Big\} \tag{2.2.1}$$

where π_j is the inclusion probability for unit j. Estimator (2.2.1) is the ratio of the Horvitz-Thompson estimator of the proportion of units in the finite population that have values of y less than or equal to $\dot{y}$, $\big\{ \sum_{j \in \mathbb{A}} \pi_j^{-1} I(y_j \leqslant \dot{y}) \big\}$, to the Horvitz-Thompson estimator of the population size, $\big\{ \sum_{j \in \mathbb{A}} \pi_j^{-1} \big\}$. In some designs the denominator part is equal to N for any sample in the sample space $\mathcal{A}$. Note that the Horvitz-Thompson estimator (2.2.1) of the distribution function does not use auxiliary concomitant variables at the estimation stage. Sometimes auxiliary information is used at the design stage of a sampling survey, and for such designs, the auxiliary information may be implicit in the inclusion probabilities π_j that appear in (2.2.1).

Francisco and Fuller (1991) considered the problem of distribution function and quantile estimation for complex designs. Restrictions on the sampling design are specified and limiting results for the estimator (2.2.1) are established for stratified cluster sampling. A method for constructing confidence intervals for superpopulation quantiles based on test inversion of the distribution function is presented. An expression for the joint limiting distribution of a vector of sample quantiles is given.

2.3 Distribution Function Estimation using Auxiliary Information

Chambers and Dunstan (1986) introduced a model based method to incorporate auxiliary information from a variable x when its value is known for all units in the population. Chambers and Dunstan assumed a superpopulation model of the form of model (2.1.4), with $h(x) = x^{1/2}$. The resulting model,

$$Y_k = x_k\beta + x_k^{1/2}U_k, \tag{2.3.1}$$

corresponds to the customary "ratio" model in survey sampling. The distribution of the random variables Y_k and U_k are specified in the superpopulation model for $k \in \mathbb{U}$. A realization of the U_k generates N particular values for the residuals that are denoted by $u_1, u_2, \ldots, u_N$. A set of N residuals $u_1, u_2, \ldots, u_N$ and the auxiliary information $x_1, x_2, \ldots, x_N$ generate a set of values of y, that is, a particular finite population $\mathcal{F}_{yN}$.

The distribution function $F_N(\dot{y})$ defined in (2.1.2) can be written as

$$F_N(\dot{y}) = N^{-1}\Big[\sum_{j\in\mathbb{A}} I(y_j \leqslant \dot{y}) \;+\; \sum_{i\in\mathbb{A}^c} I(y_i \leqslant \dot{y})\Big] \tag{2.3.2}$$

where the unknowns in formula (2.3.2) are in the last term of the sum. Letting $h_i = h(x_i)$, Chambers and Dunstan estimate the last term of (2.3.2) by observing that under model (2.3.1)

$$E\big[I(Y_i \leqslant \dot{y}) \mid x = x_i\big] = \mathrm{P}\big[Y_i \leqslant \dot{y} \mid x = x_i\big] = G\big(h_i^{-1}[\dot{y} - x_i\beta]\big) \tag{2.3.3}$$

where $G(\cdot)$ is the distribution function of U defined in (2.1.4). An estimator of the term $\sum_{i\in\mathbb{A}^c} I(y_i \leqslant \dot{y})$ can be derived by estimating $\sum_{i\in\mathbb{A}^c} G\big(h_i^{-1}[\dot{y} - x_i\beta]\big)$. Chambers and Dunstan presented the following estimator:

$$\widehat{F}_{CD}(\dot{y}) = N^{-1}\Big[\sum_{j\in\mathbb{A}} I(Y_j \leqslant \dot{y}) \;+\; \sum_{i\in\mathbb{A}^c} G_n\big(h_i^{-1}[\dot{y} - x_i b_n]\big)\Big] \tag{2.3.4}$$

where

$$b_n = \Big\{\sum_{j\in\mathbb{A}} h_j^{-2} Y_j x_j\Big\}\Big\{\sum_{j\in\mathbb{A}} h_j^{-2} x_j^2\Big\}^{-1} = \Big\{\sum_{j\in\mathbb{A}} Y_j\Big\}\Big\{\sum_{j\in\mathbb{A}} x_j\Big\}^{-1}$$

is an estimator of β, and G_n is a sample-based estimator of the distribution function of U in (2.3.1). The G_n is a function of the sample residuals, $\widehat{u}_j = Y_j - x_j b_n$, and is equal to

$$\begin{aligned} G_n\big(h_i^{-1}[\dot{y} - x_i b_n]\big) &= n^{-1}\sum_{j\in\mathbb{A}} I\big(h_j^{-1}[Y_j - x_j b_n] \leqslant h_i^{-1}[\dot{y} - x_i b_n]\big) \\ &= n^{-1}\sum_{j\in\mathbb{A}} I\big(x_i b_n + h_i h_j^{-1}[Y_j - x_j b_n] \leqslant \dot{y}\big). \end{aligned}$$

The Chambers and Dunstan Monte Carlo study, done with a population of 338 sugar cane farms that seems to follow model (2.3.1), shows that the estimator $\widehat{F}_{CD}$ can be considerably more efficient than the design based estimator (2.2.1) when the model is true.

Model based asymptotic results for $\widehat{F}_{CD}$ are based on Randles (1982). Chambers and Dunstan study the estimation error for $\widehat{F}_{CD}$ under model (2.3.1), where the error is

$$\widehat{F}_{CD}(\dot{y}) - F_N(\dot{y}) = N^{-1}\Big[\sum_{i\in\mathbb{A}^c} G_n\big(h_i^{-1}[\dot{y} - x_i b_n]\big) - \sum_{i\in\mathbb{A}^c} I(Y_i \leqslant \dot{y})\Big]. \tag{2.3.5}$$

Note that the first term on the right hand side of (2.3.5) depends on the Y random variables of the sample,

$$\sum_{i\in\mathbb{A}^c} G_n\big(h_i^{-1}[\dot{y} - x_i b_n]\big) = \sum_{i\in\mathbb{A}^c} n^{-1} \sum_{j\in\mathbb{A}} I\big(h_j^{-1}[Y_j - x_j b_n] \leqslant h_i^{-1}[\dot{y} - x_i b_n]\big),$$

while the second term on the right hand side of (2.3.5) depends on the Y of the nonsample units. If we condition on $\mathbb{A}$, the sample indices, the two terms in (2.3.5) are independent under the model. Let

$$F_r^*(\dot{y}, b_n) = (N-n)^{-1} \sum_{i\in\mathbb{A}^c} G_n\big(h_i^{-1}[\dot{y} - x_i b_n]\big) \tag{2.3.6}$$

and

$$F_r(\dot{y}) = (N-n)^{-1} \sum_{i\in\mathbb{A}^c} I(Y_i \leqslant \dot{y}), \tag{2.3.7}$$

where $F_r^*(\dot{y}, b_n)$ is the part of the estimation error that depends on the sample, and $F_r(\dot{y})$ is the proportion of nonselected units with Y_i less than or equal to $\dot{y}$. Both random quantities, $F_r^*(\dot{y}, b_n)$ and $F_r(\dot{y})$, are restricted to be between 0 and 1. Then, the estimation error $\widehat{F}_{CD}(\dot{y}) - F_N(\dot{y})$ can be seen as a difference between two random variables that are conditionally independent under the model multiplied by $N^{-1}(N-n)$,

$$\widehat{F}_{CD}(\dot{y}) - F_N(\dot{y}) = N^{-1}(N-n)\big[F_r^*(\dot{y}, b_n) - F_r(\dot{y})\big].$$

The conditional variance of the estimation error (2.3.5) is then

$$\begin{aligned} V\big(\widehat{F}_{CD}(\dot{y}) - F_N(\dot{y}) \mid \mathbb{A}\big) &= \Big\{N^{-1}(N-n)\Big\}^2 \Big\{V\big(F_r^*(\dot{y}, b_n) \mid \mathbb{A}\big) + V\big(F_r(\dot{y}) \mid \mathbb{A}\big)\Big\} \\ &= \Big\{1 - nN^{-1}\Big\}^2 \Big\{V\big(F_r^*(\dot{y}, b_n) \mid \mathbb{A}\big) + V\big(F_r(\dot{y}) \mid \mathbb{A}\big)\Big\}, \end{aligned} \tag{2.3.8}$$

because $F_r^*(\dot{y}, b_n)$ and $F_r(\dot{y})$ are conditionally independent given $\mathbb{A}$. Chambers and Dunstan denote the first variance on the right hand side of (2.3.8) as $\mathbf{W}_r^*(\dot{y}, \beta)$ and the second variance on the right hand side of (2.3.8) as $\mathbf{W}_r(\dot{y}, \beta)$, that is,

$$\mathbf{W}_r^*(\dot{y}, \beta) = V\big(F_r^*(\dot{y}, b_n) \bigm| \mathbb{A}\big),$$

and

$$\mathbf{W}_r(\dot{y}, \beta) = V\big(F_r(\dot{y}) \bigm| \mathbb{A}\big).$$

Based on Theorem 2.13 in Randles (1982), Chambers and Dunstan write the first variance on the right hand side of (2.3.8) as

$$V\big(F_r^*(\dot{y}, b_n) \bigm| \mathbb{A}\big) = \mathbf{W}_r^*(\dot{y}, \beta) = \mathbf{D}_r(\dot{y}, \beta)\ \mathbf{V}^*(\dot{y}, \beta)\ \mathbf{D}_r'(\dot{y}, \beta),$$

where $\mathbf{D}_r(\dot{y}, \beta)$ is the row vector

$$\mathbf{D}_r(\dot{y}, \beta) = \Big(1, n^{-1}(N-n)^{-1} \sum_{j \in \mathbb{A}} \sum_{i \in \mathbb{A}^c} \big\{[h_j^{-1} x_j - h_i^{-1} x_i] g\big(h_i^{-1}[\dot{y} - x_i \beta]\big)\big\}\Big)$$

and $\mathbf{V}^*(\dot{y}, \beta)$ is the conditional variance matrix of the vector $\big(F_r^*(\dot{y}, \beta),\ b_n\big)'$,

$$\mathbf{V}^*(\dot{y}, \beta) = V\big[\big(F_r^*(\dot{y}, \beta),\ b_n\big)' \bigm| \mathbb{A}\big] = \big(V_{ij}^*\big).$$

The elements of $\mathbf{V}^*(\dot{y}, \beta)$ are

$$\begin{aligned}
V_{11}^* &= n^{-1}(N-n)^{-2}\Big\{ \sum_{i \in \mathbb{A}^c} \sum_{k \in \mathbb{A}^c} G\big\{\min\big(h_i^{-1}[\dot{y} - x_i\beta], h_k^{-1}[\dot{y} - x_k\beta]\big)\big\} \\
&\qquad\qquad - \Big[\sum_{i \in \mathbb{A}^c} G\big(h_i^{-1}[\dot{y} - x_i\beta]\big)\Big]^2 \Big\} \\
V_{12}^* &= V_{21}^* = \Big[\sum_{j \in \mathbb{A}} h_j^{-2} x_j^2\Big]^{-1} n^{-1}(N-n)^{-1} \sum_{j \in \mathbb{A}} \sum_{i \in \mathbb{A}^c} h_j^{-1} x_j \times \\
&\qquad \times E\Big[h_j^{-1}[Y_j - x_j\beta]\ I\big(x_i\beta + h_i h_j^{-1}[Y_j - x_j\beta] \leqslant \dot{y}\big) \bigm| \mathbb{A}\Big] \\
V_{22}^* &= \Big[\sum_{j \in \mathbb{A}} h_j^{-2} x_j^2\Big]^{-1} V(U).
\end{aligned}$$

The second variance on the right hand side of (2.3.8) is

$$\begin{aligned} V\big(F_r(\dot{y}) \mid \mathbb{A}\big) &= \mathbf{W}_r(\dot{y}, \beta) \\ &= V\Big([N-n]^{-1} \sum_{i \in \mathbb{A}^c} I[Y_i \leqslant \dot{y}] \mid \mathbb{A}\Big) \\ &= (N-n)^{-2} \sum_{i \in \mathbb{A}^c} V\big(I[Y_i \leqslant \dot{y}] \mid \mathbb{A}\big) \\ &= (N-n)^{-2} \sum_{i \in \mathbb{A}^c} G\big(h_i^{-1}[\dot{y} - x_i\beta]\big)\big\{1 - G\big(h_i^{-1}[\dot{y} - x_i\beta]\big)\big\}, \end{aligned}$$

since the Y_i are conditionally independent given $\mathbb{A}$, $V\big(I[Y_i \leqslant \dot{y}] \mid \mathbb{A}\big) = V\big(I[Y_i \leqslant \dot{y}]\big) = \mathrm{P}(Y_i \leqslant \dot{y})[1 - \mathrm{P}(Y_i \leqslant \dot{y})]$ and $\mathrm{P}(Y_i \leqslant \dot{y}) = G\big(h_i^{-1}[\dot{y} - x_i\beta]\big)$.

Chambers and Dunstan use the following set of assumptions to find the limiting distribution of the estimation error (2.3.5). The notation of Section 2.1.2 is used.

(CD.1) $\lim_{N\to\infty} n_N N^{-1} = c$, where $c \in (0,1)$,

(CD.2) $G(u)$, the distribution function of U in model (2.3.1), is differentiable, with derivative $g(u)$,

(CD.3) there exist $M_1, M_2 < \infty$, such that $|x_k| < M_1$ and $h_k < M_2$ for all N and all $k \in \mathbb{U}_N$,

(CD.4) for arbitrary b, $0 < \lim_{N\to\infty} F_r^*(\dot{y}, b) < 1$, where $F_r^*(\dot{y}, b)$ is defined in (2.3.6),

(CD.5) b_n is asymptotically normal under (2.3.1), that is, $[V(b_n)]^{-1/2}[b_n - \beta]$ converges in distribution to a standard normal distribution as $N \to \infty$.

If conditions (CD.1) through (CD.5) hold, the Chambers and Dunstan result is that

$$\Big\{(1 - nN^{-1})^2\Big[\mathbf{W}_r^*(\dot{y}, \beta) + \mathbf{W}_r(\dot{y}, \beta)\Big]\Big\}^{-1/2}\Big\{\widehat{F}_{CD}(\dot{y}) - F_N(\dot{y}) \mid \mathbb{A}\Big\} \longrightarrow \mathrm{N}(0,1) \tag{2.3.9}$$

in distribution. Note that there are no direct assumptions on the sample design, only on the properties of the sample.

Chambers and Dunstan consider the possibility of misspecification of the model when the mean part of (2.3.1) holds but the variance function is $h_a(x) \neq h(x)$. In general, $\widehat{F}_{CD}(\dot{y}) - F_N(\dot{y})$ is still asymptotically normal but with nonzero mean. The asymptotic bias is close to zero if the sample is such that $[h(x_j)h_a(x_i)][h(x_i)h_a(x_j)]^{-1}$ is approximately one for all $j \in \mathbb{A}^c$.

Dunstan and Chambers (1989) extend the model based approach to the case when only histogram summary information is known for the auxiliary variable. The distribution function estimator and an estimator of its variance are adapted to the case of "limited information". Dunstan and Chambers' Monte Carlo results suggest that the "limited information" estimator is almost as efficient as the corresponding "full information" one, and that the confidence intervals generated by either of these model based methods have better coverage properties than confidence intervals for the design based estimator (2.2.1).

Rao, Kovar and Mantel (1990) suggested design based ratio and difference estimators of the distribution function that also make use of the auxiliary information at the estimation stage. Rao, Kovar and Mantel emphasize that the Chambers and Dunstan estimator is not design unbiased and that it is not robust against model misspecification. The variable $I(x_i\widetilde{\beta} \leqslant \dot{y})$ is used as an auxiliary variable for $I(y_i \leqslant \dot{y})$. The customary ratio and difference estimators are then defined as:

$$\widehat{F}_{RKMr}(\dot{y}) = N^{-1}\Big\{\sum_{j\in\mathbb{A}} \pi_j^{-1} I(y_j \leqslant \dot{y})\Big\}\Big\{\sum_{j\in\mathbb{A}} \pi_j^{-1} I(x_j\widetilde{\beta} \leqslant \dot{y})\Big\}^{-1}\Big\{\sum_{i\in\mathbb{U}} I(x_i\widetilde{\beta} \leqslant \dot{y})\Big\} \tag{2.3.10}$$

and

$$\widehat{F}_{RKMd}(\dot{y}) = N^{-1}\Big\{\sum_{j\in\mathbb{A}}\pi_j^{-1}I(y_j \leqslant \dot{y}) \;+\; \Big[\sum_{i\in\mathbb{U}}I(x_i\widetilde{\beta} \leqslant \dot{y}) - \sum_{j\in\mathbb{A}}\pi_j^{-1}I(x_j\widetilde{\beta} \leqslant \dot{y})\Big]\Big\} \tag{2.3.11}$$

where $\widetilde{\beta} = [\sum_{j\in\mathbb{A}}\pi_j^{-1}h_j^{-2}y_jx_j][\sum_{j\in\mathbb{A}}\pi_j^{-1}h_j^{-2}x_j^2]^{-1}$. When the variance function $h(x)$ is specified to be $h(x) = x^{1/2}$, $\widetilde{\beta} = [\sum_{j\in\mathbb{A}}\pi_j^{-1}x_j]^{-1}[\sum_{j\in\mathbb{A}}\pi_j^{-1}y_j]$. Estimator (2.3.11) is design unbiased, and estimator (2.3.10) is approximately design unbiased. The total variation of the N quantities $x_i\beta$ is in general smaller than the total variation of y_i. Estimators (2.3.10) and (2.3.11) are not model unbiased. If model (2.3.1) holds $E[Y_i] = x_i\beta$, but $E[I(Y_i \leqslant \dot{y})] = \mathrm{P}(Y_i \leqslant \dot{y}) \neq I(E[Y_i] \leqslant \dot{y}) = I(x_i\beta \leqslant \dot{y})$. Therefore $E[\widehat{F}_{RKMr}(\dot{y}) - F_N(\dot{y}) \mid \mathbb{A}] \neq 0$ and $E[\widehat{F}_{RKMd}(\dot{y}) - F_N(\dot{y}) \mid \mathbb{A}] \neq 0$. Also,

$$\lim_{n,N\to\infty} N^{-1}\sum_{i\in\mathbb{U}_N} I(x_i\beta \leqslant \dot{y}) \neq \lim_{n,N\to\infty} F_N(\dot{y}),$$

unless model (2.3.1) fits exactly with $V(U) = 0$.

An asymptotically design unbiased, model unbiased estimator is based on the distribution of U. Assume that for $i \in \mathbb{U}$ we know the quantities

$$G_i = N^{-1}\sum_{k\in\mathbb{U}} I\big(h_k^{-1}[y_k - x_kb_n] \leqslant h_i^{-1}[\dot{y} - x_ib_n]\big),$$

that is, the finite population distribution function of u evaluated at $h_i^{-1}[\dot{y} - x_ib_n]$. A design unbiased, asymptotically model unbiased estimator of $F_N(\dot{y})$ is then

$$F^*_{RKMdm}(\dot{y}) = N^{-1}\Big(\sum_{j\in\mathbb{A}}\pi_j^{-1}I(y_j \leqslant \dot{y}) \;+\; \{\sum_{i\in\mathbb{U}}G_i - \sum_{j\in\mathbb{A}}\pi_j^{-1}G_i\}\Big). \tag{2.3.12}$$

Estimator (2.3.12) is a difference estimator that uses the auxiliary variable G_i, hence, estimator (2.3.12) is design unbiased. If model (2.3.1) holds,

$$E\big(\sum_{j\in\mathbb{A}}\pi_j^{-1}I(y_j \leqslant \dot{y}) \mid \mathbb{A}_N\big) = E\big(\sum_{j\in\mathbb{A}}\pi_j^{-1}G_i \mid \mathbb{A}_N\big),$$

thus, estimator (2.3.12) is model unbiased. The G_i can not be computed, since β is unknown, and we only know y_j for $j \in \mathbb{A}$. A feasible estimator of F_N is constructed by using

$$\widehat{G}_i = \Big\{\sum_{j\in\mathbb{A}} \pi_j^{-1}\Big\}^{-1} \Big\{\sum_{j\in\mathbb{A}} \pi_j^{-1} I\big(h_i h_j^{-1}(y_j - x_j\widetilde{\beta}) + x_i\widetilde{\beta} \leqslant \dot{y}\big)\Big\}$$

as an design consistent estimator of G_i and

$$\widehat{G}_{ic} = \Big\{\sum_{j\in\mathbb{A}} (\pi_{ij}/\pi_i)^{-1}\Big\}^{-1} \Big\{\sum_{j\in\mathbb{A}} (\pi_{ij}/\pi_i)^{-1} I\big(h_i h_j^{-1}(y_j - x_j\widetilde{\beta}) + x_i\widetilde{\beta} \leqslant \dot{y}\big)\Big\}$$

as a design consistent estimator of G_i, conditional on $i \in \mathbb{A}$, where π_{ij}/π_i is the conditional probability of selecting units i and j given that unit i has been selected. The estimator,

$$\widehat{F}_{RKMdm}(\dot{y}) = N^{-1}\Big\{\sum_{j\in\mathbb{A}} \pi_j^{-1} I(y_j \leqslant \dot{y}) \;+\; \Big(\sum_{i\in\mathbb{U}} \widehat{G}_i - \sum_{j\in\mathbb{A}} \pi_j^{-1}\widehat{G}_{ic}\Big)\Big\} \tag{2.3.13}$$

is design consistent and asymptotically model unbiased.

Rao, Kovar and Mantel used two populations in a Monte Carlo study: the "Chambers and Dunstan" population, the population with 338 sugar cane farms that seems to follow (2.3.1), and the "Hansen, Madow and Tepping" population generated by the model

$$Y_k = 0.40 + 0.25x_k + x_k^{3/4} U_k \tag{2.3.14}$$

where the U_k are independent and identically distributed with zero mean. This model "was designed to make it not distinguishable from model (2.3.1)" (see Hansen et al., 1983). When evaluated at the 25th, 50th and 75th population quantiles ($\dot{y}_\alpha : F_N(\dot{y}_\alpha) = \alpha,\ \alpha = 0.25, 0.50, 0.75$), estimator (2.3.13) performs better than estimators (2.3.10) and (2.3.11) in both populations. The Chambers and Dunstan estimator (2.3.4) shows larger bias than estimator (2.3.13) (even in the first population). The mean square error for the Chambers and Dunstan estimator is smaller for the first population, but estimator (2.3.13) has smaller mean square error for $\alpha = 0.25$ and $\alpha = 0.75$ in the second population. The designs used in the Monte Carlo study were simple random sampling and stratified with proportional allocation for the "Chambers and Dunstan" population and a stratified design with ten strata and equal sample size in each stratum for the "Hansen, Madow and Tepping" population. The strata for the "Hansen, Madow and Tepping" population were created such that sum of x is approximately equal for each stratum.

The three estimators presented by Rao, Kovar and Mantel are functions of Horvitz-Thompson estimators of totals. Although the estimators involve the estimated parameter $\widetilde{\beta}$, standard Horvitz-Thompson estimators of the variance can be applied. The variance estimator for estimator (2.3.13), $\widehat{F}_{RKMdm}(\dot{y})$, requires computation of third order inclusion probabilities π_{ijk}, which may be cumbersome to compute for some designs.

For quantile estimation, Rao, Kovar and Mantel use ratio and difference estimators that make use of the sample and population quantiles for the x variable in order to improve precision over the sample quantiles for y. Note that $\widehat{F}_{RKMdm}(\dot{y})$ and $\widehat{F}_{RKMd}(\dot{y})$ may not be monotone increasing. Rao, Kovar and Mantel do not present quantile estimators that rely on inversion of the estimated distribution function.

Dorfman (1993) discusses estimators (2.3.4) and (2.3.13), and proposes a modified version of (2.3.13) that is less dependent on design based ingredients and does not need computation of second order probabilities. Dorfman observes that the estimator (2.3.4) is preferable when "reasonably careful modeling" analysis has been conducted.

Rao and Liu (1992) distinguish four different approaches for the use of auxiliary information at the estimation stage. With respect to distribution function estimation, the four approaches are:

- design based approach, which leads to estimators like (2.2.1), that do not use auxiliary information, and estimators (2.3.10) and (2.3.11), that use auxiliary information. Rao and Liu note that in general "the correlation between $I(y_j \leqslant \dot{y})$ and $I(x_j\beta \leqslant \dot{y})$ appears to be weaker than the correlation between y_j and x_j."
- prediction approach, which includes the model based estimator (2.3.4).
- model assisted approach, which leads to the estimator (2.3.13).
- conditional approach, where the estimator is constructed relying only on the knowledge of μ_x, and not on all values of $x_j,\ j \in \mathbb{U}$.

An interesting comparison of the properties of the estimators (2.3.4) and (2.3.13) is given in Chambers, Dorfman and Hall (1992). The large-sample mean square errors of both estimators are considered from a theoretical point of view. None of the estimators dominates the others for all $\dot{y} \in \Re$. Chambers, Dorfman and Hall consider the model

$$Y_k = a + bx_k + U_k,$$

where the *design points* x_k are the realization of a random variable with expected value μ_x, variance τ^2, and *design density* $d(x)$. The design density $d(x)$ used by Chambers, Dorfman and Hall (1992) is the limiting probability density function of the x in the superpopulation model. The parameters a and b are unknown, and the U_k are independent identically distributed with mean zero and density $g(u)$. Chambers, Dorfman and Hall (1992) show that the difference

$$V\big\{\widehat{F}_{RKMdm}(\dot{y}) - F_N(\dot{y})\big\} - V\big\{\widehat{F}_{CD}(\dot{y}) - F_N(\dot{y})\big\} \tag{2.3.15}$$

is positive when $g(u)$ is symmetric about zero and $d(x)$ is symmetric about μ_x, and the model used to construct $\widehat{F}_{RKMdm}$ and $\widehat{F}_{CD}$ is true. The difference (2.3.15) can be negative under different specifications of $g(u)$ and $d(x)$. For instance, under the following conditions,

- $g(u)$ bounded on a compact support, with the exception of a pole at u_0, where $g(u) \propto |u-u_0|^{-3/4}$ as $u \to u_0$,
- the design points have a density bounded on a compact support, with the exception of a pole at x_0, where $d(x) \propto |x-x_0|^{-3/4}$ as $x \to x_0$,
- $\mu_x = 0$,
- $\dot{y}_0 = a + bx_0 + u_0$,

the difference (2.3.15) becomes negative for $\dot{y} = \dot{y}_0$. Even though the previously described situation is very extreme, it shows that from a theoretical viewpoint there are situations where $\widehat{F}_{RKMdm}$ outperforms $\widehat{F}_{CD}$ even when the model used in the construction of both estimators is correctly specified.

Wang and Dorfman (1996) combine estimators $\widehat{F}_{CD}$ and $\widehat{F}_{RKMdm}$ in a weighted average:

$$\widehat{F}_{WD}(\dot{y}) = w_{\dot{y}}\widehat{F}_{CD}(\dot{y}) \ + \ (1-w_{\dot{y}})\widehat{F}_{RKMdm}(\dot{y}) \tag{2.3.16}$$

The weight $w_{\dot{y}}$ is estimated from the sample in order to minimize the asymptotic mean square error. Wang and Dorfman show that estimators (2.3.4) and (2.3.13) have some components that are negatively correlated. Wang and Dorfman take advantage of this negative correlation in the method for selecting $w_{\dot{y}}$. An estimator of the variance of $\widehat{F}_{WD}$ is given. This variance estimator uses quantities that are calculated in the computation of the optimum $w_{\dot{y}}$. An interesting byproduct of the computation of estimator (2.3.16) is the value that $w_{\dot{y}}$ assumes for each value of $\dot{y}$. These values of $w_{\dot{y}}$ give "an idea of the relative position of $\widehat{F}_{WD}$ between $\widehat{F}_{CD}$ and $\widehat{F}_{RKMdm}$". Small values of $w_{\dot{y}}$ indicate that $\widehat{F}_{RKMdm}$ is preferred over $\widehat{F}_{WD}$ in the construction of $\widehat{F}_{WD}$. Values of $w_{\dot{y}}$ close to one, on the other hand, indicate that $\widehat{F}_{WD}$ is preferred over $\widehat{F}_{RKMdm}$ in the construction of estimator (2.3.16).

Wang and Dorfman claim that the new estimator $\widehat{F}_{WD}$ is preferable to both the Chambers and Dunstan and the Rao, Kovar and Mantel estimators, in the sense that "losses of efficiency in the worst cases are marginal and gains can be appreciable."

2.4 Nonparametric Estimation

This section refers to the group of techniques for nonparametric estimation of functions known as *kernel smoothing.* Kernel smoothing permits one to explore relationships in data sets without imposing a full parametric model.

Let $\gamma(x_k) = E(Y_k \bigm| X_k = x_k)$ be the conditional expectation of Y given x_k. The conditional expectation $\gamma(x)$ is usually called the *regression* of Y on x. The regression function minimizes the mean square error $E[Y_k - \ell(x_k)]^2$ over all functions $\ell(x)$. The approach used in *nonparametric regression* is to approximate $\gamma(x)$ by a function $m(x)$ that is not restricted to belong to a fixed finite parameter family.

The method of *local polynomial kernel estimators* estimates the regression function at a point x_0 by fitting a pth degree polynomial to the data using weighted least squares. The weights are computed using the *kernel function.* The kernel function is usually selected to be a density symmetric about x_0 with a scaling parameter b called the *bandwidth.* The weights used in the least squares fitting are

$$w(x_k) = b^{-1}K\bigl(b^{-1}[x_0 - x_k]\bigr) \tag{2.4.1}$$

where K is the kernel function. Let K_b denote the rescaled kernel function

$$K_b(x_0 - x) = b^{-1}K\bigl(b^{-1}[x_0 - x]\bigr).$$

Normally K is selected to be a symmetric unimodal density that assigns larger weights to points close to x_0. Points closer to x_0 then have more influence on the estimation of

$m(x_0)$ than points that are farther from x_0. The relative distance between x_0 and other points is controlled by the bandwidth. For small b, $\widehat{m}(x_0)$ depends more heavily on the points closest to x_0 and the regression curve is a more wiggly, *undersmoothed* estimate. As $b \to 0$, $\widehat{m}(x)$ tends to an interpolation between the points $(y_j, x_j), j \in \mathbb{A}$. When b is large an *oversmoothed estimate* is produced, the weights tend to be approximately equal and the estimate tends towards the ordinary least squares fit (Wand and Jones, 1995, page 117). The terms "small" and "large" b are relative to some optimum b. If we select a value of b smaller than the optimum b we have a undersmoothed estimate. If we select a value of b larger than the optimum b we have a oversmoothed estimate.

The local polynomial kernel smoothing method is carried out as follows. Consider a pth degree polynomial

$$\beta_0 + \beta_1(x_k - x_0) + \ldots + \beta_p(x_k - x_0)^p$$

on the values of x centered at x_0. Let $\widehat{\boldsymbol{\beta}} = (\widehat{\beta}_0, \widehat{\beta}_1, \ldots, \widehat{\beta}_p)'$ be the weighted least squares estimator of $\boldsymbol{\beta}$, using the weights $w(x_k)$ defined in (2.4.1). Since the independent variable is centered at x_0,

$$\widehat{m}(x_0) = \widehat{\beta}_0.$$

Recall that $\widehat{m}(x_0)$ is also a function of b and K. A simple formula for $\widehat{m}(x_0)$ is available when $p = 0$

$$\widehat{m}(x_0) = \Big[\sum_{j \in \mathbb{A}} K_b(x_0 - x_j)\Big]^{-1}\Big[\sum_{j \in \mathbb{A}} K_b(x_0 - x_j)\; Y_j\Big] \tag{2.4.2}$$

Estimator (2.4.2) is known as the Nadaraya-Watson estimator.

In the finite population setting the sample weights are used by some authors to obtain consistent estimators of the finite population regression fit. The weights used in the weighted least squares estimator of $\boldsymbol{\beta}$ are then $\pi_k^{-1} w(x_k)$, where π_k is the first order probability for unit k and $w(x_k)$ is defined in (2.4.1). The formula for the Nadaraya-Watson estimator (2.4.2) is then

$$\widehat{m}(x_0) = \Big[\sum_{j\in\mathbb{A}} \pi_j^{-1} K_b(x_0 - x_j)\Big]^{-1} \Big[\sum_{j\in\mathbb{A}} \pi_j^{-1} K_b(x_0 - x_j)\ Y_j\Big] \qquad (2.4.3)$$

if sampling weights are used.

A measure of performance of the kernel smoothing estimator is the mean square error

$$MSE\big(\widehat{m}(x_0)\big) = E[\widehat{m}(x_0) - m(x_0)]^2 \qquad (2.4.4)$$

for the point x_0. The mean square error (2.4.4) is conditional on the x values, that is, considering the x fixed. A consolidated error measure for the whole range of x is the *weighted mean integrated square error*

$$MISE(\widehat{m}) = E\Big\{\int [\widehat{m}(x) - m(x)]^2 d(x)\ \mathrm{d}x\Big\}, \qquad (2.4.5)$$

where $d(x)$ is the design density defined in Section 2.3. Assume that "b" corresponds to the value of the nth term of a sequence of bandwidths. Asymptotic considerations usually assume that

$$\lim_{n\to\infty} b = 0$$

$$\lim_{n\to\infty} nb = \infty.$$

The mean square error can be decomposed into two terms

$$MSE\big(\widehat{m}(x_0)\big) = V\big(\widehat{m}(x_0)\big) + \big(E[\widehat{m}(x_0)] - m(x_0)\big)^2 \qquad (2.4.6)$$

The bias part in (2.4.6), $E[\widehat{m}(x_0)] - m(x_0)$, increases with b, so we need b small to have small bias contribution to the mean square error. On the other hand, the variance part in equation (2.4.6) is $O\big((nb)^{-1}\big)$ therefore reducing b will increase the variance contribution to $MSE\big(\widehat{m}(x_0)\big)$. Section 5.3 of Wand and Jones (1995) considers this *variance-bias trade-off* problem in determining the value of b for the linear case, $p = 1$.

Choosing a particular kernel function K and an appropriate bandwidth b are some of the central issues in kernel smoothing. The selection of the bandwidth has a much larger effect on the performance of the estimator than the selection of K. The methods of selection of the bandwidth based on the data are called *bandwidth selectors*. One bandwidth selector often used chooses b to minimize the $MISE(\widehat{m})$,

$$b_{MISE} = \min_{b} \big(MISE(\widehat{m}) \big).$$

The weighted mean integrated square error is unknown, for it depends on $m(x)$. An estimator of b_{MISE} is used in practice.

Another point of concern is the performance of $\widehat{m}(\cdot)$ near the boundaries of the x values, that is, close to $\min(x_j, j \in \mathbb{A})$ and $\max(x_j, j \in \mathbb{A})$, where the kernel window may be partially empty of data. The problem of estimating $\widehat{m}(\cdot)$ near the boundaries is known as the *boundary bias* problem. The boundary bias problem deals with the difference in the orders of magnitude of the bias in an interior point and near the boundary of the x data.

2.5 Nonparametric Estimation of the Distribution Function

Kuo (1988) and Kuk (1993) use nonparametric estimators of the joint distribution function of x and y, as instruments for estimating the distribution function of y. Kuo proposed an estimator of the joint distribution function

$$\widehat{F}_{Kuo,xyN}(\dot{x}, \dot{y}) = N^{-1} \Big\{ \sum_{j \in \mathbb{A}} I(x_j \leqslant \dot{x}, y_j \leqslant \dot{y}) + \sum_{i \in \mathbb{A}^c} \sum_{j \in \mathbb{A}} W_{ij} I(x_i \leqslant \dot{x}, y_j \leqslant \dot{y}) \Big\} \tag{2.5.1}$$

where the weights W_{ij} can be computed using one of the following methods:

1. naive estimator: $W_{ij} = \big[I(|x_i - x_j| < \epsilon) \big] \big[\sum_{j \in \mathbb{A}} I(|x_i - x_j| < \epsilon) \big]^{-1}$.

2. kernel method: $W_{ij} = \{K[b^{-1}(x_i - x_j)]\}\{\sum_{j\in\mathbb{A}} K[b^{-1}(x_i - x_j)]\}^{-1}$, for some function K such that $\int K(x)\mathrm{d}x = 1$.

3. k nearest neighbors method: $W_{ij} = k^{-1}$ if x_j is one of the k nearest neighbors of x_i, $W_{ij} = 0$ otherwise.

Kuo does not discuss methods to select optimal ϵ or k, and refers the reader to Silverman (1985) for the selection of b. Estimator (2.5.1) uses the y values in the sample to impute for the unobserved y. Note that $\sum_{j\in\mathbb{A}} W_{ij} = 1$ for all $i \in \mathbb{A}^c$. For each $i \in \mathbb{A}^c$, $\sum_{j\in\mathbb{A}} W_{ij} I(x_i \leqslant \dot{x}, y_j \leqslant \dot{y})$ is a quantity between 0 and 1 that tries to predict the unknown $I(x_i \leqslant \dot{x}, y_i \leqslant \dot{y})$. Special care in selecting ϵ and b is required to avoid undefined weights in methods 1 and 2. The estimator of the finite population distribution function of y is the corresponding marginal of (2.5.1):

$$\widehat{F}_{Kuo}(\dot{y}) = N^{-1}\Big\{\sum_{j\in\mathbb{A}} I(y_j \leqslant \dot{y}) + \sum_{i\in\mathbb{A}^c}\sum_{j\in\mathbb{A}} W_{ij} I(y_j \leqslant \dot{y})\Big\} \qquad (2.5.2)$$

A Monte Carlo study comparing estimators (2.5.1), (2.5.2) and (2.2.1) is presented for three populations and three sampling designs. Estimator (2.5.2) does not perform much better that the Horvitz-Thompson estimator (2.2.1) in the Monte Carlo study.

Kuk estimates the joint distribution function using kernel smoothing to obtain

$$\widehat{F}_{Kuk,xyN}(\dot{x}, \dot{y}) = \Big\{\sum_{j\in\mathbb{A}} \pi_j^{-1}\Big\}^{-1}\Big\{\sum_{j\in\mathbb{A}} \pi_j^{-1} W[b^{-1}(\dot{x} - x_j)] W[b^{-1}(\dot{y} - y_j)]\Big\}^{-1} \qquad (2.5.3)$$

where $W(u) = e^u(1 + e^u)^{-1}$ is the standard logistic distribution function. The estimator of the bivariate density corresponding to (2.5.3) is

$$\widehat{f}_{Kuk,xy}(\dot{x}, \dot{y}) = b^{-2}\Big\{\sum_{j\in\mathbb{A}} \pi_j^{-1}\Big\}^{-1}\Big\{\sum_{j\in\mathbb{A}} \pi_j^{-1} w[b^{-1}(\dot{x} - x_j)] w[b^{-1}(\dot{y} - y_j)]\Big\}^{-1}$$

where $w(u) = e^u(1+e^u)^{-2}$. A smoothed estimator of the conditional distribution of y given x is

$$\widehat{F}_{Kuk,y|x}(\dot{y} \mid \dot{x}) = \Big\{ \sum_{j\in\mathbb{A}} \pi_j^{-1} w[b^{-1}(\dot{x}-x_j)] \Big\}^{-1} \Big\{ \sum_{j\in\mathbb{A}} \pi_j^{-1} w[b^{-1}(\dot{x}-x_j)] W[b^{-1}(\dot{y}-y_j)] \Big\}^{-1}$$

Since the distribution function of x, denoted by $F_{xN}(\cdot)$, is assumed to be known, the estimator of the distribution function of y can be computed as

$$\widehat{F}_{Kuk}(\dot{y}) = \int \widehat{F}_{Kuk,y|x}(\dot{y} \mid \dot{x}) \mathrm{d}F_{xN}(\dot{x}) = \sum_{i\in\mathbb{U}} \widehat{F}_{Kuk,y|x}(\dot{y}|x_i) \qquad (2.5.4)$$

Note that only one bandwidth parameter b is used for both x and y. Kuk recommends pre-scaling of x and y to similar ranges. The value for the bandwidth parameter is selected as $b = n^{-1}R_x$, where R_x denotes the range of x: $R_x = \max\{x_i, i \in \mathbb{U}\} - \min\{x_i, i \in \mathbb{U}\}$. An expression for the variance of (2.5.4) and a discussion on design consistency of $\widehat{F}_{Kuk}(\dot{y})$ are provided. A Monte Carlo study (based on 200 samples) shows that estimator (2.5.4) is robust against misspecification of the model. Kuk states that based on empirical evidence, estimator (2.5.4) "is more efficient than the estimators suggested by Rao et al. (1990)."

Chambers, Dorfman and Wehrly (1993) use nonparametric regression to estimate $F_r(\dot{y})$, the distribution function of the unobserved units

$$F_r(\dot{y}) = (N-n)^{-1} \sum_{i\in\mathbb{A}^c} I(y_i \leqslant \dot{y}). \qquad (2.5.5)$$

Chambers, Dorfman and Wehrly assume a working model with conditional expectation equal to

$$E\big[I(Y_j \leqslant \dot{y}) \bigm| x_j\big] = \eta(x_j). \qquad (2.5.6)$$

with the possibility that the conditional expectation of Y given x is proportional to x, that is

$$E\big[Y_j \bigm| x_j\big] = x_j\beta,$$

as in the case of model (2.3.1).

Chambers, Dorfman and Wehrly present two nonparametric estimators of (2.5.5). The Nadaraya-Watson estimator using a uniform kernel $U(x_i - b, x_i + b)$ is

$$\widehat{F}^{NW}_{CDWr}(\dot{y}) = \sum_{j\in\mathbb{A}} m^{NW}_j I(y_j \leqslant \dot{y}) \tag{2.5.7}$$

where

$$m^{NW}_j = (N-n)^{-1} \sum_{i\in\mathbb{A}^c} I(x_i - b \leqslant x_j \leqslant x_i + b)\Big\{\sum_{k\in\mathbb{A}} I(x_i - b \leqslant x_k \leqslant x_i + b)\Big\}^{-1}.$$

The other nonparametric estimator of $F_r(\cdot)$ is the Gasser-Müller kernel smoother. Assume that the population labels are ordered by increasing value of x: $x_1 \leqslant x_2 \leqslant \ldots \leqslant x_N$, and that $j_1 < j_2 < \ldots < j_n$ are the labels of the units in the sample. Define $a_0 = -\infty$; $a_\ell = 2^{-1}(x_{j_\ell} + x_{j_{\ell+1}})$ for $\ell = 1, \ldots, n-1$; $a_n = +\infty$. The Gasser-Müller estimator of (2.5.5) is

$$\widehat{F}^{GM}_{CDWr}(\dot{y}) = \sum_{j\in\mathbb{A}} m^{GM}_j I(y_j \leqslant \dot{y}) \tag{2.5.8}$$

where

$$m^{GM}_{j_\ell} = b^{-1}(N-n)^{-1} \int_{a_{\ell-1}}^{a_\ell} \sum_{i\in\mathbb{A}^c} K[b^{-1}(x_i - u)]\mathrm{d}u.$$

As a special example of (2.5.8), consider the kernel function $U(-1,1)$. As $b \to 0$, the weights $m^{GM}_{j_\ell}$ go to $(N-n)^{-1}$ times the number of elements in the set $\{i \in \mathbb{A}^c : a_{\ell-1} < x_i \leqslant a_\ell\}$. That is, the weights are equal to the proportion of nonselected units with an x value "close" to x_j for $j \in \mathbb{A}$. Estimators (2.5.7) and (2.5.8) can be incorporated as part of a distribution function estimator. An estimator of the distribution function based on the Nadaraya-Watson estimator (2.5.7) is

$$\widehat{F}^{NW}_{CDW}(\dot{y}) = N^{-1}\Big[\sum_{j\in\mathbb{A}} I(y_j \leqslant \dot{y}) + (N-n)\widehat{F}^{NW}_{CDWr}(\dot{y})\Big].$$

A similar estimator of the distribution function can be derived from the Gasser-Müller estimator (2.5.8).

Consider, in what follows, the Nadaraya-Watson estimator (2.5.7). Similar results apply to the Gasser-Müller estimator (2.5.8). Under (2.5.6), the conditional prediction bias for the Nadaraya-Watson estimator of $F_r(\cdot)$ is

$$E\big[\widehat{F}^{NW}_{CDWr}(\dot{y}) - F_r(\dot{y}) \mid \mathbb{A}_N, \mathcal{F}_{xN}\big] = \sum_{j\in\mathbb{A}} m_j^{NW}\eta(x_j) - (N-n)^{-1}\sum_{i\in\mathbb{A}^c}\eta(x_i), \tag{2.5.9}$$

where $\eta(x_i) = E[I(Y_i \leqslant \dot{y}) \mid x_i]$. Chambers, Dorfman and Wehrly suggest estimating (2.5.9) under model (2.3.1) to produce a calibrated version of estimator (2.5.7),

$$\widehat{E}\big[\widehat{F}^{NW}_{CDWr}(\dot{y}) - F_r(\dot{y}) \,|\, \mathbb{A}_N, \mathcal{F}_{xN}\big] = \sum_{j\in\mathbb{A}} m_j^{NW} G_n\big(h_j^{-1}[\dot{y} - x_j b_n]\big) - \sum_{j\in\mathbb{A}} m_j^{NW} I(y_j \leqslant \dot{y}) \tag{2.5.10}$$

where G_n is defined in (2.3.4) as part of the Chambers and Dunstan estimator. Subtracting the estimate of the bias (2.5.10) from estimator (2.5.7) we get the bias calibrated estimator,

$$\widetilde{F}^{NW}_{CDWr}(\dot{y}) = \widehat{F}^{NW}_{CDWr}(\dot{y}) + \sum_{j\in\mathbb{A}} m_j^{NW}\Big[I(y_j \leqslant \dot{y}) - G_n\big(h_j^{-1}[\dot{y} - x_j b_n]\big)\Big] \tag{2.5.11}$$

Chambers, Dorfman and Wehrly maintain that in the event that model (2.3.1) is approximately true $\widetilde{F}^{NW}_{CDWr}$ should perform better than $\widehat{F}^{NW}_{CDWr}$.

A finite sample model based approach in the bandwidth selector is used by Chambers, Dorfman and Wehrly. The summation in $F_r(\cdot)$ is already a form of smoothing, and using criteria that minimize the integrated square error would lead to oversmoothed results. The bandwidth b for estimator (2.5.7) is chosen to minimize an estimate of the mean square error of prediction under model (2.3.1): $V_s^2 + B_s^2$, where

$$V_s^2 = \sum_{j\in\mathbb{A}} (m_j^{NW})^2 G_n\big(h_j^{-1}[\dot{y} - x_j b_n]\big)\big\{1 - G_n\big(h_j^{-1}[\dot{y} - x_j b_n]\big)\big\}$$

$$B_s = \sum_{j\in\mathbb{A}} m_j^{NW} G_n\big(h_j^{-1}[\dot{y} - x_j b_n]\big) - \sum_{i\in\mathbb{A}^c} G_n\big(h_i^{-1}[\dot{y} - x_i b_n]\big).$$

The bandwidth for estimator (2.5.11) is selected to minimize V_s^2. In practice, Chambers, Dorfman and Wehrly suggest to use a grid of potential bandwidth values and choose the b that minimizes $V_s^2 + B_s^2$ for $\widehat{F}_{CDWr}^{NW}$ and that minimizes V_s^2 for $\widetilde{F}_{CDWr}^{NW}$.

Chambers, Dorfman and Wehrly study a total of 17 estimators in a Monte Carlo simulation involving a population of 430 farms with 50 or more beef cattle. Two models, one that fits the data poorly and one based on transformed x and y that has a better fit are used. Nonparametric regression estimators and bias calibrated estimators are compared to estimators (2.2.1), (2.3.4) and (2.3.13). The performance, measured in mean square error, of the calibrated estimators is, in general, better than the corresponding nonparametric regression estimators. The best results, in terms of mean square error, are achieved by the Chambers and Dunstan estimator (2.3.4) under the better fitting model.

The paper by Dorfman and Hall (1993) works on the large-sample theory for several estimators of the distribution function under simple random sampling without replacement. Three different "schemas" are considered to describe the relationship between y and x:

(1) y has a well defined relation to x, for instance, $Y_i = a + bx_i + \epsilon_i$, where $E(\epsilon_i) = 0$, and all ϵ_i have a common distribution G,

(2) y has an ill defined but smooth relationship to x of the form $Y_i = m(x_i) + \epsilon_i$, with ϵ_i as above,

(3) the function $I(y_i \leqslant \dot{y})$ is more closely related to x_i than y_i. We may have, for instance, $E\{I(Y_i \leqslant \dot{y}) \mid x_i\} = H(x_i)$.

The list of estimators studied under these three schemas includes

- Horvitz-Thompson estimator (2.2.1)
- estimators (2.3.4), $\widehat{F}_{CD}$, and (2.3.13), $\widehat{F}_{RKMdm}$ under schema (1)
- nonparametric regression versions of $\widehat{F}_{CD}$ and $\widehat{F}_{RKMdm}$ under schema (2)
- estimator (2.5.2) proposed by Kuo under schema (3)
- design adjusted Kuo estimator, an analogue to Rao, Kovar and Mantel estimator under schema (3),

$$\widehat{F}_{Kuo,da}(\dot{y}) = \widehat{F}_{HT}(\dot{y}) + N^{-1}\sum_{i\in\mathbb{U}}\widehat{H}(x_i) - (n^{-1} - N^{-1})\sum_{j\in\mathbb{A}}\widehat{H}_j(x_j)$$

where

$$\widehat{H}(x_i) = \sum_{j\in\mathbb{A}} W_{ij} I(y_j \leqslant \dot{y})$$
$$\widehat{H}_j(x_j) = \sum_{k\in\mathbb{A}, k\neq j} W^{kj} I(y_k \leqslant \dot{y}),$$

W_{ij} is defined in (2.5.1) and

$$W^{kj} = K\big(b^{-1}[x_k - x_j]\big)\Big[\sum_{i\in\mathbb{A}, i\neq j} K\big(b^{-1}[x_i - x_j]\big)\Big]^{-1}$$

- nonparametric calibration estimator introduced by Dorfman and Hall

$$\widehat{F}_{DH}(\dot{y}) = N^{-1}\Big[\sum_{j\in\mathbb{A}} I(y_j \leqslant \dot{y}) + \sum_{i\in\mathbb{A}^c} G_n(\dot{y} - \widehat{a} - \widehat{b}x_i) + C_{DH}\Big]$$
$$C_{DH} = \sum_{i\in\mathbb{A}^c}\sum_{j\in\mathbb{A}} W_{ij}\left[I(y_j \leqslant \dot{y}) - G_n(\dot{y} - \widehat{a} - \widehat{b}x_i)\right]$$

Expressions for the asymptotic model bias and model variance of these estimators are computed (see Dorfman and Hall, 1993, Table 1). Dorfman and Hall observe that although some estimators perform better under certain conditions, there are no clear winners. The group of “bias-vulnerable” estimators includes Chambers and Dunstan’s estimator (2.3.4), the naive estimator (2.2.1) and the Rao, Kovar and Mantel’s estimator (2.3.13). The Rao, Kovar and Mantel’s estimator (2.3.13) is also called bias-vulnerable because it is conditionally model biased, conditional on the sample indexes.

2.6 Poststratified Estimation

Nascimento-Silva and Skinner (1995) consider a poststratified estimator of the distribution function with poststrata defined by intervals of x. Poststratified estimation is a widely used method in survey sampling. Furthermore, poststratified estimation may be more robust than model based procedures because it does not depend on a specific model. Consider a partition of $\mathbb{U}$ into G groups $\mathbb{U}_1, \mathbb{U}_2, \ldots, \mathbb{U}_G$ where $i \in \mathbb{U}_g$ if $x_{[g-1]} < x_i \leqslant x_{[g]}$; $x_{[0]} = -\infty$, $x_{[G]} = +\infty$ and $x_{[1]} < x_{[2]} < \ldots < x_{[G-1]}$ are some fixed values. Let $\mathbb{A}_g = \mathbb{A} \cap \mathbb{U}_g, g = 1, 2, \ldots, G$. Let N_g be the number of elements in $\mathbb{U}_g$, and let $\widehat{N}_g = \sum_{j \in \mathbb{A}_g} \pi_j^{-1}$. Then the poststratified estimator of $F_N(\dot{y})$ is

$$\widehat{F}_{PS}(\dot{y}) = N^{-1} \sum_{g=1}^{G} N_g \widehat{N}_g^{-1} \sum_{j \in \mathbb{A}_g} \pi_j^{-1} I(y_j \leqslant \dot{y}) = N^{-1} \sum_{g=1}^{G} N_g \widehat{F}_g(\dot{y}) \tag{2.6.1}$$

where $\widehat{F}_g(\dot{y}) = \widehat{N}_g^{-1} \sum_{j \in \mathbb{A}_g} \pi_j^{-1} I(y_j \leqslant \dot{y})$. In practice, any poststrata without observations are combined with non-empty adjacent strata.

The poststratified estimator $\widehat{F}_{PS}$ defined in (2.6.1) is compared theoretically and in a Monte Carlo study with estimators $\widehat{F}_{HT}$, $\widehat{F}_{CD}$, $\widehat{F}_{RKMr}$, $\widehat{F}_{RKMd}$, $\widehat{F}_{RKMdm}$, $\widehat{F}_{Kuo}$ and $\widehat{F}_{Kuk}$ defined in (2.2.1), (2.3.4), (2.3.10), (2.3.11), (2.3.13), (2.5.2) and (2.5.4) respectively. Let $\widehat{F}$ denote any of the estimators $\widehat{F}_{PS}$, $\widehat{F}_{HT}$, $\widehat{F}_{CD}$, $\widehat{F}_{RKMr}$, $\widehat{F}_{RKMd}$, $\widehat{F}_{RKMdm}$, $\widehat{F}_{Kuo}$ and $\widehat{F}_{Kuk}$ mentioned above. The criteria used for the comparison are:

- Is $\widehat{F}$ monotone, with $\lim_{\dot{y} \to -\infty} \widehat{F}(\dot{y}) = 0$ and $\lim_{\dot{y} \to +\infty} \widehat{F}(\dot{y}) = 1$? The three estimators proposed in Rao et al. (1990): $\widehat{F}_{RKMr}$, $\widehat{F}_{RKMd}$, and $\widehat{F}_{RKMdm}$, fail to meet the monotonicity criterion.

- Does $y = x$ imply $\widehat{F} = F_N$? When $y = x$, estimator (2.6.1) is equal to the distribution function for $\dot{y} = x[g], g = 0, 1, \ldots, G$, but not for general values of $\dot{y}$.

The property holds for $\widehat{F}_{CD}$, $\widehat{F}_{RKMr}$, $\widehat{F}_{RKMd}$ and $\widehat{F}_{RKMdm}$. The estimators that do not equal the distribution function when $y = x$ are $\widehat{F}_{HT}$, $\widehat{F}_{Kuo}$ and $\widehat{F}_{Kuk}$.

- Is there flexibility in the use of auxiliary information? The poststratified estimator $\widehat{F}_{PS}$ only needs the N_g values to be known, that is, the number of units in the population with values of x in certain intervals. The Chambers and Dunstan estimator can be computed when summary information about the number of elements per interval of x is available, as shown in Dunstan and Chambers (1989). When more than one auxiliary variable is available, the estimators that can be extended easily to include such variables are $\widehat{F}_{PS}$, $\widehat{F}_{CD}$, $\widehat{F}_{RKMd}$ and $\widehat{F}_{RKMdm}$. The extension of $\widehat{F}_{RKMr}$, $\widehat{F}_{Kuo}$ and $\widehat{F}_{Kuk}$ to include additional auxiliary variables is not straightforward.

- Is the computation simple? Nascimento-Silva and Skinner adopt the convention that an estimator is simple to compute when the estimator can be written as

$$\widehat{F}(\dot{y}) = \sum_{j \in \mathbb{A}} w_j I(y_j \leqslant \dot{y})$$

  where the w_j do not depend on y_j or on $\dot{y}$. Only the estimators $\widehat{F}_{HT}$, $\widehat{F}_{Kuo}$ and $\widehat{F}_{PS}$ have this property. Ease of computation may become an important issue if we want to compute the distribution function for several values of $\dot{y}$.

- Automatic definition of the estimator. The only estimators for which no models, bandwidths or scaling factors are necessary are $\widehat{F}_{HT}$, $\widehat{F}_{RKMr}$ and $\widehat{F}_{RKMd}$. The poststratified estimator (2.6.1) requires the definition of G and the $x_{[g]}, g = 1, 2, \ldots, G-1$ values.

- Bias. Asymptotic model unbiasedness for estimators $\widehat{F}_{CD}$ and $\widehat{F}_{RKMdm}$ is shown in Chambers and Dunstan (1986) and Rao et al. (1990) respectively. The poststratified estimator is design unbiased provided $N_g > 0$ for all groups.

- Variance. An approximate expression for the variance of estimator (2.6.1) is

$$V\{\widehat{F}_{PS}(\dot{y})\} \doteq N^{-2} \sum_{i<j \in \mathbb{U}} (\pi_i \pi_j - \pi_{ij})(a_i \pi_i^{-1} - a_j \pi_j^{-1})^2 \tag{2.6.2}$$

where $a_i = I(y_i \leqslant \dot{y}) - F_{g_i}(\dot{y})$ and g_i is the group to which unit i belongs. Note that the variance of $\widehat{F}_{PS}$ is zero if all of the y values in each group are either below $\dot{y}$ or above $\dot{y}$, indicating that it is not possible to have a single poststratification that minimizes (2.6.2) for all values of $\dot{y}$. An estimator of (2.6.2) is

$$\widehat{V}\{\widehat{F}_{PS}(\dot{y})\} \doteq N^{-2} \sum_{i<j \in \mathbb{U}} (\pi_i \pi_j - \pi_{ij})\pi_{ij}^{-1}(a_i \pi_i^{-1} - a_j \pi_j^{-1})^2.$$

A Monte Carlo study of the performance of estimators $\widehat{F}_{HT}$, $\widehat{F}_{CD}$, $\widehat{F}_{RKMr}$, $\widehat{F}_{RKMd}$, $\widehat{F}_{RKMdm}$, $\widehat{F}_{Kuo}$, $\widehat{F}_{Kuk}$ and $\widehat{F}_{PS}$ is presented in Nascimento-Silva and Skinner (1995). Two populations are used: the 338 sugar cane farms used in Chambers and Dunstan (1986) and the population of 430 beef cattle farms used in Chambers et al. (1993). The distribution function and the estimators listed are computed at eleven quantiles y_α, corresponding to $\alpha = 1/12,\ 2/12,\ \ldots,\ 11/12$. The average bias, average mean square error, aggregated average bias for the 11 quantiles, aggregated average mean square error for the 11 quantiles and maximum absolute deviation between each estimator and the finite population distribution function are computed. Simple random samples of size 30 and 50 are selected. Three schema of poststratification are considered:

- equal number of units in each poststrata,
- equal aggregate square root of x in each poststrata,
- equal aggregate of x in each poststrata.

The performance of the poststratified estimator is not very good for samples of size 30 and 50. As shown in other papers, the Chambers and Dunstan estimator outperforms the others in terms of aggregate mean square error. The bias component has a relatively large contribution to the mean square error of $\widehat{F}_{CD}$. Nascimento-Silva and Skinner observe that if the bias does not decrease at the same rate as the variance of the Chambers and Dunstan estimator, the relative contribution of the bias to the mean square error will be larger as the sample size increases. In fact, for samples of size 300 from the beef cattle farms population, the performance of the poststratified estimator is roughly the same as the performance of the Chambers and Dunstan estimator.

Fuller (1966) gives an alternative for collapsing empty poststrata with adjacent non-empty poststrata that produces unbiased estimators. Suppose that the population is divided into two poststrata: $\mathbb{U}_1$ and $\mathbb{U}_2$. A simple random sample is selected from the population. Let $\mathbb{A}_g = \mathbb{U}_g \cap \mathbb{A}$, $g = 1, 2$. After the sample is selected, two possible situations are considered:

- case I. Both $\mathbb{A}_1$ and $\mathbb{A}_2$ are non-empty.
- case II. One of the $\mathbb{A}_g$ has no elements.

The poststratified estimator of the population mean $\mu_y = N^{-1} \sum_{i \in \mathbb{U}} y_i$ for case I is

$$\widehat{\mu}_{FPS} = N^{-1}[N_1 \bar{y}_1 + N_2 \bar{y}_2] \tag{2.6.3}$$

where N_g is the number of elements in $\mathbb{U}_g$ and $\bar{y}_g$ is the sample mean for $\mathbb{A}_g$, $g = 1, 2$,

$$\bar{y}_g = n_g^{-1} \sum_{j \in \mathbb{A}_g} y_j$$

and n_g is the number of units in $\mathbb{A}_g$. Estimator (2.6.3) is unbiased for μ_y given that case I has occurred. If one of the strata is empty, estimator $\widehat{\mu}_{FPS}$ is based on the sample mean of the non-empty strata.

$$\begin{aligned} \widehat{\mu}_{FPS} &= D_1 \bar{y}_1 \quad \text{if } \mathbb{A} = \mathbb{A}_1 \\ &= D_2 \bar{y}_2 \quad \text{if } \mathbb{A} = \mathbb{A}_2 \end{aligned}$$

where D_1 and D_2 are chosen such that $\widehat{\mu}_{FPS}$ is conditionally unbiased under case II:

$$P_1^* D_1 \mu_{y1} + P_2^* D_2 \mu_{y2} = \mu_y$$

where P_g^* is the probability that $\mathbb{A}_g$ is non-empty given case II, and $\mu_{yg} = N_g^{-1} \sum_{i \in \mathbb{U}_g} y_i$, $g = 1, 2$. For $\widehat{\mu}_{FPS}$ to be conditionally unbiased given case II for all μ_{y1} and μ_{y2}, we need to have

$$D_g = P_g (P_g^*)^{-1}$$

where $P_g = N^{-1} N_g$ is the proportion of units in stratum g, $g = 1, 2$. Fuller notes that it is unclear under which conditions estimator (2.6.3) will perform better than the customary poststratified estimator formed by collapsing empty strata. According to Fuller, if the procedure is generalized to more than two strata "it is very possible that the unbiased estimator would have a smaller mean square error than the biased collapsing estimator."

Note that if $N_1 = N_2$, under simple random sampling without replacement, $P_1^* = P_2^* = 1/2$. If $N_1 = N_2$, estimator (2.6.3) and the biased collapsing estimator are the same.

The extension of the procedure to more than two strata is done by repeatedly dividing the population into groups of two. In the example presented in Fuller (1966) the population is divided initially into two groups, 1 and 2. Group 1 is divided into strata 11 and 12. Group 2 is divided into groups 21 and 22. Finally, group 22 is subdivided into strata 221 and 222. The five resulting strata are then identified as: 11, 12, 21, 221 and 222. The method described in Fuller (1966) is then applied iteratively to sets of two strata. First, the two strata with 3 digit identification numbers are considered. Strata 221 and 222 are then combined to reconstruct "stratum" 22. The method applied to strata 21 and 22 gives an estimation for stratum 2, and the method applied to strata 11 and 12 gives an estimation for stratum 1. At each stage we are dealing with only 2 strata so that an analogue of estimator (2.6.3) can be used for unbiased estimation. Estimator (2.6.3) can be changed to an estimator of the distribution function by replacing the variable y with the function $I(y \leqslant \dot{y})$.

Wey (1966) presents two estimators for the population mean μ_y that use the ranks of the variable x as auxiliary information. The design used is simple random sampling with a sample of size n selected from a population of size N. In order to simplify notation, assume that the sample is sorted by some auxiliary variable x, and that $\mathbb{A} = \{1, 2, \dots, n\}$. That is, the elements in the sample are labeled 1 to n, and $x_1 < x_2 < \dots < x_n$. Let $z_j, j \in \mathbb{A}$ be the ranks in the population of elements in the sample.

The first estimator is constructed using the x values in the sample to stratify the population into n strata with boundaries defined by $\left[2^{-1}(z_{i-1} + z_i),\ 2^{-1}(z_i + z_{i+1})\right]$ for $i = 1, \dots, n$, where $z_0 = 1 - z_1$ and $z_{n+1} = 2N + 1 - z_n$. The number of elements in Stratum i is $N_i^{[1]} = 2^{-1}(z_{i+1} - z_{i-1})$ for $i = 2, \dots, n-1$. The number of elements in stratum 1 is $N_1^{[1]} = 2^{-1}(z_1 + z_2 - 1)$ and the number of elements in stratum n is $N_n^{[1]} = 2^{-1}(2N + 1 - z_n - z_{n-1})$. Note that some of the stratum sizes may be noninteger. The sum of the stratum sizes is N. The $N_i^{[1]}$ are random variables that depend on the $z_i, i = 1, \dots, n$. The pseudo-poststratified estimator of μ_y for strata of size 1 is defined as

$$\widehat{\mu}_{pW}^{[1]} = n^{-1}(N+1)^{-1}(n+1)\Big\{ \sum_{i=2}^{n-1} N_i^{[1]} y_i + (N+1)(n+1)^{-1}(y_1 + y_n) \Big\} \tag{2.6.4}$$

where the weight given to observations 1 and n has the purpose of reducing the bias in (2.6.4) as an estimator of μ_y. Wey generalizes estimator $\widehat{\mu}_{pW}^{[1]}$ by allowing each stratum to have r sample elements. Assume $n = mr$ and let the stratum boundaries be defined by $\left[2^{-1}(z_{(h-1)r} + z_{(h-1)r+1}),\ 2^{-1}(z_{hr} + z_{hr+1})\right], h = 1, \dots, m$ with $z_0 = 1 - z_1$ and $z_{n+1} = 2N + 1 - z_n$. The number of elements in stratum h is then $N_h^{[r]} = 2^{-1}(z_{hr} + z_{hr+1} -$

$z_{(h-1)r} - z_{(h-1)r+1}$). The pseudo-poststratified estimator for strata of size r is defined as

$$\widehat{\mu}_{pW}^{[r]} = n^{-1}(N+1)^{-1}(n+1)\Big\{\sum_{i=2}^{m-1} N_i^{[r]}\bar{y}_h + r(N+1)(n+1)^{-1}(\bar{y}_1 + \bar{y}_m)\Big\} \tag{2.6.5}$$

where $\bar{y}_h$ is the sample mean of the hth stratum. Wey uses the following linear model to study the properties of $\widehat{\mu}_{pW}^{[r]}$

$$y_i = A + Bz_i + e_i, \tag{2.6.6}$$

where the e_i are uncorrelated random variables given z_i with $E[e_i \,|zi] = 0$ and $V[e_i \,|z_i] = S_e^2$. An expression for the approximate variance of estimator (2.6.5) under model (2.6.6) is given. The optimum value for r can then be determined, under model (2.6.6), by minimizing the approximate variance of $\widehat{\mu}_{pW}^{[r]}$ (Wey, 1966, page 86).

An alternative unbiased estimator of μ_y using the ranks of x as auxiliary information is constructed by averaging conditionally unbiased estimators. Assume that $n = 2m+1$ for some $m \geqslant 2$. To construct an unbiased estimator of μ_y, condition on the even numbered observations $2i, i = 1, \ldots, m$. If observation 2 is given, observation 1 can be seen as a simple random sample of size 1 from the set of observations with ranks of x in the set $\{1, 2, \ldots, z_2 - 1\}$. Given observations 2 and 4, the third observation can be seen as a simple random sample of size 1 from the set of observations with ranks in the set $\{z_2 + 1, \ldots, z_4 - 1\}$. Proceed similarly with the rest of the odd numbered observations. An unbiased estimator of the mean can be constructed by weighting observations $2i+1, i = 2, \ldots, m-1$ by the number of elements in the population that lie between the two contiguous even numbered observations: $(z_{2i+2} - z_{2i} - 1)$. The weights used for observations 1 and n are $(z_2 - 1)$ and $(N - z_{n-1})$ respectively. The estimator

$$\widetilde{\mu}_{W1}^{[2]} = N^{-1}\Big\{\widetilde{T}_{W1} + \sum_{i=1}^{m} y_{2i}\Big\} \tag{2.6.7}$$

is conditionally unbiased for μ_y given observations $2, 4, \ldots, 2m$, where

$$\widetilde{T}_{W1} = \Big[(z_2 - 1)y_1 + \sum_{i=1}^{m-1}(z_{2i+2} - z_{2i} - 1)y_{2i+1} + (N - z_{n-1})y_n\Big].$$

Similarly, we can condition on the odd numbered observations $2i+1, i = 2, 3, \ldots, m-1$ to construct another unbiased estimator of μ_y. The first and last observations are not used in the conditioning operation. Given observation 3, the first two sample elements can be seen as a simple random sample of size two of the $(z_3 - 1)$ elements with $x < x_3$, that is, with ranks in the set $\{1, 2, \ldots, z_3 - 1\}$. The last two observations can be viewed as a simple random sample of size 2 selected from the set of observations with ranks in the set $\{z_{n-2} + 1, z_{n-2} + 2, \ldots, N\}$. The rest of the even numbered observations $2i, i = 2, 3, \ldots, m-1$ are treated as samples of size 1 from the sets with $(z_{2i+1} - z_{2i-1} - 1)$ observations whose values of x are between x_{2i-1} and x_{2i+1} respectively. A second unbiased estimator of μ_y is then

$$\widetilde{\mu}^{[2]}_{W2} = N^{-1}\Big\{\widetilde{T}_{W2} + \sum_{i=1}^{m-1} y_{2i+1}\Big\} \tag{2.6.8}$$

where

$$\widetilde{T}_{W2} = \Big[(z_3 - 1)2^{-1}(y_1 + y_2) + \sum_{i=2}^{m-2}(z_{2i+1} - z_{2i-1} - 1)y_{2i} + (N - z_{n-2})2^{-1}(y_{n-1} + y_n)\Big].$$

Wey suggests to use the mean of estimators (2.6.7) and (2.6.8) as an estimator for μ_y,

$$\widehat{\mu}^{[2]}_{uW} = 2^{-1}(\widetilde{\mu}^{[2]}_{W1} + \widetilde{\mu}^{[2]}_{W2}). \tag{2.6.9}$$

where the superscript [2] denotes that estimator (2.6.9) is the mean of two conditionally unbiased estimators. Estimator (2.6.9) is design unbiased for μ_y.

Wey generalizes estimator (2.6.9) to the case where the unbiased estimator for μ_y is constructed by averaging k conditionally unbiased poststratified estimators, where k may vary between 2 and n. In the extreme case of $k = n$, for instance, each one of the

n unbiased estimators is constructed conditioning on just observation $i, i = 1, 2, \ldots, n$. Assume that $n = km + k - 1$. One of the conditionally unbiased estimators is obtained by conditioning on observations $ki + 1, i = 1, \ldots, m$ as

$$\widetilde{\mu}_{W1}^{[k]} = N^{-1}\Big\{\widetilde{T}_{W1}^{[k]} + \sum_{i=1}^{m} y_{ki+1}\Big\}$$

where

$$\widetilde{T}_{W1}^{[k]} = k^{-1}(z_{k+1} - 1)\sum_{j=1}^{k} y_j + (k-1)^{-1}\sum_{i=2}^{m}(z_{ki+1} - z_{ki-k+1})\sum_{j=1}^{k-1} y_{ki-k+1+j}$$
$$+ (k-2)^{-1}(N - z_{km+1})\sum_{j=1}^{k-2} y_{km+1+j}$$

Similarly, a second conditionally unbiased estimator, $\widetilde{\mu}_{W2}^{[k]}$, is constructed by conditioning on observations $ki, i = 1, \ldots, m$. A third conditionally unbiased estimator, $\widetilde{\mu}_{W3}^{[k]}$, is constructed by conditioning on observations $ki - 1, i = 1, \ldots, m$. The procedure is repeated until each observation, except the first and last, have been used once and only once in the conditioning set (Wey, 1966, page 62). The general unbiased estimator for μ_y is then constructed by averaging the $\widetilde{\mu}_{Wi}^{[k]}$ as

$$\widehat{\mu}_{uW}^{[k]} = k^{-1}\sum_{i=1}^{k}\widetilde{\mu}_{Wi}^{[k]}. \tag{2.6.10}$$

An expression for the approximate variance of estimator (2.6.10) under model (2.6.6) is given. The optimum k can then be determined by minimizing the variance of $\widehat{\mu}_{uW}^{[k]}$ with respect to k (Wey, 1966, page 98). Estimators for the variances of $\widehat{\mu}_{pW}^{[r]}$ and $\widehat{\mu}_{uW}^{[k]}$ are given in Section V of Wey (1966).

2.7 Comments

Some comments on the different articles described in this chapter, particularly those in Section 2.3, are possible.

- The Chambers and Dunstan estimator (2.3.4) outperforms the others when the model used to construct it closely describes the relation between y and x. This estimator does not recognize the sampling design, and performance breaks down when the model is incorrectly specified.

- Model misspecification can occur for: (1) the mean function for y in terms of x, (2) the variance function of the residuals, $V\{h(x_k)U_k\}$ in (2.3.1), and (3) the specification of a common distribution function. Correct specification of the variance may be more difficult to achieve than the correct specification of the mean function.

- Intensive computations seem to be unavoidable in the estimation of the distribution function. The Chambers and Dunstan method requires the computation of $n(N-n)$ imputed values for the variable y as presented in (3.1.1). The $n(N-n)$ imputed values for the variable y may be used to estimate the distribution function at as many points as desired.

3 LOCAL-RESIDUALS ESTIMATOR

3.1 Introduction

The Chambers and Dunstan estimator (2.3.4) can be seen as a weighted average of $n + n(N-n)$ indicator functions. The construction of the estimator is composed of several steps. First the regression coefficient b_n is computed. Then for each element $i \in \mathbb{A}^c$, $\widehat{y}_i = x_i b_n$ is computed. Then n new y values are created for each $i \in \mathbb{A}^c$ by adding to $\widehat{y}_i$ each of the n sample residuals $r_j = y_j - x_j b_n$ multiplied by $h(x_i)h(x_j)^{-1}$ for $j \in \mathbb{A}$. Thus, the new imputed values are

$$\begin{aligned} \widehat{y}_{ij}^{CD} &= x_i b_n + h_i h_j^{-1}(y_j - x_j b_n) \\ &= \widehat{y}_i + h_i h_j^{-1} r_j \qquad (3.1.1) \end{aligned}$$

for $i \in \mathbb{A}$, $j \in \mathbb{A}^c$, where $h_i = h(x_i)$. Estimator (2.3.4) can be computed as

$$\widehat{F}_{CD}(\dot{y}) = N^{-1}\Big[\sum_{j\in\mathbb{A}} I(y_j \leqslant \dot{y}) \; + \; \sum_{i\in\mathbb{A}^c}\sum_{j\in\mathbb{A}} n^{-1} I(\widehat{y}_{ij}^{CD} \leqslant \dot{y})\Big]. \qquad (3.1.2)$$

The same idea of an "imputed" population is used for the local-residuals estimator developed in this section. Instead of using the n residuals for each nonsampled unit i, only residuals from sampled units that are "close" to i are used. Assume that model (2.1.4) holds and that the value for an auxiliary variable, x, is available for each element in the population. For the units in the sample both y and x are available. For simplicity we will assume that there are no duplicate x values.

Suppose that the sample is divided into B groups according to the sorted values of x and that $k = n/B$ is an integer. Each group, denoted by $\mathbb{A}_\ell$, contains k elements of the sample such that $x_{([\ell-1]k+1)}, \ldots, x_{(\ell k)} \in \mathbb{A}_\ell$, $\ell = 1, \ldots, B$, where $x_{(j)}$ denotes the jth sample order statistic of x,

$$\underbrace{x_{(1)}, \ldots, x_{(k)}}_{\text{Group 1}}, \underbrace{x_{(k+1)}, \ldots, x_{(2k)}}_{\text{Group 2}}, \ldots, \underbrace{x_{([B-1]k+1)}, \ldots, x_{(n)}}_{\text{Group } B}. \tag{3.1.3}$$

Elements in the population are divided similarly into B *bins*, denoted by $\mathbb{U}_\ell$. The boundary between bins $\mathbb{U}_\ell$ and $\mathbb{U}_{\ell+1}$ is $(x_{(\ell k)} + x_{(\ell k+1)})/2$, $\ell = 1, \ldots, B-1$. For each $i \in \mathbb{U}$, let ℓ_i be the index of the bin containing unit i. That is, for unit i, $\ell_i = \ell$ if

$$\begin{cases} & x_i \leqslant [x_{(k)} + x_{(k+1)}]/2 & \text{for } \ell = 1 \\ (x_{([\ell-1]k)} + x_{([\ell-1]k+1)})/2 \leqslant & x_i \leqslant [x_{(\ell k)} + x_{(\ell k+1)}]/2 & \text{for } 1 < \ell < B \\ (x_{([B-1]k)} + x_{([B-1]k+1)})/2 \leqslant & x_i & \text{for } \ell = B. \end{cases}$$

Because we are assuming that x is a continuous variable, the inequalities that define the bins are strict inequalities. Similar definition for populations with ties are possible, but in such cases the number of sample elements in each bin will not necessarily be the same.

The local-residuals estimator is defined as

$$\widehat{F}_L(\dot{y}) = N^{-1}\Big[\sum_{j\in\mathbb{A}} I(Y_j \leqslant \dot{y}) + \sum_{i\in\mathbb{A}^c} \widehat{G}_{L\ell_i}(h_i^{-1}[\dot{y} - x_i\widehat{\beta}])\Big] \tag{3.1.4}$$

where

$$\widehat{\beta} = \Big[\sum_{j\in\mathbb{A}} \pi_j^{-1} h_j^{-2} Y_j x_j\Big]\Big[\sum_{j\in\mathbb{A}} \pi_j^{-1} h_j^{-2} x_j^2\Big]^{-1}, \tag{3.1.5}$$

is the Horvitz-Thompson estimator of the finite population parameter β_y, where

$$\beta_y = \Big[\sum_{i\in\mathbb{U}} h_i^{-2} y_i x_i\Big]\Big[\sum_{i\in\mathbb{U}} h_i^{-2} x_i^2\Big]^{-1}.$$

When we consider the finite population as a sample of size N from the superpopulation model, β_y can be regarded as an estimator of the parameter β that appears in model (2.1.4). The estimator of the distribution function of U used in (3.1.4) is

$$\begin{aligned} \widehat{G}_{L\ell_i}(h_i^{-1}[\dot{y} - x_i\widehat{\beta}]) &= \sum_{j\in\mathbb{A}_{\ell_i}} \omega_{ij} I(h_j^{-1}[Y_j - x_j\widehat{\beta}] \leqslant h_i^{-1}[\dot{y} - x_i\widehat{\beta}]) \\ &= \sum_{j\in\mathbb{A}_{\ell_i}} \omega_{ij} I(x_i\widehat{\beta} + h_i h_j^{-1}[Y_j - x_j\widehat{\beta}] \leqslant \dot{y}) \end{aligned} \tag{3.1.6}$$

with

$$\omega_{ij} = \pi_j^{-1}\Big[\sum_{j'\in\mathbb{A}_{\ell_i}} \pi_{j'}^{-1}\Big]^{-1}. \tag{3.1.7}$$

Each of the estimators $\widehat{G}_{L\ell}(\cdot), \ell = 1, \ldots, B$, uses k observations from the sample. The estimator $G_n(\cdot)$ of the distribution function of U, introduced in the Chambers and Dunstan estimator (2.3.4), uses all n observations from the sample. When estimating $G(h_i^{-1}[\dot{y} - x_i\beta])$ for each $i \in \mathbb{A}^c$, estimator $\widehat{G}_{L\ell}(\cdot)$ is expected to be more robust than estimator $G_n(\cdot)$ against model misspecifications since $\widehat{G}_{L\ell}(\cdot)$ only uses the k sampled observations that are close to unit i.

Imputed values of y can be computed as in (3.1.1),

$$\widehat{y}_{ij} = x_i\widehat{\beta} + h_i h_j^{-1}[y_j - x_j\widehat{\beta}]. \tag{3.1.8}$$

Using $\widehat{y}_{ij}$, estimator (3.1.4) can be rewritten as

$$\widehat{F}_L(\dot{y}) = N^{-1}\Big[\sum_{j\in\mathbb{A}} I(y_j \leqslant \dot{y}) + \sum_{i\in\mathbb{A}^c}\sum_{j\in\mathbb{A}_{\ell_i}} \omega_{ij} I(\widehat{y}_{ij} \leqslant \dot{y})\Big]. \tag{3.1.9}$$

For each $i \in \mathbb{A}^c$, k new $\widehat{y}_{ij}$ are imputed using the k residuals from the sample points with $j \in \mathbb{A}_{\ell_i}$. A total of $k(N-n)$ imputed values are computed. Note that if all π_j are equal, as in simple random sampling, $\omega_{ij} \equiv k^{-1}$. Furthermore, if $k = n$, i.e., $B = 1$, estimators (3.1.2) and (3.1.9) are identical.

3.2 Limiting Distribution of $\widehat{F}_L(\dot{y}) - F_N(\dot{y})$ Conditioning on the Sample

We first study the sampling distribution of the local-residuals estimator conditional on $\mathbb{A}_N$ under alternative superpopulation model assumptions. More precisely, since from a superpopulation perspective, both $\widehat{F}_L(\dot{y})$ and $F_N(\dot{y})$ are random variables, we will study the distribution of the estimation error

$$\widehat{F}_L(\dot{y}) - F_N(\dot{y}), \tag{3.2.1}$$

for a fixed $\dot{y}$, where $\widehat{F}_L(\dot{y})$ is defined in (3.1.4) and $F_N(\dot{y})$ is defined in (2.1.2). In subsections 3.2.1 through 3.2.4 we will consider $\widehat{F}_L(\dot{y})$ and $F_N(\dot{y})$ to be functions of the random variables $Y_i, i \in \mathbb{U}$.

3.2.1 Case A: $E(Y_i) = x_i\beta$ and $V(Y_i) = h(x_i)^2\sigma^2$; β and $h(x_i)$ known

We first derive the limiting distribution of the estimation error under the assumption that the parameters of the superpopulation model are known. Theorem 3.2.1 presents the model mean and model variance of the estimation error $\widehat{F}_L(\dot{y}) - F_N(\dot{y})$ for a fixed point $\dot{y}$. In Theorem 3.2.2 model consistency of $\widehat{F}_L(\dot{y})$ for $F_N(\dot{y})$ and the limiting distribution of $\widehat{F}_L(\dot{y}) - F_N(\dot{y})$ are shown.

Theorem 3.2.1 *Let $\mathbb{A}_N$ be a sample of size n selected from the finite population $\mathbb{U}_N$ of size N. Assume that the sample is divided into B groups, each of size k, as described in (3.1.3). Assume the superpopulation model*

$$Y_i = x_i\beta + h(x_i)U_i \tag{3.2.2}$$

for $i \in \mathbb{U}$, with $h(x)$ known. Let $h_i = h(x_i)$. The U_i are independent and identically distributed with $E(U_i) = 0$, $V(U_i) = \sigma^2$ and distribution function $G(u)$. Let $\widetilde{F}_{L\beta}(\dot{y})$ be the estimator (3.1.4) with the true β used in (3.1.6). Then for a fixed point $\dot{y}$,

(a) Estimator $\widetilde{F}_{L\beta}(\dot{y})$ is model unbiased for the finite population distribution function $F_N(\dot{y})$ defined in (2.1.2).

(b) The model variance of the estimation error is

$$V\big(\widetilde{F}_{L\beta}(\dot{y}) - F_N(\dot{y}) \bigm| \mathbb{A}_N\big) = V\big(T_L \bigm| \mathbb{A}_N\big) + V\big(T_N \bigm| \mathbb{A}_N\big) \tag{3.2.3}$$

where

$$V\big(T_L \bigm| \mathbb{A}_N\big) = N^{-2} \sum_{\ell=1}^{B} \sum_{j \in \mathbb{A}_\ell} \sum_{i_1, i_2 \in \mathbb{U}_\ell - \mathbb{A}_\ell} \omega_{i_1 j} \omega_{i_2 j} \big[G\big(\min[\dot{u}_{i_1}, \dot{u}_{i_2}]\big) - G\big(\dot{u}_{i_1}\big) G\big(\dot{u}_{i_2}\big)\big]$$

and

$$V\big(T_N \bigm| \mathbb{A}_N\big) = N^{-2} \sum_{i \in \mathbb{A}^c} G(\dot{u}_i)\big[1 - G(\dot{u}_i)\big],$$

with $\omega_{ij} = \pi_j^{-1} \sum_{j' \in \mathbb{A}_{\ell_i}} \pi_{j'}^{-1}$ defined in (3.1.7) and $\dot{u}_i = h_i^{-1}(\dot{y} - x_i\beta)$.

Proof. Part *(a)*. The error in the estimated distribution function is

$$\widetilde{F}_{L\beta}(\dot{y}) - F_N(\dot{y}) = N^{-1}\Big[\sum_{i \in \mathbb{A}^c} \widehat{G}_{L\ell_i}(h_i^{-1}[\dot{y} - x_i\beta]) - \sum_{i \in \mathbb{A}^c} I(Y_i \leqslant \dot{y})\Big], \tag{3.2.4}$$

where $\widehat{G}_{L\ell_i}$ is defined in (3.1.6). Note that the first term depends only on the sample observations while the second term depends only on nonsampled observations. Under model (3.2.2), the Y_i, $i \in \mathbb{U}$, are conditionally independent given $\mathbb{A}_N$. Let the two terms in (3.2.4) be

$$T_L = N^{-1} \sum_{i \in \mathbb{A}^c} \widehat{G}_{L\ell_i}(h_i^{-1}[\dot{y} - x_i\beta]) \tag{3.2.5}$$

and

$$T_N = N^{-1} \sum_{i\in\mathbb{A}^c} I(Y_i \leqslant \dot{y}), \tag{3.2.6}$$

where the subindex L or N denotes whether the term comes from the definition of $\widetilde{F}_{L\beta}$ or from the definition of F_N respectively.

The conditional expected value of (3.2.4) under model (3.2.2) is

$$E\big[\widetilde{F}_{L\beta}(\dot{y}) - F_N(\dot{y}) \bigm| \mathbb{A}_N\big] = E\big[T_L \bigm| \mathbb{A}_N\big] - E\big[T_N \bigm| \mathbb{A}_N\big]. \tag{3.2.7}$$

The second term in (3.2.7) is

$$\begin{aligned} E\big[T_N \bigm| \mathbb{A}_N\big] &= N^{-1} \sum_{i\in\mathbb{A}^c} E\big[I\big(Y_i \leqslant \dot{y}\big) \bigm| \mathbb{A}_N\big] \\ &= N^{-1} \sum_{i\in\mathbb{A}^c} E\big[I\big(Y_i \leqslant \dot{y}\big)\big], \end{aligned}$$

since $E\big[I\big(Y_i \leqslant \dot{y}\big)\big]$ does not depend on whether unit i belongs to the sample or not. Then

$$\begin{aligned} E\big[T_N \bigm| \mathbb{A}_N\big] &= N^{-1} \sum_{i\in\mathbb{A}^c} E\big[I\big(h_i^{-1}[Y_i - x_i\beta] \leqslant h_i^{-1}[\dot{y} - x_i\beta]\big)\big] \\ &= N^{-1} \sum_{i\in\mathbb{A}^c} E\big[I\big(U_i \leqslant \dot{u}_i\big)\big] \\ &= N^{-1} \sum_{i\in\mathbb{A}^c} G(\dot{u}_i). \end{aligned} \tag{3.2.8}$$

The first term in (3.2.7) is

$$\begin{aligned} E\big[T_L \bigm| \mathbb{A}_N\big] &= N^{-1} \sum_{i\in\mathbb{A}^c} E\big[\widehat{G}_{L\ell_i}(h_i^{-1}[\dot{y} - x_i\beta]) \bigm| \mathbb{A}_N\big] \\ &= N^{-1} \sum_{i\in\mathbb{A}^c} \sum_{j\in\mathbb{A}_{\ell_i}} \omega_{ij} E\big[I\big(U_j \leqslant \dot{u}_i\big) \bigm| \mathbb{A}_N\big] \\ &= N^{-1} \sum_{i\in\mathbb{A}^c} \sum_{j\in\mathbb{A}_{\ell_i}} \omega_{ij} E\big[I\big(U_j \leqslant \dot{u}_i\big)\big] \\ &= N^{-1} \sum_{i\in\mathbb{A}^c} \sum_{j\in\mathbb{A}_{\ell_i}} \omega_{ij} G(\dot{u}_i) \\ &= N^{-1} \sum_{i\in\mathbb{A}^c} G(\dot{u}_i) \end{aligned} \tag{3.2.9}$$

since $\sum_{j\in\mathbb{A}_{\ell_i}}\omega_{ij}=1$ by construction and $E\big[I(U_j\leqslant\dot u_i)\big]=E\big[I(U_i\leqslant\dot u_i)\big]=G(\dot u_i)$ by model (3.2.2). Combining results (3.2.8) and (3.2.9) we have that the local-residuals estimator is conditionally unbiased for $F_N(\dot y)$, that is,

$$E\big[\widetilde F_{L\beta}(\dot y)-F_N(\dot y)\;\big|\;\mathbb{A}_N\big]=0, \tag{3.2.10}$$

which proves part *(a)*. In the survey sampling literature this property is called model unbiasedness.

Part *(b)*. Since T_L and T_N defined in (3.2.5) and (3.2.6) are conditionally independent given $\mathbb{A}_N$, the conditional variance of (3.2.4) under model (3.2.2) is

$$V\big(\widetilde F_{L\beta}(\dot y)-F_N(\dot y)\;\big|\;\mathbb{A}_N\big)=V\big(T_L\;\big|\;\mathbb{A}_N\big)+V\big(T_N\;\big|\;\mathbb{A}_N\big) \tag{3.2.11}$$

which proves (3.2.3). The second variance in (3.2.11) is

$$\begin{aligned}
V\big(T_N\;\big|\;\mathbb{A}_N\big) &= N^{-2}V\big\{\sum_{i\in\mathbb{A}^c}I(Y_i\leqslant\dot y)\;\big|\;\mathbb{A}_N\big\}\\
&= N^{-2}\sum_{i\in\mathbb{A}^c}V\big\{I(Y_i\leqslant\dot y)\big\}\\
&= N^{-2}\sum_{i\in\mathbb{A}^c}V\big\{I(U_i\leqslant\dot u_i\big\})\\
&= N^{-2}\sum_{i\in\mathbb{A}^c}G(\dot u_i)\big[1-G(\dot u_i)\big].
\end{aligned} \tag{3.2.12}$$

The first term on the right hand side of (3.2.11) is

$$\begin{aligned}
V\big(T_L\;\big|\;\mathbb{A}_N\big) &= N^{-2}V\big(\sum_{i\in\mathbb{A}^c}\widehat G_{L\ell_i}\big(h_i^{-1}[\dot y-x_i\widehat\beta]\big)\;\big|\;\mathbb{A}_N\big)\\
&= N^{-2}V\big(\sum_{i\in\mathbb{A}^c}\widehat G_{L\ell_i}\big(\dot u_i\big)\;\big|\;\mathbb{A}_N\big)\\
&= N^{-2}V\big(\sum_{i\in\mathbb{A}^c}\sum_{j\in\mathbb{A}_{\ell_i}}\omega_{ij}I\big(U_j\leqslant\dot u_i\big)\;\big|\;\mathbb{A}_N\big)\\
&= N^{-2}V\big(\sum_{\ell=1}^{B}\sum_{j\in\mathbb{A}_\ell}\sum_{i\in\mathbb{U}_\ell-\mathbb{A}_\ell}\omega_{ij}I\big(U_j\leqslant\dot u_i\big)\;\big|\;\mathbb{A}_N\big)\\
&= N^{-2}\sum_{\ell=1}^{B}\sum_{j\in\mathbb{A}_\ell}V\big(\sum_{i\in\mathbb{U}_\ell-\mathbb{A}_\ell}\omega_{ij}I\big(U_j\leqslant\dot u_i\big)\;\big|\;\mathbb{A}_N\big),
\end{aligned}$$

because under model (3.2.2) the U_j are conditionally independent. The set $\mathbb{U}_\ell - \mathbb{A}_\ell = \mathbb{U}_\ell \cap \mathbb{A}^c$ contains the subindices of nonsampled elements in bin ℓ. Then

$$\begin{aligned} V(T_L \mid \mathbb{A}_N) &= N^{-2}\sum_{\ell=1}^{B}\sum_{j\in\mathbb{A}_\ell}\sum_{i_1,i_2\in\mathbb{U}_\ell-\mathbb{A}_\ell} \omega_{i_1 j}\omega_{i_2 j}\mathrm{Cov}\big[I(U_j \leqslant \dot{u}_{i_1}), I(U_j \leqslant \dot{u}_{i_2}) \mid \mathbb{A}_N\big] \\ &= N^{-2}\sum_{\ell=1}^{B}\sum_{j\in\mathbb{A}_\ell}\sum_{i_1,i_2\in\mathbb{U}_\ell-\mathbb{A}_\ell} \omega_{i_1 j}\omega_{i_2 j}\mathrm{Cov}\big[I(U_j \leqslant \dot{u}_{i_1}), I(U_j \leqslant \dot{u}_{i_2})\big] \\ &= N^{-2}\sum_{\ell=1}^{B}\sum_{j\in\mathbb{A}_\ell}\sum_{i_1,i_2\in\mathbb{U}_\ell-\mathbb{A}_\ell} \omega_{i_1 j}\omega_{i_2 j}\big[G(\min[\dot{u}_{i_1}, \dot{u}_{i_2}]) - G(\dot{u}_{i_1})G(\dot{u}_{i_2})\big] \end{aligned} \tag{3.2.13}$$

By combining (3.2.13) and (3.2.12) we have the result. ▲

Theorem 3.2.1 shows that the local-residuals estimator (3.1.4) is model unbiased for the finite population distribution function and has model variance given by (3.2.3). No asymptotic assumptions are made in Theorem 3.2.1. In Theorem 3.2.2 we consider a sequence of samples and finite populations indexed by N as described in Section 2.1.2. The superpopulation model (3.2.14) assumed in Theorem 3.2.2 does not change with N.

Theorem 3.2.2 *Let $\{\mathbb{A}_N\}$ be a sequence of samples selected from the sequence of finite populations $\{\mathbb{U}_N\}$. Assume that the sample $\mathbb{A}_N$ is divided into B_N groups, each of size k_N, as described in (3.1.3). Assume the superpopulation model*

$$Y_i = x_i\beta + h(x_i)U_i \tag{3.2.14}$$

for $i \in \mathbb{U}_N$, with $h(x)$ known. The U_i are independent and identically distributed with $E(U_i) = 0$, $V(U_i) = \sigma^2$ and distribution function $G(u)$. Let $\dot{y}$ be a fixed point. Let $h_i = h(x_i)$ and $\dot{u}_i = h_i^{-1}[\dot{y} - x_i\beta]$. Assume

A.1a *The number of indices in $\mathbb{A}_N$, denoted by n_N, is such that*

$$n_{N+1} \geqslant n_N,$$

$$0 \leqslant \lim_{N\to\infty} N^{-1} n_N = f_c < 1.$$

A.2a *The number of bins and the number of sample elements per bin satisfy*

$$B_N^{-1} = O(N^{-\alpha})$$

$$k_N^{-1} = O(N^{-(1-\alpha)})$$

where $n_N = B_N k_N$ *and* $0 < \alpha < 1$.

A.3a *There exist* L_1 *and* L_2 *such that the number of population elements per bin,* $K_{N\ell}$*, satisfy*

$$0 < L_1 k_N < K_{N\ell} < L_2 k_N < \infty$$

for all $\ell = 1, \ldots, B_N$*, and*

$$\sum_{\ell=1}^{B_N} K_{N\ell} = N.$$

A.4a *There exist* L_1^* *and* L_2^* *such that*

$$0 < L_1^* k_N^{-1} < \omega_{ij} < L_2^* k_N^{-1} < \infty,$$

for $i \in \mathbb{A}_N^c$, $j \in \mathbb{A}_{\ell_i}$, *where* $\omega_{ij} = \pi_j^{-1}\Big[\sum_{j' \in \mathbb{A}_{\ell_i}} \pi_{j'}^{-1}\Big]^{-1}$ *is defined in (3.1.7), and* π_j^{-1} *are the sample weights.*

A.5a *The term* $\Big\{(N - n_N)^{-1} \sum_{i \in \mathbb{A}_N^c} G(\dot{u}_i)[1 - G(\dot{u}_i)]\Big\}$ *is positive for all* N.

A.6a *The term* $\Big\{N^{-1} \sum_{j \in \mathbb{A}_N} V(Z_j \bigm| \mathbb{A}_N)\Big\}$ *is positive for all* N*, where*

$$V(Z_j \bigm| \mathbb{A}_N) = \sum_{i_1, i_2 \in \mathbb{U}_\ell - \mathbb{A}_\ell} \omega_{i_1 j} \omega_{i_2 j} \big[G\big(\min[\dot{u}_{i_1}, \dot{u}_{i_2}]\big) - G\big(\dot{u}_{i_1}\big) G\big(\dot{u}_{i_2}\big)\big],$$

$Z_j = \sum_{i \in \mathbb{U}_\ell - \mathbb{A}_\ell} \omega_{ij} I\big(U_j \leqslant \dot{u}_i\big)$ *for* $j \in \mathbb{A}_N$, ℓ *is the bin that contains unit* j, $\mathbb{U}_\ell$ *is the set of indices in bin* ℓ *and* $\mathbb{A}_\ell = \mathbb{A}_N \cap \mathbb{U}_\ell$. *The subindex* N *has been omitted in* $\mathbb{U}_\ell$ *and* $\mathbb{A}_\ell$ *to simplify notation.*

Let $\widetilde{F}_{L\beta}(\dot{y})$ be the estimator (3.1.4) with the true β used in (3.1.6). The subindex N has been omitted in $\widetilde{F}_{L\beta}(\dot{y})$ to simplify notation. Then,

(a) The estimator $\widetilde{F}_{L\beta}(\dot{y})$ satisfies

$$\lim_{N\to\infty} P\big(\big|\widetilde{F}_{L\beta}(\dot{y}) - F_N(\dot{y})\big| > \epsilon \;\big|\; \mathbb{A}_N\big) = 0$$

for all $\epsilon > 0$, where $F_N(\dot{y})$ is defined in (2.1.2).

(b) The sequence

$$\Big\{V\big(\widetilde{F}_{L\beta}(\dot{y}) - F_N(\dot{y}) \;\big|\; \mathbb{A}_N\big)\Big\}^{-1/2}\Big\{\widetilde{F}_{L\beta}(\dot{y}) - F_N(\dot{y})\Big\}$$

converges in distribution to a $N(0,1)$ random variable given $\mathbb{A}_N$, where $V\big(\widetilde{F}_{L\beta}(\dot{y}) - F_N(\dot{y}) \;\big|\; \mathbb{A}_N\big)$ is given in Theorem 3.2.1 for a population of size N.

Proof. Part *(a)*. By Theorem 3.2.1 we have that

$$E\Big[\widetilde{F}_{L\beta}(\dot{y}) - F_N(\dot{y}) \;\Big|\; \mathbb{A}_N\Big] = 0.$$

Then, to prove model consistency of $\widetilde{F}_{L\beta}(\dot{y}) - F_N(\dot{y})$ we need to show that the model variance of $\widetilde{F}_{L\beta}(\dot{y}) - F_N(\dot{y})$ given in (3.2.3) converges to zero as $N \to \infty$. The two variances that appear in (3.2.3), $V\Big[T_L \;\Big|\; \mathbb{A}_N\Big]$ and $V\Big[T_N \;\Big|\; \mathbb{A}_N\Big]$, are given in (3.2.13) and (3.2.12) respectively. The second component of (3.2.3) is

$$\begin{aligned} V\Big[T_N \;\Big|\; \mathbb{A}_N\Big] &= N^{-2}\sum_{i\in\mathbb{A}^c} G(\dot{u}_i)\big[1 - G(\dot{u}_i)\big] \\ &\leqslant N^{-2}\sum_{i\in\mathbb{A}^c} 4^{-1} = N^{-2}(N-n)4^{-1} = O(N^{-1}), \end{aligned} \tag{3.2.15}$$

since $G(\dot{u}_i)\big[1 - G(\dot{u}_i)\big] \leqslant 4^{-1}$ for all $i \in \mathbb{A}^c$ and the number of terms in $\sum_{i\in\mathbb{A}^c}$ is $N - n$. The first term in (3.2.3), given in (3.2.13), is

$$V\Big[T_L \;\Big|\; \mathbb{A}_N\Big] = N^{-2}\sum_{\ell=1}^{B}\sum_{j\in\mathbb{A}_\ell}\sum_{i_1,i_2\in\mathbb{U}_\ell-\mathbb{A}_\ell} \omega_{i_1j}\omega_{i_2j}\big[G\big(\min[\dot{u}_{i_1},\dot{u}_{i_2}]\big) - G\big(\dot{u}_{i_1}\big)G\big(\dot{u}_{i_2}\big)\big].$$

The factors $G(\min[\dot{u}_{i_1},\dot{u}_{i_2}])-G(\dot{u}_{i_1})G(\dot{u}_{i_2})$ are the covariances between two indicator functions, $I(U_j \leqslant \dot{u}_{i_1})$ and $I(U_j \leqslant \dot{u}_{i_2})$, then,

$$\Big|G(\min[\dot{u}_{i_1},\dot{u}_{i_2}])-G(\dot{u}_{i_1})G(\dot{u}_{i_2})\Big| \leqslant 4^{-1}.$$

The model variance of T_L is then,

$$\begin{aligned}
V\Big[T_L \mid \mathbb{A}_N\Big] &= N^{-2}\sum_{\ell=1}^{B}\sum_{j\in\mathbb{A}_\ell}\sum_{i_1,i_2\in\mathbb{U}_\ell-\mathbb{A}_\ell}\omega_{i_1j}\omega_{i_2j}\Big[G(\min[\dot{u}_{i_1},\dot{u}_{i_2}])-G(\dot{u}_{i_1})G(\dot{u}_{i_2})\Big] \\
&= \Big|N^{-2}\sum_{\ell=1}^{B}\sum_{j\in\mathbb{A}_\ell}\sum_{i_1,i_2\in\mathbb{U}_\ell-\mathbb{A}_\ell}\omega_{i_1j}\omega_{i_2j}\Big[G(\min[\dot{u}_{i_1},\dot{u}_{i_2}])-G(\dot{u}_{i_1})G(\dot{u}_{i_2})\Big]\Big| \\
&\leqslant N^{-2}\sum_{\ell=1}^{B}\sum_{j\in\mathbb{A}_\ell}\sum_{i_1,i_2\in\mathbb{U}_\ell-\mathbb{A}_\ell}\omega_{i_1j}\omega_{i_2j}\Big|\Big[G(\min[\dot{u}_{i_1},\dot{u}_{i_2}])-G(\dot{u}_{i_1})G(\dot{u}_{i_2})\Big]\Big| \\
&\leqslant N^{-2}\sum_{\ell=1}^{B}\sum_{j\in\mathbb{A}_\ell}\sum_{i_1,i_2\in\mathbb{U}_\ell-\mathbb{A}_\ell}\omega_{i_1j}\omega_{i_2j}4^{-1} \\
&= 4^{-1}N^{-2}\sum_{\ell=1}^{B}\sum_{j\in\mathbb{A}_\ell}\Big(\sum_{i\in\mathbb{U}_\ell-\mathbb{A}_\ell}\omega_{ij}\Big)^2 \qquad (3.2.16)
\end{aligned}$$

The sum $\Big(\sum_{i\in\mathbb{U}_\ell-\mathbb{A}_\ell}\omega_{ij}\Big)$ is $O(1)$ since by A.4a the ω_{ij} are $O(k_N^{-1})$ and the number of summands is $O(k_N)$ by A.3a. Since $\sum_{\ell=1}^{B}\sum_{j\in\mathbb{A}_\ell} = \sum_{j\in\mathbb{A}}$ has n terms, the order of (3.2.16) is then $O(N^{-2}n)=O(N^{-1})$. Then,

$$V\Big[T_L \mid \mathbb{A}_N\Big] = O(N^{-1}) \qquad (3.2.17)$$

Then, by (3.2.17) and (3.2.15),

$$V\Big[\widetilde{F}_{L\beta}(\dot{y})-F_N(\dot{y}) \mid \mathbb{A}_N\Big] = O(N^{-1}),$$

and

$$\lim_{N\to\infty} V\Big[\widetilde{F}_{L\beta}(\dot{y})-F_N(\dot{y}) \mid \mathbb{A}_N\Big] = 0,$$

which proves the model consistency of $\widetilde{F}_{L\beta}(\dot{y})-F_N(\dot{y})$ stated in part *(a).*

Part *(b)*. Conditioning on $\mathbb{A}_N$, T_L and T_N are independent. The term T_N is the sum of $N - n_N$ independent random variables multiplied by N^{-1}. Assumption A.5a is sufficient for the Lyapounov condition for the sum $\sum_{i\in\mathbb{A}_N^c} I(U_i \leqslant \dot{u}_i)$ because all moments exist for the indicator functions. It follows that

$$\Big\{V\Big[T_L \mid \mathbb{A}_N\Big]\Big\}^{-1/2}\Big\{T_L - E\Big[T_L \mid \mathbb{A}_N\Big]\Big\} =$$

$$= \Big\{N^{-2}\sum_{i\in\mathbb{A}_N^c} G(\dot{u}_i)[1-G(\dot{u}_i)]\Big\}^{-1/2}\Big\{N^{-1}\sum_{i\in\mathbb{A}_N^c}\big[I(U_i \leqslant \dot{u}_i) - G(\dot{u}_i)\big]\Big\}$$

$$= \Big\{\sum_{i\in\mathbb{A}_N^c} G(\dot{u}_i)[1-G(\dot{u}_i)]\Big\}^{-1/2}\Big\{\sum_{i\in\mathbb{A}_N^c}\big[I(U_i \leqslant \dot{u}_i) - G(\dot{u}_i)\big]\Big\}$$

converges in distribution to a standard normal as $N \to \infty$. Analogously, T_L is the sum of n independent random variables multiplied by N^{-1}. The random variable T_L defined in (3.2.5) can be written as

$$\begin{aligned} T_L &= N^{-1}\sum_{i\in\mathbb{A}^c} \widehat{G}_{L\ell_i}(h_i^{-1}[\dot{y} - x_i\beta]) \\ &= N^{-1}\sum_{i\in\mathbb{A}^c}\sum_{j\in\mathbb{A}_{\ell_i}} \omega_{ij} I\big(U_j \leqslant \dot{u}_i\big) \\ &= N^{-1}\sum_{\ell=1}^{B}\sum_{j\in\mathbb{A}_\ell}\sum_{i\in\mathbb{U}_\ell-\mathbb{A}_\ell} \omega_{ij} I\big(U_j \leqslant \dot{u}_i\big) \\ &= N^{-1}\sum_{\ell=1}^{B}\sum_{j\in\mathbb{A}_\ell} Z_j \\ &= N^{-1}\sum_{j\in\mathbb{A}_N} Z_j, \end{aligned}$$

where the $Z_j = \sum_{i\in\mathbb{U}_\ell-\mathbb{A}_\ell} \omega_{ij} I\big(U_j \leqslant \dot{u}_i\big)$ are defined in A.6a for $j \in \mathbb{A}_N$. Given $\mathbb{A}_N$, the Z_j are independent random variables with

$$E(Z_j \mid \mathbb{A}_N) = \sum_{i\in\mathbb{U}_\ell-\mathbb{A}_\ell} \omega_{ij} G(\dot{u}_i)$$

and

$$V(Z_j \mid \mathbb{A}_N) = \sum_{i_1,i_2\in\mathbb{U}_\ell-\mathbb{A}_\ell} \omega_{i_1 j}\omega_{i_2 j}\big[G\big(\min[\dot{u}_{i_1}, \dot{u}_{i_2}]\big) - G\big(\dot{u}_{i_1}\big)G\big(\dot{u}_{i_2}\big)\big].$$

The variables Z_j are a linear combination of indicator functions. Condition A.6a is sufficient for the Lyapounov condition for the sum $\sum_{j\in\mathbb{A}_N} Z_j$ because all moments exist for the Z_J. It follows that

$$\Big\{V\Big[T_N \mid \mathbb{A}_N\Big]\Big\}^{-1/2}\Big\{T_N - E\Big[T_N \mid \mathbb{A}_N\Big]\Big\} = \\ = \Big\{\sum_{j\in\mathbb{A}} V\Big[Z_J \mid \mathbb{A}_N\Big]\Big\}^{-1/2}\Big\{\sum_{j\in\mathbb{A}} Z_j - \sum_{j\in\mathbb{A}} E\Big[Z_j \mid \mathbb{A}_N\Big]\Big\}, \tag{3.2.18}$$

converges in distribution to a standard normal as $N \to \infty$. Finally, since T_L and T_N are conditionally independent given $\mathbb{A}_N$, and

$$\widetilde{F}_{L\beta}(\dot{y}) - F_N(\dot{y}) = T_L - T_N,$$

the sequence of random variables

$$\Big\{V\big[\widetilde{F}_{L\beta}(\dot{y}) - F_N(\dot{y}) \mid \mathbb{A}_N\big]\Big\}^{-1/2}\Big\{\widetilde{F}_{L\beta}(\dot{y}) - F_N(\dot{y})\Big\}$$

converges in distribution to a $N(0,1)$ given $\mathbb{A}_N$ as $N \to \infty$. ▲

3.2.2 Case B: $E(Y_i) = x_i\beta$ and $V(Y_i) = q(x_i)^2\sigma^2$; β known

Chambers and Dunstan (1986) showed that estimator (2.3.4) is no longer model unbiased when the variance function of Y given x is misspecified. If the conditional variance of Y given x is not the same function specified in the construction of estimator (2.3.4), $\widehat{F}_{CD}(\dot{y}) - F_N(\dot{y})$ is still asymptotically normal, but with mean different from zero (see Chambers and Dunstan, 1986, Section 3.2). In practice, the variance function of Y given x may be more difficult to specify correctly than the mean function of Y given x. Thus, it is important to study the sampling distribution of $\widehat{F}_L(\dot{y}) - F_N(\dot{y})$ under misspecification of the variance function. Theorem 3.2.3 contains some results for the misspecified case. The subindex N is often omitted in the discussion to simplify notation.

Theorem 3.2.3 *Let $\{\mathbb{A}_N\}$ be a sequence of samples selected from the sequence of finite populations $\{\mathbb{U}_N\}$. Assume that the sample $\mathbb{A}_N$ is divided into B_N groups, each of size k_N, as described in (3.1.3). Assume the superpopulation model*

$$Y_i = x_i\beta + q(x_i)U_i \tag{3.2.19}$$

for $i \in \mathbb{U}_N$, with $q(x) \neq h(x)$, where $h(x)$ is the function used in constructing estimator (3.1.4). Let $h_i = h(x_i)$ and $q_i = q(x_i)$. The $U_i = q_i^{-1}[Y_i - x_i\beta]$ are independent and identically distributed with $E(U_i) = 0$, $V(U_i) = \sigma^2$ and distribution function $G(u)$. There exists an m_h such that $0 < m_h < h(x_i) < \infty$ for $i \in \mathbb{U}_N$. There exists an M_x such that $|x_i| < M_x$ for $i \in \mathbb{U}_N$. Let $\dot{y}$ be a fixed point. Let $\dot{q}_i = q_i^{-1}[\dot{y} - x_i\beta]$. Assume A.1a *through* A.4a *from Theorem 3.2.2. Also assume*

A.5b *The term $\left\{(N - n_N)^{-1} \sum_{i \in \mathbb{A}_N^c} G(\dot{q}_i)[1 - G(\dot{q}_i)]\right\}$ is positive for all N.*

A.6b *The term $\left\{N^{-1} \sum_{j \in \mathbb{A}_N} V(Z_j^* \mid \mathbb{A}_N)\right\}$ is positive for all N, where*

$$V(Z_j^* \mid \mathbb{A}_N) = \sum_{i_1, i_2 \in \mathbb{U}_\ell - \mathbb{A}_\ell} \omega_{i_1 j}\omega_{i_2 j}\left[G\left(\min[\dot{q}_{i_1 j}^*, \dot{q}_{i_2 j}^*]\right) - G\left(\dot{q}_{i_1 j}^*\right)G\left(\dot{q}_{i_2 j}^*\right)\right],$$

$Z_j^* = \sum_{i \in \mathbb{U}_\ell - \mathbb{A}_\ell} \omega_{ij} I\left(U_j \leqslant \dot{q}_{ij}^*\right), j \in \mathbb{A}_N$, $\dot{q}_{ij}^* = \dot{q}_i + h_i^{-1}(q_j^{-1}h_j - q_i^{-1}h_i)[\dot{y} - x_i\beta]$, ℓ is the bin that contains unit j, $\mathbb{U}_\ell$ is the set of indices in bin ℓ and $\mathbb{A}_\ell = \mathbb{A}_N \cap \mathbb{U}_\ell$.

A.7b *The distribution function $G(u)$ is differentiable and there exists an M_g such that $|\partial G(u)/\partial u| < M_g$ for all u.*

A.8b *The functions $q(\cdot)$ and $h(\cdot)$ are differentiable and there exists an $\dot{M}_{qh}$ such that $|\partial[q(x)^{-1}h(x)]/\partial x| < \dot{M}_{qh}$ for all x.*

A.9b *The $\max_\ell(b_\ell) = O(B_N^{-1})$, where b_ℓ is the length of bin ℓ.*

Let $\widetilde{F}_{L\beta}(\dot{y})$ be the estimator (3.1.4) with the true β used in (3.1.6). Then, *(a) The estimator* $\widetilde{F}_{L\beta}(\dot{y})$ *satisfies*

$$\lim_{N\to\infty} P\big(\big|\widetilde{F}_{L\beta}(\dot{y}) - F_N(\dot{y})\big| > \epsilon \bigm| \mathbb{A}_N\big) = 0$$

for all $\epsilon > 0$*, where* $F_N(\dot{y})$ *is defined in (2.1.2).*

(b) If the value of α *in A.2a is greater than* 0.5*, then the sequence*

$$\Big\{V\big(\widetilde{F}_{L\beta}(\dot{y}) - F_N(\dot{y}) \bigm| \mathbb{A}_N\big)\Big\}^{-1/2}\Big\{\widetilde{F}_{L\beta}(\dot{y}) - F_N(\dot{y})\Big\} \tag{3.2.20}$$

converges in distribution to a $N(0,1)$ *random variable, where*

$$\begin{aligned} &V\big(\widetilde{F}_{L\beta}(\dot{y}) - F_N(\dot{y}) \bigm| \mathbb{A}_N\big) = \\ &\qquad = N^{-2}\sum_{\ell=1}^{B}\sum_{j\in\mathbb{A}_\ell}\sum_{i_1,i_2\in\mathbb{U}_\ell-\mathbb{A}_\ell} \omega_{i_1j}\omega_{i_2j}\big[G\big(\min[\dot{q}^*_{i_1j},\dot{q}^*_{i_2j}]\big) - G\big(\dot{q}^*_{i_1j}\big)G\big(\dot{q}^*_{i_2j}\big)\big] + \\ &\qquad\qquad + N^{-2}\sum_{i\in\mathbb{A}_N^c} G(\dot{q}_i)[1-G(\dot{q}_i)]. \end{aligned}$$

Proof. Part *(a)*. Let

$$\widetilde{F}_{L\beta}(\dot{y}) - F_N(\dot{y}) = T_L - T_N, \tag{3.2.21}$$

where T_L and T_N are defined in (3.2.5) and (3.2.6) respectively. Under model (3.2.19),

$$\begin{aligned} E\big[I(Y_i \leqslant \dot{y})\big] &= E\big\{I\big(q_i^{-1}[Y_i - x_i\beta] \leqslant q_i^{-1}[\dot{y} - x_i\beta]\big)\big\} \\ &= E\big\{I\big(U_i \leqslant \dot{q}_i\big)\big\} \\ &= G\big(\dot{q}_i\big) \end{aligned} \tag{3.2.22}$$

where $\dot{q}_i = q_i^{-1}[\dot{y} - x_i\beta]$. The summands in T_L involve $I\bigl(h_j^{-1}[Y_j - x_j\beta] \leqslant h_i^{-1}[\dot{y} - x_i\beta]\bigr)$. All of the following inequalities are equivalent,

$$\begin{aligned} h_j^{-1}[Y_j - x_j\beta] &\leqslant h_i^{-1}[\dot{y} - x_i\beta] \\ q_j^{-1}[Y_j - x_j\beta] &\leqslant q_j^{-1}h_jh_i^{-1}[\dot{y} - x_i\beta] \\ U_j &\leqslant q_i^{-1}[\dot{y} - x_i\beta] + (q_j^{-1}h_jh_i^{-1} - q_i^{-1})[\dot{y} - x_i\beta] \\ U_j &\leqslant \dot{q}_i + h_i^{-1}(q_j^{-1}h_j - q_i^{-1}h_i)[\dot{y} - x_i\beta] \\ U_j &\leqslant \dot{q}_i + \dot{\delta}_{ij}, \end{aligned} \tag{3.2.23}$$

where

$$\dot{\delta}_{ij} = h_i^{-1}(q_j^{-1}h_j - q_i^{-1}h_i)[\dot{y} - x_i\beta]. \tag{3.2.24}$$

Then, under model (3.2.19),

$$\begin{aligned} E\bigl\{I\bigl(h_j^{-1}[Y_j - x_j\beta] \leqslant h_i^{-1}[\dot{y} - x_i\beta]\bigr)\bigr\} &= E\bigl\{I\bigl(U_j \leqslant \dot{q}_i + \dot{\delta}_{ij}\bigr)\bigr\} \\ &= G\bigl(\dot{q}_i + \dot{\delta}_{ij}\bigr). \end{aligned} \tag{3.2.25}$$

Using results (3.2.22) and (3.2.25) we can compute

$$E[T_L \mid \mathbb{A}_N] = N^{-1} \sum_{i \in \mathbb{A}^c} \sum_{j \in \mathbb{A}_{\ell_i}} \omega_{ij} G\bigl(\dot{q}_i + \dot{\delta}_{ij}\bigr), \tag{3.2.26}$$

and

$$\begin{aligned} E[T_N \mid \mathbb{A}_N] &= N^{-1} \sum_{i \in \mathbb{A}^c} G\bigl(\dot{q}_i\bigr) \\ &= N^{-1} \sum_{i \in \mathbb{A}^c} \sum_{j \in \mathbb{A}_{\ell_i}} w_{ij} G\bigl(\dot{q}_i\bigr), \end{aligned} \tag{3.2.27}$$

since $\sum_{j \in \mathbb{A}_{\ell_i}} w_{ij} = 1$ for all i. Then

$$E\bigl[\widetilde{F}_{L\beta}(\dot{y}) - F_N(\dot{y}) \mid \mathbb{A}_N\bigr] = N^{-1} \sum_{i \in \mathbb{A}^c} \sum_{j \in \mathbb{A}_{\ell_i}} \omega_{ij} \bigl\{G\bigl(\dot{q}_i + \dot{\delta}_{ij}\bigr) - G\bigl(\dot{q}_i\bigr)\bigr\}, \tag{3.2.28}$$

which will in general be different from zero.

We study the magnitude of the model bias of $\widetilde{F}_{L\beta}(\dot{y})$ as an estimator of $F_N(\dot{y})$. By the mean value theorem and A.7b,

$$\Big|G(\dot{q}_i + \dot{\delta}_{ij}) - G(\dot{q}_i)\Big| \leqslant |\dot{\delta}_{ij}| M_g.$$

Then,

$$\begin{aligned} \Big|E\big[\widetilde{F}_{L\beta}(\dot{y}) - F_N(\dot{y}) \mid \mathbb{A}_N\big]\Big| &= \Big|N^{-1} \sum_{i \in \mathbb{A}^c} \sum_{j \in \mathbb{A}_{\ell_i}} \omega_{ij} \{G(\dot{q}_i + \dot{\delta}_{ij}) - G(\dot{q}_i)\}\Big| \\ &\leqslant N^{-1} \sum_{i \in \mathbb{A}^c} \sum_{j \in \mathbb{A}_{\ell_i}} \omega_{ij} \Big|\{G(\dot{q}_i + \dot{\delta}_{ij}) - G(\dot{q}_i)\}\Big| \\ &\leqslant N^{-1} \sum_{i \in \mathbb{A}^c} \sum_{j \in \mathbb{A}_{\ell_i}} \omega_{ij} \Big|\dot{\delta}_{ij}\Big| M_g. \end{aligned} \qquad (3.2.29)$$

The quantities $\dot{\delta}_{ij}$ in (3.2.29) are

$$\dot{\delta}_{ij} = h_i^{-1}(q_j^{-1}h_j - q_i^{-1}h_i)[\dot{y} - x_i\beta]$$

as defined in (3.2.24). Under model (3.2.19),

$$h_i^{-1} = h(x_i)^{-1} < m_h^{-1} = O(1) \qquad (3.2.30)$$

for all $i \in \mathbb{U}_N$. By A.8b and applying the mean value theorem,

$$\Big|q_j^{-1}h_j - q_i^{-1}h_i\Big| < \Big|x_i - x_j\Big|\dot{M}_{qh}. \qquad (3.2.31)$$

If the units i and j of (3.2.31) belong to the same bin, then

$$\Big|x_i - x_j\Big|\dot{M}_{qh} \leqslant \big(\max_{l=1,\dots,B_N} b_\ell\big)\dot{M}_{qh} = O(B_N^{-1}) \qquad (3.2.32)$$

by A.9b. Finally,

$$\dot{y} - x_i\beta = O(1) \qquad (3.2.33)$$

because $\dot{y}$ and β are fixed, and $|x_i| < M_x$ under model (3.2.19). Combining (3.2.30), (3.2.32) and (3.2.33) we have that

$$\dot{\delta}_{ij} = O(B_N^{-1}). \tag{3.2.34}$$

Then, since $\sum_{j \in \mathbb{A}_{\ell_i}} \omega_{ij} = 1$,

$$\begin{aligned} \Big| E\big[\widetilde{F}_{L\beta}(\dot{y}) - F_N(\dot{y}) \mid \mathbb{A}_N\big] \Big| &\leqslant N^{-1} \sum_{i \in \mathbb{A}^c} \sum_{j \in \mathbb{A}_{\ell_i}} \omega_{ij} \left| \dot{\delta}_{ij} \right| M_g \\ &\leqslant N^{-1} \sum_{i \in \mathbb{A}^c} \sum_{j \in \mathbb{A}_{\ell_i}} \omega_{ij} O(B_N^{-1}) = O(B_N^{-1}) \end{aligned} \tag{3.2.35}$$

Then, by A.2a, the model bias of $\widetilde{F}_{L\beta}(\dot{y})$ as an estimator of $F_N(\dot{y})$ decreases at a rate $N^{-\alpha}$. Then, under model (3.2.19), the local-residuals estimator (3.1.4) is asymptotically model unbiased. Note that assumption A.9b is essential in proving (3.2.35). Asymptotic model unbiasedness of Chambers and Dunstan estimator (2.3.4) does not hold when the variance function of Y given x is misspecified since $B = 1$ is used in constructing $\widehat{F}_{CD}(\dot{y})$. With $B = 1$ the model bias in $\widehat{F}_{CD}(\dot{y})$ as an estimator of $F_N(\dot{y})$ does not converge to zero as $N \to \infty$.

The model variance of $\widetilde{F}_{L\beta}(\dot{y}) - F_N(\dot{y})$ is

$$V\big(\widetilde{F}_{L\beta}(\dot{y}) - F_N(\dot{y}) \mid \mathbb{A}_N\big) = V\big(T_L \mid \mathbb{A}_N\big) + V\big(T_N \mid \mathbb{A}_N\big) \tag{3.2.36}$$

because the $Y_i, i \in \mathbb{U}_N$ are independent given $\mathbb{A}_N$. Under model (3.2.19), the indicator functions $I\Big(Y_i \leqslant \dot{y}\Big)$ that appear in T_N have

$$\begin{aligned} V\Big[I\Big(Y_i \leqslant \dot{y}\Big)\Big] &= V\Big[I\Big(q_i^{-1}[Y_i - x_i\beta] \leqslant q_i^{-1}[\dot{y} - x_i\beta]\Big)\Big] \\ &= V\Big[I\Big(U_I \leqslant \dot{q}_i\Big)\Big] \\ &= G(\dot{q}_i)[1 - G(\dot{q}_i)] \end{aligned}$$

Then,

$$\begin{aligned} V\Big(T_N \mid \mathbb{A}_N\Big) &= V\Big(N^{-1}\sum_{i\in\mathbb{A}_N^c} I(Y_i \leqslant \dot{y}) \mid \mathbb{A}_N\Big) \\ &= N^{-2}\sum_{i\in\mathbb{A}_N^c} V\Big(I(Y_i \leqslant \dot{y}) \mid \mathbb{A}_N\Big) \\ &= N^{-2}\sum_{i\in\mathbb{A}_N^c} G(\dot{q}_i)[1-G(\dot{q}_i)]. \qquad (3.2.37) \end{aligned}$$

By (3.2.23) we have that,

$$\begin{aligned} V\Big(T_L \mid \mathbb{A}_N\Big) &= V\Big(N^{-1}\sum_{\ell=1}^{B}\sum_{j\in\mathbb{A}_\ell}\sum_{i\in\mathbb{U}_\ell-\mathbb{A}_\ell} \omega_{ij} I\big(h_j^{-1}[Y_j - x_j\beta] \leqslant h_i^{-1}[\dot{y} - x_i\beta]\big) \mid \mathbb{A}_N\Big) \\ &= V\Big(N^{-1}\sum_{\ell=1}^{B}\sum_{j\in\mathbb{A}_\ell}\sum_{i\in\mathbb{U}_\ell-\mathbb{A}_\ell} \omega_{ij} I\big(U_j \leqslant \dot{q}_i + \dot{\delta}_{ij}\big) \mid \mathbb{A}_N\Big) \\ &= V\Big(N^{-1}\sum_{\ell=1}^{B}\sum_{j\in\mathbb{A}_\ell}\sum_{i\in\mathbb{U}_\ell-\mathbb{A}_\ell} \omega_{ij} I\big(U_j \leqslant \dot{q}_{ij}^*\big) \mid \mathbb{A}_N\Big), \end{aligned}$$

where $\dot{q}_{ij}^* = \dot{q}_i + \dot{\delta}_{ij}$ is defined in A.6b. Then, since under model (3.2.19) the U_j are independent,

$$\begin{aligned} V\Big(T_L \mid \mathbb{A}_N\Big) &= V\Big(N^{-1}\sum_{\ell=1}^{B}\sum_{j\in\mathbb{A}_\ell}\sum_{i\in\mathbb{U}_\ell-\mathbb{A}_\ell} \omega_{ij} I\big(U_j \leqslant \dot{q}_{ij}^*\big) \mid \mathbb{A}_N\Big) \\ &= N^{-2}\sum_{\ell=1}^{B}\sum_{j\in\mathbb{A}_\ell} V\Big(\sum_{i\in\mathbb{U}_\ell-\mathbb{A}_\ell} \omega_{ij} I\big(U_j \leqslant \dot{q}_{ij}^*\big) \mid \mathbb{A}_N\Big) \\ &= N^{-2}\sum_{\ell=1}^{B}\sum_{j\in\mathbb{A}_\ell}\sum_{i_1,i_2\in\mathbb{U}_\ell-\mathbb{A}_\ell} \omega_{i_1 j}\omega_{i_2 j}\mathrm{Cov}\big[I\big(U_j \leqslant \dot{q}_{i_1 j}^*\big), I\big(U_j \leqslant \dot{q}_{i_2 j}^*\big) \mid \mathbb{A}_N\big] \\ &= N^{-2}\sum_{\ell=1}^{B}\sum_{j\in\mathbb{A}_\ell}\sum_{i_1,i_2\in\mathbb{U}_\ell-\mathbb{A}_\ell} \omega_{i_1 j}\omega_{i_2 j}\big[G\big(\min[\dot{q}_{i_1 j}^*, \dot{q}_{i_2 j}^*]\big) - G\big(\dot{q}_{i_1 j}^*\big)G\big(\dot{q}_{i_2 j}^*\big)\big]. \\ & \qquad (3.2.38) \end{aligned}$$

To study the asymptotic properties of (3.2.37) and (3.2.38), recall that the variances $G(\dot{q}_i)[1-G(\dot{q}_i)]$ that appear in (3.2.37) are bounded by

$$G(\dot{q}_i)[1-G(\dot{q}_i)] \leqslant 0.25$$

and that the covariances $G\big(\min[\dot{q}^*_{i_1j}, \dot{q}^*_{i_2j}]\big) - G\big(\dot{q}^*_{i_1j}\big)G\big(\dot{q}^*_{i_2j}\big)$ that appear in (3.2.38) are bounded, in absolute value, by

$$\left|G\big(\min[\dot{q}^*_{i_1j}, \dot{q}^*_{i_2j}]\big) - G\big(\dot{q}^*_{i_1j}\big)G\big(\dot{q}^*_{i_2j}\big)\right| \leqslant 0.25.$$

Then,

$$V\Big(T_N \mid \mathbb{A}_N\Big) \leqslant 4^{-1}N^{-1} = O(N^{-1}) \tag{3.2.39}$$

and

$$V\Big(T_L \mid \mathbb{A}_N\Big) \leqslant 4^{-1}N^{-2}\sum_{\ell=1}^{B}\sum_{j\in\mathbb{A}_\ell}\Big(\sum_{i\in\mathbb{U}_\ell-\mathbb{A}_\ell}\omega_{ij}\Big)^2$$

as in (3.2.16). Then, by the same argument that is used in (3.2.17),

$$V\Big(T_L \mid \mathbb{A}_N\Big) = O(N^{-1}). \tag{3.2.40}$$

Combining (3.2.40) and (3.2.39) we have that

$$V\Big\{\widetilde{F}_{L\beta}(\dot{y}) - F_N(\dot{y}) \mid \mathbb{A}_N\Big\} = O(N^{-1}). \tag{3.2.41}$$

By result (3.2.35), $\widetilde{F}_{L\beta}(\dot{y})$ is asymptotically model unbiased for $F_N(\dot{y})$, and, by (3.2.41), the model variance of $\widetilde{F}_{L\beta}(\dot{y}) - F_N(\dot{y})$ converges to zero as $N \to \infty$. Thus, we have that the local-residuals estimator is model consistent for the finite population distribution function under model (3.2.19).

Part *(b)*. We will use an argument similar to the one used in proving part *(b)* of Theorem 3.2.2. The terms T_L and T_N that appear in (3.2.21) are independent given $\mathbb{A}_N$. By assumption A.5b,

$$\begin{aligned}\Big\{V(T_N \mid \mathbb{A}_N)\Big\}^{-1/2}&\Big\{T_N - E(T_N \mid \mathbb{A}_N)\Big\} = \\ &= \Big\{N^{-2}\sum_{i\in\mathbb{A}_N^c} G(\dot{u}_i)[1-G(\dot{u}_i)]\Big\}^{-1/2}\Big\{N^{-1}\sum_{i\in\mathbb{A}_N^c}\big[I(U_i \leqslant \dot{u}_i) - G(\dot{u}_i)\big]\Big\} \\ &= \Big\{\sum_{i\in\mathbb{A}_N^c} G(\dot{u}_i)[1-G(\dot{u}_i)]\Big\}^{-1/2}\Big\{\sum_{i\in\mathbb{A}_N^c}\big[I(U_i \leqslant \dot{u}_i) - G(\dot{u}_i)\big]\Big\}\end{aligned} \tag{3.2.42}$$

converges in distribution to a standard normal. The term T_L can be written as

$$\begin{aligned} T_L &= N^{-1} \sum_{i \in \mathbb{A}_N^c} \sum_{j \in \mathbb{A}_{\ell_i}} \omega_{ij} I\Big(h_j^{-1}[Y_j - x_j\beta] \leqslant h_i^{-1}[\dot{y} - x_i\beta]\Big) \\ &= N^{-1} \sum_{\ell=1}^{B} \sum_{j \in \mathbb{A}_\ell} \sum_{i \in \mathbb{U}_\ell - \mathbb{A}_\ell} \omega_{ij} I\Big(U_j \leqslant \dot{q}_{ij}^*\Big) \end{aligned}$$

by (3.2.23), where $\dot{q}_{ij}^* = \dot{q}_i + \dot{\delta}_{ij}$ is defined in A.6b. Using the Z_j^* defined in A.6b, $Z_j^* = \sum_{i \in \mathbb{U}_\ell - \mathbb{A}_\ell} \omega_{ij} I\big(U_j \leqslant \dot{q}_{ij}^*\big), j \in \mathbb{A}_N$, we have that

$$\begin{aligned} T_L &= N^{-1} \sum_{\ell=1}^{B} \sum_{j \in \mathbb{A}_\ell} Z_j^* \\ &= N^{-1} \sum_{j \in \mathbb{A}_N} Z_j^*. \end{aligned}$$

Then, by A.6b,

$$\begin{aligned} &\Big\{V(T_L \mid \mathbb{A}_N)\Big\}^{-1/2}\Big\{T_L - E(T_L \mid \mathbb{A}_N)\Big\} = \\ &\qquad = \Big\{N^{-2} \sum_{j \in \mathbb{A}_N} V(Z_j^* \mid \mathbb{A}_N)\Big\}^{-1/2}\Big\{N^{-1} \sum_{j \in \mathbb{A}_N} Z_j^* - N^{-1}E(Z_j^* \mid \mathbb{A}_N)\Big\} \\ &\qquad = \Big\{\sum_{j \in \mathbb{A}_N} V(Z_j^* \mid \mathbb{A}_N)\Big\}^{-1/2}\Big\{\sum_{j \in \mathbb{A}_N} Z_j^* - E(Z_j^* \mid \mathbb{A}_N)\Big\} \end{aligned}$$

converges in distribution to a standard normal. It follows that

$$\Big\{V\big(\widetilde{F}_{L\beta}(\dot{y}) - F_N(\dot{y}) \mid \mathbb{A}_N\big)\Big\}^{-1/2}\Big\{\widetilde{F}_{L\beta}(\dot{y}) - E\big(\widetilde{F}_{L\beta}(\dot{y}) \mid \mathbb{A}_N\big)\Big\} \tag{3.2.43}$$

converges in law to a $N(0,1)$. To find the limiting distribution of

$$\Big\{V\big(\widetilde{F}_{L\beta}(\dot{y}) - F_N(\dot{y}) \mid \mathbb{A}_N\big)\Big\}^{-1/2}\Big\{\widetilde{F}_{L\beta}(\dot{y}) - F_N(\dot{y})\Big\}$$

we consider the fact that the model expectation of $\widetilde{F}_{L\beta}(\dot{y}) - F_N(\dot{y})$ is $O(B_N^{-1})$, while the model variance is $O(N^{-1})$. Since by assumption $\alpha > 0.5$,

$$\begin{aligned} E\Big[\Big\{V\big(\widetilde{F}_{L\beta}(\dot{y}) - F_N(\dot{y}) \mid \mathbb{A}_N\big)\Big\}^{-1/2}\Big\{\widetilde{F}_{L\beta}(\dot{y}) - F_N(\dot{y})\Big\} \mid \mathbb{A}_N\Big] &= O(N^{1/2-\alpha}) \\ &= o(1). \end{aligned} \tag{3.2.44}$$

Therefore, we can replace $E\big(\widetilde{F}_{L\beta}(\dot{y}) \mid \mathbb{A}_N\big)$ in (3.2.43) with $F_N(\dot{y})$ to obtain the limiting distribution of (3.2.20). ▲

3.2.3 Case C: $Y_i \sim G_Y(\dot{y}; x_i)$; β known

In Section 3.2.2, we proved that the local-residuals estimator is robust against misspecification of the variance function. In this section we study the case when both mean and variance of Y are misspecified. The superpopulation model that describes the relation between Y and x is such that $E[Y_i \,|x_i]$ and $V(Y_i \,|x_i)$ are not restricted to be $x_i\beta$ and $h(x_i)^2\sigma^2$ respectively. The residuals $\big[V(Y_i \,|x_i)\big]^{-1/2}\big[Y_i - E(Y_i \,|x_i)\big]$ are independent, but the residuals are no longer restricted to be identically distributed. The superpopulation model (3.2.45) in Theorem 3.2.4 is specified in terms of the distribution of Y given x.

Theorem 3.2.4 *Let $\{\mathbb{A}_N\}$ be a sequence of samples selected from the sequence of finite populations $\{\mathbb{U}_N\}$. Assume that the sample $\mathbb{A}_N$ is divided into B_N groups, each of size k_N, as described in (3.1.3). Assume a superpopulation model where the Y_i are independent and*

$$P(Y_i \leqslant y \mid x_i) = G_Y(y; x_i) \tag{3.2.45}$$

for $i \in \mathbb{U}_N$. Let $h_i = h(x_i)$, where $h(\cdot)$ is the function used in constructing estimator (3.1.4). Assume that there exists an m_h such that $0 < m_h \leqslant h(x_i) < \infty$ for $i \in \mathbb{U}_N$. Let $\dot{y}$ be a fixed point. Assume A.1a *through* A.4a *from Theorem 3.2.2 and*

A.5c *For all N, $\Big\{(N - n_N)^{-1} \sum_{i \in \mathbb{A}_N^c} G_Y(\dot{y}; x_i)[1 - G_Y(\dot{y}; x_i)]\Big\}$ is positive.*

A.6c *For all N, $\Big\{N^{-1} \sum_{j \in \mathbb{A}_N} V(\ddot{Z}_j \mid \mathbb{A}_N)\Big\}$ is positive, where*

$$V(\ddot{Z}_j \,|\mathbb{A}_N) = \sum_{i_1, i_2 \in \mathbb{U}_\ell - \mathbb{A}_\ell} \omega_{i_1 j}\omega_{i_2 j}\big[G_Y\big(\min[\ddot{y}_{i_1 j}, \ddot{y}_{i_2 j}]; x_j\big) - G_Y\big(\ddot{y}_{i_1 j}; x_j\big) G_Y\big(\ddot{y}_{i_2 j}; x_j\big)\big],$$

$$\ddot{Z}_j = \sum_{i \in \mathbb{U}_\ell - \mathbb{A}_\ell} \omega_{ij} I\big(Y_j \leqslant \ddot{y}_{ij}\big), j \in \mathbb{A}_N,$$

$$\ddot{y}_{ij} = \dot{y} + h_j(h_i^{-1} - h_j^{-1})\dot{y} + h_j(h_j^{-1}x_j - h_i^{-1}x_i)\beta,$$

ℓ is the bin that contains unit j, $\mathbb{U}_\ell$ is the set of indices in bin ℓ and $\mathbb{A}_\ell = \mathbb{A}_N \cap \mathbb{U}_\ell$.

A.7c *The distribution function $G_Y(y;x)$ is differentiable in y and x, and there exists an $\dot{M}_{gg}$ such that $|\partial G_Y(y;x)/\partial y| < \dot{M}_{gg}$ and $|\partial G_Y(y;x)/\partial x| < \dot{M}_{gg}$ for all y and x.*

A.8c *The positive function $h(x)$ is differentiable and there exists an $\dot{M}_{hx}$ such that*

$$|h_j(h_i^{-1} - h_j^{-1})\dot{y} + h_j(h_j^{-1}x_j - h_i^{-1}x_i)\beta| \leqslant |x_i - x_j|\dot{M}_{hx}$$

for all x.

A.9c *The $\max_\ell(b_\ell) = O(B_N^{-1})$, where b_ℓ is the length of bin ℓ.*

Let $\widetilde{F}_{L\beta}(\dot{y})$ be the estimator (3.1.4) with the true β used in (3.1.6). Then,

(a) The estimator $\widetilde{F}_{L\beta}(\dot{y})$ satisfies

$$\lim_{N\to\infty} P\big(\big|\widetilde{F}_{L\beta}(\dot{y}) - F_N(\dot{y})\big| > \epsilon \bigm| \mathbb{A}_N\big) = 0$$

for all $\epsilon > 0$, where $F_N(\dot{y})$ is defined in (2.1.2).

(b) If the value of α in A.2a is greater than 0.5, then the sequence

$$\Big\{V\big(\widetilde{F}_{L\beta}(\dot{y}) - F_N(\dot{y}) \bigm| \mathbb{A}_N\big)\Big\}^{-1/2}\Big\{\widetilde{F}_{L\beta}(\dot{y}) - F_N(\dot{y})\Big\} \qquad (3.2.46)$$

converges in distribution to a $N(0,1)$ random variable, where

$$\begin{aligned} V\big(\widetilde{F}_{L\beta}(\dot{y}) - F_N(\dot{y}) \bigm| \mathbb{A}_N\big) = N^{-2}\sum_{\ell=1}^{B}\sum_{j\in\mathbb{A}_\ell}\sum_{i_1,i_2\in\mathbb{U}_\ell-\mathbb{A}_\ell} \omega_{i_1 j}\omega_{i_2 j} \times \\ \times \big[G_Y\big(\min[\ddot{y}_{i_1 j}, \ddot{y}_{i_2 j}]; x_j\big) - G_Y\big(\ddot{y}_{i_1 j}; x_j\big)G_Y\big(\ddot{y}_{i_2 j}; x_j\big)\big] + \\ + N^{-2}\sum_{i\in\mathbb{A}_N^c} G_Y(\dot{y}; x_i)[1 - G_Y(\dot{y}; x_i)]. \end{aligned}$$

Proof. Part *(a)*. Let

$$\widetilde{F}_{L\beta}(\dot{y}) - F_N(\dot{y}) = T_L - T_N, \tag{3.2.47}$$

where T_L and T_N are defined in (3.2.5) and (3.2.6) respectively. We will compute the expected value under model (3.2.45) of the indicator functions $I\Big(h_j^{-1}[Y_j - x_j\beta] \leqslant h_i^{-1}[\dot{y} - x_i\beta]\Big)$ and $I\Big(Y_i \leqslant \dot{y}\Big)$ that appear in the definitions of T_L and T_N respectively. The following inequalities are equivalent

$$\begin{aligned} h_j^{-1}[Y_j - x_j\beta] \leqslant h_i^{-1}[\dot{y} - x_i\beta] & \\ Y_j \leqslant x_j\beta + h_j h_i^{-1}[\dot{y} - x_i\beta] & \\ Y_j \leqslant \dot{y} + h_j(h_i^{-1} - h_j^{-1})\dot{y} + h_j(h_j^{-1}x_j - h_i^{-1}x_i)\beta & \\ Y_j \leqslant \dot{y} + \ddot{\delta}_{ij} & \qquad (3.2.48) \\ Y_j \leqslant \ddot{y}_{ij}, & \qquad (3.2.49) \end{aligned}$$

where $\ddot{\delta}_{ij}$ in (3.2.48) is

$$\ddot{\delta}_{ij} = h_j(h_i^{-1} - h_j^{-1})\dot{y} + h_j(h_j^{-1}x_j - h_i^{-1}x_i)\beta \tag{3.2.50}$$

and $\ddot{y}_{ij} = \dot{y} + \ddot{\delta}_{ij}$ is defined in A.6c. Then, under model (3.2.45) we have that

$$\begin{aligned} E\big\{I\big(h_j^{-1}[Y_j - x_j\beta] \leqslant h_i^{-1}[\dot{y} - x_i\beta]\big)\big\} &= E\big\{I\big(Y_j \leqslant \dot{y} + \ddot{\delta}_{ij}\big)\big\} \\ &= G_Y(\dot{y} + \ddot{\delta}_{ij}; x_j) \qquad (3.2.51) \end{aligned}$$

and

$$E\big[I(Y_i \leqslant \dot{y})\big] = G_Y(\dot{y}; x_i). \tag{3.2.52}$$

Using (3.2.51) and (3.2.52) we have that

$$E[T_L \mid \mathbb{A}_N] = N^{-1} \sum_{i \in \mathbb{A}_N^c} \sum_{j \in \mathbb{A}_{\ell_i}} \omega_{ij} G_Y\big(\dot{y} + \ddot{\delta}_{ij}; x_j\big), \tag{3.2.53}$$

and

$$
\begin{aligned}
E[T_N \mid \mathbb{A}_N] &= N^{-1} \sum_{i \in \mathbb{A}_N^c} G_Y(\dot{y}; x_i) \\
&= N^{-1} \sum_{i \in \mathbb{A}_N^c} \sum_{j \in \mathbb{A}_{\ell_i}} \omega_{ij} G_Y(\dot{y}; x_i), \qquad (3.2.54)
\end{aligned}
$$

because $\sum_{j \in \mathbb{A}_{\ell_i}} \omega_{ij} = 1$ for all $i \in \mathbb{A}_N^c$. Then, the model bias of $\widetilde{F}_{L\beta}(\dot{y})$ as an estimator of $F_N(\dot{y})$ is

$$
E\big[\widetilde{F}_{L\beta}(\dot{y}) - F_N(\dot{y}) \mid \mathbb{A}_N\big] = N^{-1} \sum_{i \in \mathbb{A}^c} \sum_{j \in \mathbb{A}_{\ell_i}} \omega_{ij} \big\{ G_Y(\dot{y} + \ddot{\delta}_{ij}; x_j) - G_Y(\dot{y}; x_i) \big\} \qquad (3.2.55)
$$

which in general will be different from zero. The difference between distribution functions in (3.2.55) can be written as

$$
G_Y(\dot{y} + \ddot{\delta}_{ij}; x_j) - G_Y(\dot{y}; x_i) = \big[G_Y(\dot{y} + \ddot{\delta}_{ij}; x_j) - G_Y(\dot{y}; x_j)\big] + \big[G_Y(\dot{y}; x_j) - G_Y(\dot{y}; x_i)\big].
$$

Then, by the triangular inequality,

$$
\Big|G_Y(\dot{y} + \ddot{\delta}_{ij}; x_j) - G_Y(\dot{y}; x_i)\Big| \leqslant \Big|G_Y(\dot{y} + \ddot{\delta}_{ij}; x_j) - G_Y(\dot{y}; x_j)\Big| + \Big|G_Y(\dot{y}; x_j) - G_Y(\dot{y}; x_i)\Big|. \qquad (3.2.56)
$$

Applying mean value theorem and A.7c on the right hand side of (3.2.56) we have that

$$
\begin{aligned}
\Big|G_Y(\dot{y} + \ddot{\delta}_{ij}; x_j) - G_Y(\dot{y}; x_i)\Big| &\leqslant \big(|\ddot{\delta}_{ij}| + |x_j - x_i|\big)\dot{M}_{gg} \\
&\leqslant |x_j - x_i|(\dot{M}_{hx} + 1)\dot{M}_{gg}
\end{aligned}
$$

by A.8c. Since units i and j are in the same bin, $|x_j - x_i| \leqslant \max_\ell(b_\ell)$, then

$$
\begin{aligned}
\Big|G_Y(\dot{y} + \ddot{\delta}_{ij}; x_j) - G_Y(\dot{y}; x_i)\Big| &\leqslant \max_\ell(b_\ell)(\dot{M}_{hx} + 1)\dot{M}_{gg} \\
&= O(B_N^{-1}) \qquad (3.2.57)
\end{aligned}
$$

by A.9c. The model bias of the local-residuals estimator is then

$$\begin{aligned}\Big|E\big[\widetilde{F}_{L\beta}(\dot{y}) - F_N(\dot{y}) \mid \mathbb{A}_N\big]\Big| &\leqslant N^{-1}\sum_{i\in\mathbb{A}^c}\sum_{j\in\mathbb{A}_{\ell_i}} \omega_{ij}\Big|G_Y(\dot{y}+\ddot{\delta}_{ij};x_j) - G_Y(\dot{y};x_i)\Big| \\ &= N^{-1}\sum_{i\in\mathbb{A}^c}\sum_{j\in\mathbb{A}_{\ell_i}} \omega_{ij}O(B_N^{-1}) \\ &= O(B_N^{-1}). \qquad (3.2.58)\end{aligned}$$

Since by A.2a we have that $B \to \infty$ as $N \to \infty$, we have that $\widetilde{F}_{L\beta}(\dot{y})$ is asymptotically model unbiased for $F_N(\dot{y})$. Note that (3.2.58) indicates that the model unbiasedness holds even when both the mean function and the variance function of Y given x are misspecified.

The variance of $\widetilde{F}_{L\beta}(\dot{y}) - F_N(\dot{y})$ given $\mathbb{A}_N$ is

$$V\big(\widetilde{F}_{L\beta}(\dot{y}) - F_N(\dot{y}) \mid \mathbb{A}_N\big) = V\big(T_L \mid \mathbb{A}_N\big) + V\big(T_N \mid \mathbb{A}_N\big) \qquad (3.2.59)$$

because the $Y_i, i \in \mathbb{U}_N$ are independent under model (3.2.45). The model variance of T_N is

$$\begin{aligned}V\big(T_N \mid \mathbb{A}_N\big) &= V\big(N^{-1}\sum_{i\in\mathbb{A}_N^c} I(Y_i \leqslant \dot{y}) \mid \mathbb{A}_N\big) \\ &= N^{-2}\sum_{i\in\mathbb{A}_N^c} G_Y(\dot{y};x_i)[1 - G_Y(\dot{y};x_i)] \qquad (3.2.60)\end{aligned}$$

Using (3.2.49), the model variance of T_L is

$$\begin{aligned}V\Big(T_L \mid \mathbb{A}_N\Big) &= V\Big(N^{-1}\sum_{\ell=1}^{B}\sum_{j\in\mathbb{A}_\ell}\sum_{i\in\mathbb{U}_\ell-\mathbb{A}_\ell} \omega_{ij}I\big(h_j^{-1}[Y_j - x_j\beta] \leqslant h_i^{-1}[\dot{y} - x_i\beta]\big) \mid \mathbb{A}_N\Big) \\ &= V\Big(N^{-1}\sum_{\ell=1}^{B}\sum_{j\in\mathbb{A}_\ell}\sum_{i\in\mathbb{U}_\ell-\mathbb{A}_\ell} \omega_{ij}I\big(Y_j \leqslant \ddot{y}_{ij}\big) \mid \mathbb{A}_N\Big),\end{aligned}$$

where $\ddot{y}_{ij}$ is defined in A.6c. Then, since under model (3.2.45) the Y_j are independent,

$$\begin{aligned} V\Big(T_L \mid \mathbb{A}_N\Big) &= V\Big(N^{-1}\sum_{\ell=1}^{B}\sum_{j\in\mathbb{A}_\ell}\sum_{i\in\mathbb{U}_\ell-\mathbb{A}_\ell} \omega_{ij} I\big(Y_j \leqslant \ddot{y}_{ij}\big) \mid \mathbb{A}_N\Big) \\ &= N^{-2}\sum_{\ell=1}^{B}\sum_{j\in\mathbb{A}_\ell} V\Big(\sum_{i\in\mathbb{U}_\ell-\mathbb{A}_\ell} \omega_{ij} I\big(Y_j \leqslant \ddot{y}_{ij}\big) \mid \mathbb{A}_N\Big) \\ &= N^{-2}\sum_{\ell=1}^{B}\sum_{j\in\mathbb{A}_\ell}\sum_{i_1,i_2\in\mathbb{U}_\ell-\mathbb{A}_\ell} \omega_{i_1 j}\omega_{i_2 j}\mathrm{Cov}\big[I\big(Y_j \leqslant \ddot{y}_{i_1 j}\big), I\big(Y_j \leqslant \ddot{y}_{i_2 j}\big) \mid \mathbb{A}_N\big] \\ &= N^{-2}\sum_{\ell=1}^{B}\sum_{j\in\mathbb{A}_\ell}\sum_{i_1,i_2\in\mathbb{U}_\ell-\mathbb{A}_\ell} \omega_{i_1 j}\omega_{i_2 j} \times \\ &\qquad \times \big[G_Y\big(\min[\ddot{y}_{i_1 j}, \ddot{y}_{i_2 j}]; x_j\big) - G_Y\big(\ddot{y}_{i_1 j}; x_j\big)G_Y\big(\ddot{y}_{i_2 j}; x_j\big)\big]. \end{aligned} \tag{3.2.61}$$

As in Theorem 3.2.3, to study the asymptotic properties of (3.2.60) and (3.2.61), recall that the variances $G_Y(\dot{y}; x_i)[1 - G_Y(\dot{y}; x_i)]$ that appear in (3.2.60) are bounded by

$$G_Y(\dot{y}; x_i)[1 - G_Y(\dot{y}; x_i)] \leqslant 4^{-1}$$

and that the covariances $G_Y\big(\min[\ddot{y}_{i_1 j}, \ddot{y}_{i_2 j}]; x_j\big) - G_Y\big(\ddot{y}_{i_1 j}; x_j\big)G_Y\big(\ddot{y}_{i_2 j}; x_j\big)$ that appear in (3.2.61) are bounded, in absolute value, by

$$\Big|G_Y\big(\min[\ddot{y}_{i_1 j}, \ddot{y}_{i_2 j}]; x_j\big) - G_Y\big(\ddot{y}_{i_1 j}; x_j\big)G_Y\big(\ddot{y}_{i_2 j}; x_j\big)\Big| \leqslant 4^{-1}.$$

Then,

$$V\Big(T_N \mid \mathbb{A}_N\Big) \leqslant 4^{-1}N^{-1} = O(N^{-1}) \tag{3.2.62}$$

and

$$V\Big(T_L \mid \mathbb{A}_N\Big) \leqslant 4^{-1}N^{-2}\sum_{\ell=1}^{B}\sum_{j\in\mathbb{A}_\ell}\Big(\sum_{i\in\mathbb{U}_\ell-\mathbb{A}_\ell}\omega_{ij}\Big)^2$$

as in (3.2.16). Then, by the same argument that is used in (3.2.17),

$$V\Big(T_L \mid \mathbb{A}_N\Big) = O(N^{-1}). \tag{3.2.63}$$

Combining (3.2.63) and (3.2.62) we have that

$$V\Big\{\widetilde{F}_{L\beta}(\dot{y}) - F_N(\dot{y}) \mid \mathbb{A}_N\Big\} = O(N^{-1}). \tag{3.2.64}$$

By result (3.2.58), $\widetilde{F}_{L\beta}(\dot{y})$ is asymptotically model unbiased for $F_N(\dot{y})$, and, by (3.2.64), the model variance of $\widetilde{F}_{L\beta}(\dot{y}) - F_N(\dot{y})$ converges to zero as $N \to \infty$. Then, we have that the local-residuals estimator is model consistent for the finite population distribution function under model (3.2.45).

Part *(b)*. We will use an argument similar to the one used in proving parts *(b)* of Theorem 3.2.2 and Theorem 3.2.3. The terms T_L and T_N that appear in (3.2.47) are independent given $\mathbb{A}_N$. By assumption A.5c,

$$\begin{aligned}\Big\{V(T_N \mid \mathbb{A}_N)\Big\}^{-1/2}\Big\{T_N - E(T_N \mid \mathbb{A}_N)\Big\} &= \\ &= \Big\{N^{-2}\sum_{i\in\mathbb{A}_N^c} G_Y(\dot{y};x_i)[1-G_Y(\dot{y};x_i)]\Big\}^{-1/2} \times \\ &\qquad \times \Big\{N^{-1}\sum_{i\in\mathbb{A}_N^c}\big[I(Y_i \leqslant \dot{y};x_i) - G_Y(\dot{y};x_i)\big]\Big\} \\ &= \Big\{\sum_{i\in\mathbb{A}_N^c} G_Y(\dot{y};x_i)[1-G_Y(\dot{y};x_i)]\Big\}^{-1/2} \times \\ &\qquad \times \Big\{\sum_{i\in\mathbb{A}_N^c}\big[I(Y_i \leqslant \dot{y};x_i) - G_Y(\dot{y};x_i)\big]\Big\}\end{aligned} \tag{3.2.65}$$

converges in distribution to a standard normal. The term T_L can be written as

$$\begin{aligned}T_L &= N^{-1}\sum_{i\in\mathbb{A}_N^c}\sum_{j\in\mathbb{A}_{\ell_i}} \omega_{ij} I\Big(h_j^{-1}[Y_j - x_j\beta] \leqslant h_i^{-1}[\dot{y} - x_i\beta]\Big) \\ &= N^{-1}\sum_{\ell=1}^{B}\sum_{j\in\mathbb{A}_\ell}\sum_{i\in\mathbb{U}_\ell - \mathbb{A}_\ell} \omega_{ij} I\Big(Y_j \leqslant \ddot{y}_{ij}\Big)\end{aligned}$$

by (3.2.49). Using the $\ddot{Z}_j$ defined in A.6c, $\ddot{Z}_j = \sum_{i\in\mathbb{U}_\ell-\mathbb{A}_\ell}\omega_{ij}I\big(Y_j \leqslant \ddot{y}_{ij}\big), j \in \mathbb{A}_N$, we have that

$$\begin{aligned} T_L &= N^{-1}\sum_{\ell=1}^{B}\sum_{j\in\mathbb{A}_\ell}\ddot{Z}_j \\ &= N^{-1}\sum_{j\in\mathbb{A}_N}\ddot{Z}_j. \end{aligned}$$

Then, by A.6c,

$$\begin{aligned} &\Big\{V(T_L \mid \mathbb{A}_N)\Big\}^{-1/2}\Big\{T_L - E(T_L \mid \mathbb{A}_N)\Big\} = \\ &\qquad = \Big\{N^{-2}\sum_{j\in\mathbb{A}_N}V(\ddot{Z}_j \mid \mathbb{A}_N)\Big\}^{-1/2}\Big\{N^{-1}\sum_{j\in\mathbb{A}_N}\ddot{Z}_j - N^{-1}E(\ddot{Z}_j \mid \mathbb{A}_N)\Big\} \\ &\qquad = \Big\{\sum_{j\in\mathbb{A}_N}V(\ddot{Z}_j \mid \mathbb{A}_N)\Big\}^{-1/2}\Big\{\sum_{j\in\mathbb{A}_N}\ddot{Z}_j - E(\ddot{Z}_j \mid \mathbb{A}_N)\Big\} \end{aligned}$$

converges in distribution to a standard normal.

To find the limiting distribution of

$$\Big\{V\big(\widetilde{F}_{L\beta}(\dot{y}) - F_N(\dot{y}) \mid \mathbb{A}_N\big)\Big\}^{-1/2}\Big\{\widetilde{F}_{L\beta}(\dot{y}) - F_N(\dot{y})\Big\}$$

recall that the model expectation of $\widetilde{F}_{L\beta}(\dot{y}) - F_N(\dot{y})$ is $O(B_N^{-1})$, while the model variance is $O(N^{-1})$. By A.2a, $B_N = O(N^{\alpha})$, and by the assumption that $\alpha > 0.5$,

$$\begin{aligned} E\Big[\Big\{V\big(\widetilde{F}_{L\beta}(\dot{y}) - F_N(\dot{y}) \mid \mathbb{A}_N\big)\Big\}^{-1/2}\Big\{\widetilde{F}_{L\beta}(\dot{y}) - F_N(\dot{y})\Big\} \mid \mathbb{A}_N\Big] &= O(N^{1/2-\alpha}) \\ &= o(1), \end{aligned} \tag{3.2.66}$$

and the distribution result follows. ▲

Part *(a)* of Theorem 3.2.4 shows that the local-residuals estimator of the finite population distribution function is model consistent even in the case when both the conditional mean and the conditional variance of Y given x are misspecified. The assumption that the maximum of the bin lengths goes to zero as N increases is crucial for the local-residuals estimator to be model consistent under misspecification of the conditional mean and the conditional variance of Y given x. We need $\alpha > 0.5$ for the bias in the sum to go to zero faster than the standard error as $N \to \infty$. Although, $\alpha > 0.5$ is not necessary for the model consistency of $\widetilde{F}_{L\beta}(\dot{y}) - F_N(\dot{y})$. However, if α is close to 0.5 the rate of decrease in the bias is small.

We will construct a model consistent estimator of the variance of $\widehat{F}_L(\dot{y}) - F_N(\dot{y})$ based on the local-residuals estimator. Let

$$\widetilde{G}_Y(\check{y}; x_i) = \sum_{j \in \mathbb{A}_{\ell_i}} \omega_{ij} I\big(\widetilde{Y}_{ij} \leqslant \check{y}\big), \tag{3.2.67}$$

be the local estimator of the conditional distribution function of Y for $x = x_i$, evaluated at the point $\check{y}$, where

$$\widetilde{Y}_{ij} = x_i\beta + h_i h_j^{-1}[Y_j - x_j\beta]. \tag{3.2.68}$$

In Corollary 3.2.1 we demonstrate that $\widetilde{G}_Y(\check{y}; x_i)$ is model consistent for $G_Y(\check{y}; x_i)$ defined in (3.2.45). We use $\widetilde{G}_Y(\check{y}; x_i)$ to construct model consistent estimators of the components of $V\big(\widehat{F}_L(\dot{y}) - F_N(\dot{y}) \bigm| \mathbb{A}_N\big)$, where the components are

$$\begin{aligned} V\Big(T_L \bigm| \mathbb{A}_N\Big) = N^{-2} \sum_{\ell=1}^{B} \sum_{j \in \mathbb{A}_\ell} \sum_{i_1, i_2 \in \mathbb{U}_\ell - \mathbb{A}_\ell} & \omega_{i_1 j}\omega_{i_2 j} \times \\ & \times \big[G_Y\big(\min[\ddot{y}_{i_1 j}, \ddot{y}_{i_2 j}]; x_j\big) - G_Y\big(\ddot{y}_{i_1 j}; x_j\big) G_Y\big(\ddot{y}_{i_2 j}; x_j\big)\big] \end{aligned} \tag{3.2.69}$$

defined in (3.2.61) and

$$V\big(T_N \bigm| \mathbb{A}_N\big) = N^{-2} \sum_{i \in \mathbb{A}_N^c} G_Y(\dot{y}; x_i)[1 - G_Y(\dot{y}; x_i)] \tag{3.2.70}$$

defined in (3.2.60).

Corollary 3.2.1 *Assume* A.1a *through* A.4a *from Theorem 3.2.2. Assume that the value of α in A.2a satisfies $0 < \alpha < 1$. Let the model 3.2.45 hold, and assume* A.5c *through* A.9c *from Theorem 3.2.4. Then,* *(a) For any $\check{y}$ and x_i with $i \in \mathbb{U}_N$, estimator $\widetilde{G}_Y(\check{y}; x_i)$ satisfies*

$$\lim_{N\to\infty} E\Big(\Big|\widetilde{G}_Y(\check{y}; x_i) - G_Y(\check{y}; x_i)\Big| \mid \mathbb{A}_N\Big) = 0,$$

where $\widetilde{G}_Y(\check{y}; x_i)$ is defined in (3.2.67) and $G_Y(\check{y}; x_i)$ is defined in (3.2.45).

(b) The estimator

$$\widetilde{V}\Big(T_L \mid \mathbb{A}_N\Big) = N^{-2} \sum_{\ell=1}^{B} \sum_{j\in\mathbb{A}_\ell} \sum_{i_1,i_2\in\mathbb{U}_\ell-\mathbb{A}_\ell} \omega_{i_1 j}\omega_{i_2 j} \times \\ \times \big[\widetilde{G}_Y\big(\min[\ddot{y}_{i_1 j}, \ddot{y}_{i_2 j}]; x_j\big) - \widetilde{G}_Y\big(\ddot{y}_{i_1 j}; x_j\big)\widetilde{G}_Y\big(\ddot{y}_{i_2 j}; x_j\big)\big] \tag{3.2.71}$$

of $V(T_L \mid \mathbb{A}_N)$ given by (3.2.69), satisfies

$$\lim_{N\to\infty} E\Big(N\Big|\widetilde{V}(T_L \mid \mathbb{A}_N) - V(T_L \mid \mathbb{A}_N)\Big| \mid \mathbb{A}_N\Big) = 0$$

for any $\epsilon > 0$.

(c) The estimator

$$\widetilde{V}\big(T_N \mid \mathbb{A}_N\big) = N^{-2} \sum_{i\in\mathbb{A}_N^c} \widetilde{G}_Y(\dot{y}; x_i)[1 - \widetilde{G}_Y(\dot{y}; x_i)] \tag{3.2.72}$$

of $V(T_N \mid \mathbb{A}_N)$ given by (3.2.60), satisfies

$$\lim_{N\to\infty} E\Big(N\Big|\widetilde{V}(T_N \mid \mathbb{A}_N) - V(T_N \mid \mathbb{A}_N)\Big| \mid \mathbb{A}_N\Big) = 0$$

for any $\epsilon > 0$.

Proof. Part *(a)*. By (3.2.48), (3.2.50) and (3.2.68), all of the following inequalities are equivalent,

$$\widetilde{Y}_{ij} \leqslant \check{y}$$

$$x_i\beta + h_i h_j^{-1}[Y_j - x_j\beta] \leqslant \check{y}$$

$$h_j^{-1}[Y_j - x_j\beta] \leqslant h_i^{-1}[\check{y} - x_i\beta]$$

$$Y_j \leqslant \check{y} + \ddot{\delta}_{ij(\beta)}.$$

Thus, the estimator (3.2.67) can be written as

$$\begin{aligned}\widetilde{G}_Y(\check{y}; x_i) &= \sum_{j\in\mathbb{A}_{\ell_i}} \omega_{ij} I\big(\widetilde{Y}_{ij} \leqslant \check{y}\big)\\ &= \sum_{j\in\mathbb{A}_{\ell_i}} \omega_{ij} I\big(Y_j \leqslant \check{y} + \ddot{\delta}_{ij(\beta)}\big).\end{aligned}$$

Under model (3.2.45), the conditional expectation of $\widetilde{G}_Y(\check{y}; x_i)$ is,

$$\begin{aligned}E\Big(\widetilde{G}_Y(\check{y}; x_i) \mid \mathbb{A}_N\Big) &= E\Big(\sum_{j\in\mathbb{A}_{\ell_i}} \omega_{ij} I\big(Y_j \leqslant \check{y} + \ddot{\delta}_{ij(\beta)}\big) \mid \mathbb{A}_N\Big)\\ &= \sum_{j\in\mathbb{A}_{\ell_i}} \omega_{ij} G_Y\big(\check{y} + \ddot{\delta}_{ij(\beta)}; x_j\big).\end{aligned}$$

By (3.2.57) we have that $\Big[G_Y\big(\check{y}+\ddot{\delta}_{ij(\beta)}; x_j\big) - G_Y(\check{y}; x_i)\Big]$ is $O(B_N^{-1})$ for all $\ell = 1, \ldots, B_N$, $i \in \mathbb{U}_\ell - \mathbb{A}_\ell$ and $j \in \mathbb{A}_\ell$. Because the $O(B_N^{-1})$ upper bound is given for the maximum difference between x_i and x_j when i and j are in the same bin, by (3.2.57) we have

$$\max_{\ell=1,\ldots,B_N}\Big\{\max_{i\in\mathbb{U}_\ell-\mathbb{A}_\ell,\ j\in\mathbb{A}_\ell} G_Y(\check{y} + \ddot{\delta}_{ij}; x_j) - G_Y(\check{y}; x_i)\Big\} = O(B_N^{-1}). \tag{3.2.73}$$

Then,

$$\begin{aligned}E\Big(\widetilde{G}_Y(\check{y}; x_i) - G_Y(\check{y}; x_i) \mid \mathbb{A}_N\Big) &= \sum_{j\in\mathbb{A}_{\ell_i}} \omega_{ij}\Big[G_Y\big(\check{y} + \ddot{\delta}_{ij(\beta)}; x_j\big) - G_Y(\check{y}; x_i)\Big]\\ &= \sum_{j\in\mathbb{A}_{\ell_i}} \omega_{ij} O(B_N^{-1}) = O(B_N^{-1}). \qquad (3.2.74)\end{aligned}$$

Furthermore, by A.4a, we have

$$0 < L_1^* k_N^{-1} < \omega_{ij} < L_2^* k_N^{-1} < \infty,$$

where $\omega_{ij} = \pi_j^{-1}\Big[\sum_{j'\in\mathbb{A}_{\ell_i}} \pi_{j'}^{-1}\Big]^{-1}$. Therefore, it follows that $\omega_{ij}^2 = O(k_N^{-2})$ for all $i \in \mathbb{U}_N$, $j \in \mathbb{A}_N$. Then,

$$\begin{aligned}
V\Big(\widetilde{G}_Y(\check{y};x_i) - G_Y(\check{y};x_i) \mid \mathbb{A}_N\Big) &= V\Big(\widetilde{G}_Y(\check{y};x_i) \mid \mathbb{A}_N\Big) \\
&= V\Big(\sum_{j\in\mathbb{A}_{\ell_i}} \omega_{ij} I\big(Y_j \leqslant \check{y} + \ddot{\delta}_{ij(\beta)}\big) \mid \mathbb{A}_N\Big) \\
&= \sum_{j\in\mathbb{A}_{\ell_i}} \omega_{ij}^2 V\Big(I\big(Y_j \leqslant \check{y} + \ddot{\delta}_{ij(\beta)}\big) \mid \mathbb{A}_N\Big) \\
&= \sum_{j\in\mathbb{A}_{\ell_i}} \omega_{ij}^2 G_Y(\check{y} + \ddot{\delta}_{ij(\beta)}; x_j)\big[1 - G_Y(\check{y} + \ddot{\delta}_{ij(\beta)}; x_j)\big] \\
&\leqslant \sum_{j\in\mathbb{A}_{\ell_i}} O(k_N^{-2})4^{-1} = O(k_N^{-1}). \qquad (3.2.75)
\end{aligned}$$

Then, by Jensen's inequality, (3.2.74) and (3.2.75),

$$\begin{aligned}
\Big\{E\Big(\Big|\widetilde{G}_Y(\check{y};x_i) - G_Y(\check{y};x_i)\Big| \mid \mathbb{A}_N\Big)\Big\}^2 &\leqslant E\Big(\Big|\widetilde{G}_Y(\check{y};x_i) - G_Y(\check{y};x_i)\Big|^2 \mid \mathbb{A}_N\Big) \\
&= \Big[E\Big(\widetilde{G}_Y(\check{y};x_i) - G_Y(\check{y};x_i) \mid \mathbb{A}_N\Big)\Big]^2 \\
&\quad + V\Big(\widetilde{G}_Y(\check{y};x_i) - G_Y(\check{y};x_i) \mid \mathbb{A}_N\Big) \\
&= O(B_N^{-2}) + O(k_N^{-1}), \qquad (3.2.76)
\end{aligned}$$

hence,

$$E\Big(\Big|\widetilde{G}_Y(\check{y};x_i) - G_Y(\check{y};x_i)\Big| \mid \mathbb{A}_N\Big) = O\big[max(B_N^{-1}, k_N^{-1/2})\big]. \qquad (3.2.77)$$

Because, by assumption $0 < \alpha < 1$ for the value of α in A.2a, $B_N \to \infty$ and $k_N \to \infty$ as $N \to \infty$,

$$E\Big(\Big|\widetilde{G}_Y(\check{y};x_i) - G_Y(\check{y};x_i)\Big| \mid \mathbb{A}_N\Big) = o(1), \qquad (3.2.78)$$

and the result follows.

Part *(b)*. Results (3.2.74) and (3.2.75) are independent of $\check{y}$ and of the indexes i, and ℓ. The order of (3.2.74) depends on assumption A.9c, that the $max_\ell(b_\ell) = O(B_N^{-1})$, which is independent of ℓ, $\check{y}$ and x_i. The order of (3.2.75) depends on assumption A.4a about the order of ω_{ij} as $N \to \infty$, which is also independent of ℓ, $\check{y}$ and x_i. Because of (3.2.78), we have that

$$\begin{aligned} E\Big(\Big|\big[\widetilde{G}_Y\big(\min[\ddot{y}_{i_1j}, \ddot{y}_{i_2j}]; x_j\big) - \widetilde{G}_Y\big(\ddot{y}_{i_1j}; x_j\big)\widetilde{G}_Y\big(\ddot{y}_{i_2j}; x_j\big)\big] \\ - \big[G_Y\big(\min[\ddot{y}_{i_1j}, \ddot{y}_{i_2j}]; x_j\big) - G_Y\big(\ddot{y}_{i_1j}; x_j\big)G_Y\big(\ddot{y}_{i_2j}; x_j\big)\big]\Big| \mid \mathbb{A}_N\Big) = o(1), \end{aligned} \tag{3.2.79}$$

for any units i_1, i_2 and j that belong to the same bin ℓ. Then,

$$\begin{aligned} E\Big(N\big[\widetilde{V}(T_L \mid \mathbb{A}_N) - V(T_L \mid \mathbb{A}_N)\big] \mid \mathbb{A}_N\Big) = & \\ = E\Big(\Big|N^{-1}\sum_{\ell=1}^{B}\sum_{j\in\mathbb{A}_\ell}\sum_{i_1,i_2\in\mathbb{U}_\ell-\mathbb{A}_\ell} & \omega_{i_1j}\omega_{i_2j} \times \\ \times \Big\{\big[\widetilde{G}_Y\big(\min[\ddot{y}_{i_1j}, \ddot{y}_{i_2j}]; x_j\big) - & \widetilde{G}_Y\big(\ddot{y}_{i_1j}; x_j\big)\widetilde{G}_Y\big(\ddot{y}_{i_2j}; x_j\big)\big] \\ - \big[G_Y\big(\min[\ddot{y}_{i_1j}, \ddot{y}_{i_2j}]; x_j\big) - & G_Y\big(\ddot{y}_{i_1j}; x_j\big)G_Y\big(\ddot{y}_{i_2j}; x_j\big)\big]\Big\}\Big| \mid \mathbb{A}_N\Big) \\ \leqslant N^{-1}\sum_{\ell=1}^{B}\sum_{j\in\mathbb{A}_\ell}\sum_{i_1,i_2\in\mathbb{U}_\ell-\mathbb{A}_\ell} & \omega_{i_1j}\omega_{i_2j} \times \\ \times E\Big(\Big|\big[\widetilde{G}_Y\big(\min[\ddot{y}_{i_1j}, \ddot{y}_{i_2j}]; x_j\big) - & \widetilde{G}_Y\big(\ddot{y}_{i_1j}; x_j\big)\widetilde{G}_Y\big(\ddot{y}_{i_2j}; x_j\big)\big] \\ - \big[G_Y\big(\min[\ddot{y}_{i_1j}, \ddot{y}_{i_2j}]; x_j\big) - & G_Y\big(\ddot{y}_{i_1j}; x_j\big)G_Y\big(\ddot{y}_{i_2j}; x_j\big)\big]\Big| \mid \mathbb{A}_N\Big) \\ = N^{-1}\sum_{\ell=1}^{B}\sum_{j\in\mathbb{A}_\ell}\sum_{i_1,i_2\in\mathbb{U}_\ell-\mathbb{A}_\ell} & \omega_{i_1j}\omega_{i_2j} \times o(1) \\ = o(1)N^{-1}\sum_{\ell=1}^{B}\sum_{j\in\mathbb{A}_\ell}\Big(\sum_{i\in\mathbb{U}_\ell-\mathbb{A}_\ell} & \omega_{ij}\Big)^2. \end{aligned} \tag{3.2.80}$$

By arguments similar to those used in (3.2.17), we have that the order of the right hand side of (3.2.80) is $o(1)$. Thus, $N\widetilde{V}(T_L \mid \mathbb{A}_N)$ converges to $NV(T_L \mid \mathbb{A}_N)$ in L_1.

Part *(c)*. As in part *(b)*,

$$E\Big(\Big|\widetilde{G}_Y(\dot{y}; x_i)[1 - \widetilde{G}_Y(\dot{y}; x_i)] - G_Y(\dot{y}; x_i)[1 - G_Y(\dot{y}; x_i)]\Big| \mid \mathbb{A}_N\Big) = o(1).$$

Then,

$$
\begin{aligned}
E\Big(\Big|N[\widetilde{V}(T_N \mid \mathbb{A}_N) - V(T_N \mid \mathbb{A}_N)]\Big| \mid \mathbb{A}_N\Big) = \\
&= E\Big(\Big|N^{-1}\sum_{i\in\mathbb{A}_N^c}\Big\{\widetilde{G}_Y(\dot{y};x_i)[1-\widetilde{G}_Y(\dot{y};x_i)] \\
&\qquad - G_Y(\dot{y};x_i)[1-G_Y(\dot{y};x_i)]\Big\}\Big| \mid \mathbb{A}_N\Big) \\
&\leqslant N^{-1}\sum_{i\in\mathbb{A}_N^c} E\Big(\Big|\widetilde{G}_Y(\dot{y};x_i)[1-\widetilde{G}_Y(\dot{y};x_i)] \\
&\qquad - G_Y(\dot{y};x_i)[1-G_Y(\dot{y};x_i)]\Big| \mid \mathbb{A}_N\Big) \\
&= N^{-1}\sum_{i\in\mathbb{A}_N^c} o(1) \\
&= o(1).
\end{aligned}
$$

Thus, $N\widetilde{V}(T_N \mid \mathbb{A}_N)$ converges to $NV(T_N \mid \mathbb{A}_N)$ in L_1. ▲

In section 3.3 we will use the results *(b)* and *(c)* from Corollary 3.2.1 to construct a variance estimator for $\widetilde{F}_{L\beta}(\dot{y}) - F_N(\dot{y})$.

3.2.4 Case D: $Y_i \sim G_Y(\dot{y};x_i)$; β estimated

In this section we study the properties and distribution of $\widehat{F}_L(\dot{y})$ when the parameter β is estimated by $\widehat{\beta}$ given in (3.1.5). As in Section 3.2.3, the assumptions about the superpopulation model are specified in terms of the conditional distribution of Y given x. We will prove in Theorem 3.2.5 that the results of Theorem 3.2.4 hold when the parameter β is estimated from the data. Thus, the local-residuals estimator is model consistent for the finite population distribution function and approximately normally distributed, even if both the mean and the variance of Y given x are misspecified in the superpopulation model.

Theorem 3.2.5 *Let $\{\mathbb{A}_N\}$ be a sequence of samples selected from the sequence of finite populations $\{\mathbb{U}_N\}$. Assume that the sample $\mathbb{A}_N$ is divided into B_N groups, each of size k_N, as described in (3.1.3). Assume a superpopulation model where the Y_i are independent and*

$$P(Y_i \leqslant y \mid x_i) = G_Y(y; x_i) \tag{3.2.81}$$

for $i \in \mathbb{U}_N$. The set $\{x_1, x_2, \ldots, x_N\}$ is assumed fixed and known. Let $h_i = h(x_i)$, where $h(\cdot)$ is the function used in constructing estimator (3.1.4). Assume that there exists an m_h such that $0 < m_h \leqslant h(x_i) < \infty$ for $i \in \mathbb{U}_N$. Assume A.1a *through* A.4a *from Theorem 3.2.2, and* A.5c *through* A.7c *from Theorem 3.2.4. Also assume*

A.8d *The positive function $h(x)$ is differentiable and there exists an $\dot{M}_{hx}$ such that*

$$|h_j(h_j^{-1}x_j - h_i^{-1}x_i) \leqslant |x_i - x_j|\dot{M}_{hx}$$

for all x.

A.9d *The* $\max_\ell(b_\ell) = O(B_N^{-1})$, *where b_ℓ is the length of bin ℓ.*

A.10d *The sequence of $\{x_i : i \in \mathbb{A}_N\}$ is such that $|\widehat{\beta} - \beta| = O_p(N^{-1/2})$.*

Let $\widehat{F}_L(\dot{y})$ be the estimator (3.1.4) with the estimated β used in (3.1.6) and let $\widetilde{F}_{L\beta}(\dot{y})$ be the estimator (3.1.4) with the true β used in (3.1.6). Then,

(a) The sequence

$$N^{1/2}\Big\{\widehat{F}_L(\dot{y}) - \widetilde{F}_{L\beta}(\dot{y})\Big\}$$

converges to zero in probability for a fixed point $\dot{y}$.

(b) If the α of assumption A.2a is greater than 0.5, then the sequence

$$\Big\{V\big(\widetilde{F}_{L\beta}(\dot{y}) - F_N(\dot{y}) \mid \mathbb{A}_N\big)\Big\}^{-1/2}\Big\{\widehat{F}_L(\dot{y}) - F_N(\dot{y})\Big\}$$

converges in distribution to a standard normal random variable, where $V\big(\widetilde{F}_{L\beta}(\dot{y}) - F_N(\dot{y}) \bigm| \mathbb{A}_N\big)$ *is given in Theorem 3.2.4.*

Proof. Part *(a)*. The estimation error $\Big\{\widehat{F}_L(\dot{y}) - F_N(\dot{y})\Big\}$ can be decomposed into two parts,

$$\widehat{F}_L(\dot{y}) - F_N(\dot{y}) = \Big\{\widehat{F}_L(\dot{y}) - \widetilde{F}_{L\beta}(\dot{y})\Big\} + \Big\{\widetilde{F}_{L\beta}(\dot{y}) - F_N(\dot{y})\Big\}. \tag{3.2.82}$$

For $N^{1/2}\big[\widehat{F}_L(\dot{y}) - \widetilde{F}_{L\beta}(\dot{y})\big]$ to converge to zero in probability, we need to show that for any $\lambda > 0$ and $\epsilon > 0$, there exists $N_{\lambda\epsilon}$ such that $N > N_{\lambda\epsilon}$ implies that

$$P\big(N^{1/2}|\widehat{F}_L(\dot{y}) - \widetilde{F}_{L\beta}(\dot{y})| > \lambda \bigm| \mathbb{A}_N\big) < \epsilon. \tag{3.2.83}$$

Let D_N be the event $D_N = \Big\{N^{1/2}|\widehat{F}_L(\dot{y}) - \widetilde{F}_{L\beta}(\dot{y})| > \lambda\Big\}$. By assumption A.10d, for any $\epsilon > 0$ we can find $\eta = O(N^{-1/2})$ and $N_{\eta\epsilon}$ such that for $N > N_{\eta\epsilon}$,

$$P\big(|\widehat{\beta} - \beta| \geqslant \eta\big) < \epsilon/2.$$

Then for all $N > N_{\eta\epsilon}$,

$$\begin{aligned} P\big(D_N \bigm| \mathbb{A}_N\big) &= P\big(|\widehat{\beta} - \beta| \geqslant \eta\big)P\big(D_N \bigm| \mathbb{A}_N, |\widehat{\beta} - \beta| \geqslant \eta\big) + \\ &\quad P\big(|\widehat{\beta} - \beta| < \eta\big)P\big(D_N \bigm| \mathbb{A}_N, |\widehat{\beta} - \beta| < \eta\big) \\ &< \epsilon/2 + \\ &\quad P\big(|\widehat{\beta} - \beta| < \eta\big)P\big(D_N \bigm| \mathbb{A}_N, |\widehat{\beta} - \beta| < \eta\big) \\ &\leqslant \epsilon/2 + P\big(D_N \bigm| \mathbb{A}_N, |\widehat{\beta} - \beta| < \eta\big). \end{aligned} \tag{3.2.84}$$

We extend the notation for $\ddot{\delta}_{ij}$ defined in (3.2.50) to

$$\ddot{\delta}_{ij(b)} = h_j(h_i^{-1} - h_j^{-1})\dot{y} + h_j(h_j^{-1}x_j - h_i^{-1}x_i)b, \tag{3.2.85}$$

to make explicit whether $\ddot{\delta}_{ij}$ is computed using β or $\widehat{\beta}$. Let

$$\Delta_{ij} = I\big(Y_j \leqslant \dot{y} + \ddot{\delta}_{ij(\widehat{\beta})}\big) - I\big(Y_j \leqslant \dot{y} + \ddot{\delta}_{ij(\beta)}\big).$$

By (3.2.48), we have that

$$\begin{aligned} N^{1/2}\Big|\widehat{F}_L(\dot{y}) - \widetilde{F}_{L\beta}(\dot{y})\Big| &= N^{1/2}\Big|N^{-1}\sum_{i\in\mathbb{A}_N^c}\sum_{j\in\mathbb{A}_{\ell_i}}\omega_{ij}I\big(Y_j \leqslant \dot{y} + \ddot{\delta}_{ij(\widehat{\beta})}\big) - \\ &\qquad - N^{-1}\sum_{i\in\mathbb{A}_N^c}\sum_{j\in\mathbb{A}_{\ell_i}}\omega_{ij}I\big(Y_j \leqslant \dot{y} + \ddot{\delta}_{ij(\beta)}\big)\Big| \\ &\leqslant N^{-1/2}\sum_{i\in\mathbb{A}_N^c}\sum_{j\in\mathbb{A}_{\ell_i}}\omega_{ij}\Big|I\big(Y_j \leqslant \dot{y} + \ddot{\delta}_{ij(\widehat{\beta})}\big) - I\big(Y_j \leqslant \dot{y} + \ddot{\delta}_{ij(\beta)}\big)\Big| \\ &= N^{-1/2}\sum_{i\in\mathbb{A}_N^c}\sum_{j\in\mathbb{A}_{\ell_i}}\omega_{ij}\Big|\Delta_{ij}\Big|. \end{aligned} \tag{3.2.86}$$

We will prove that (3.2.86) converges to zero in L_1 conditional on $\mathbb{A}_N$ and $|\widehat{\beta} - \beta| < \eta$, that is,

$$E\Big[N^{-1/2}\sum_{i\in\mathbb{A}^c}\sum_{j\in\mathbb{A}_{\ell_i}}\omega_{ij}|\Delta_{ij}| \;\big|\; \mathbb{A}_N, |\widehat{\beta} - \beta| < \eta\Big] \longrightarrow 0 \tag{3.2.87}$$

as $N \to \infty$. Note that, conditional on $\mathbb{A}_N$ and $|\widehat{\beta} - \beta| < \eta$,

- $|\Delta_{ij}|$ can only take the values 0 or 1,
- $|\Delta_{ij}| = 1$ only when

  $\dot{y} + \ddot{\delta}_{ij(\widehat{\beta})} < Y_j \leqslant \dot{y} + \ddot{\delta}_{ij(\beta)}$ or

  $\dot{y} + \ddot{\delta}_{ij(\beta)} < Y_j \leqslant \dot{y} + \ddot{\delta}_{ij(\widehat{\beta})}$,
- $\ddot{\delta}_{ij(b)}$ is a monotone function of b,
- $\ddot{\delta}_{ij(\widehat{\beta})}$ is then restricted to be between

$$m_\eta = \min\big(\ddot{\delta}_{ij(\beta+\eta)}, \ddot{\delta}_{ij(\beta-\eta)}\big) \text{ and } M_\eta = \max\big(\ddot{\delta}_{ij(\beta+\eta)}, \ddot{\delta}_{ij(\beta-\eta)}\big),$$

- by (3.2.50), the distance between m_η and M_η is

$$\begin{aligned} M_\eta - m_\eta &= |\ddot{\delta}_{ij(\beta+\eta)} - \ddot{\delta}_{ij(\beta-\eta)}| \\ &= |(\beta+\eta) - (\beta-\eta)| \times |h_j(h_j^{-1}x_j - h_i^{-1}x_i)| \\ &= 2\eta|h_j(h_j^{-1}x_j - h_i^{-1}x_i)|. \end{aligned} \tag{3.2.88}$$

The expected value of (3.2.86) given $|\widehat{\beta}-\beta| < \eta$ and $\mathbb{A}_N$ is

$$\begin{aligned} E\big[N^{-1/2}\sum_{i\in\mathbb{A}^c}\sum_{j\in\mathbb{A}_{\ell_i}} \omega_{ij}|\Delta_{ij}| \bigm| \mathbb{A}_N, |\widehat{\beta}-\beta| < \eta\big] &= \\ &= N^{-1/2}\sum_{i\in\mathbb{A}^c}\sum_{j\in\mathbb{A}_{\ell_i}} \omega_{ij}E\big(|\Delta_{ij}| \bigm| |\widehat{\beta}-\beta| < \eta\big) \\ &= N^{-1/2}\sum_{i\in\mathbb{A}^c}\sum_{j\in\mathbb{A}_{\ell_i}} \omega_{ij}P\big(|\Delta_{ij}| = 1 \bigm| |\widehat{\beta}-\beta| < \eta\big), \end{aligned} \tag{3.2.89}$$

with

$$\begin{aligned} P\big(|\Delta_{ij}| = 1 \bigm| |\widehat{\beta}-\beta| < \eta\big) &= P\big(\dot{y} + m_\eta < Y_j \leqslant \dot{y} + \ddot{\delta}_{ij(\beta)}\big) + \\ & P\big(\dot{y} + \ddot{\delta}_{ij(\beta)} < Y_j \leqslant \dot{y} + M_\eta\big) \\ &= G_Y(\dot{y} + M_\eta; x_j) - G_Y(\dot{y} + m_\eta; x_j). \end{aligned} \tag{3.2.90}$$

An upper bound for (3.2.90) is

$$\begin{aligned} G_Y(\dot{y} + M_\eta; x_j) - G_Y(\dot{y} + m_\eta; x_j) &\leqslant (M_\eta - m_\eta)\dot{M}_{gg} && \text{by A.7d} \\ &= 2\eta|h_j(h_j^{-1}x_j - h_i^{-1}x_i)|\dot{M}_{gg} && \text{by (3.2.88)} \\ &\leqslant 2\eta|x_i - x_j|\dot{M}_{hx}\dot{M}_{gg} && \text{by A.8d} \\ &\leqslant 2\eta(\max_\ell b_\ell)\dot{M}_{hx}\dot{M}_{gg}. \end{aligned} \tag{3.2.91}$$

The order of (3.2.91) depends on the order of η and $\max_\ell b_\ell$, since for a fixed $\dot{y}$ both $\dot{M}_{hx}$ and $\dot{M}_{gg}$ are constants. By A.10d, $\eta = O(N^{-1/2})$ and, by A.9d, $\max_\ell b_\ell = O(B_N^{-1})$.

Substituting (3.2.91) into (3.2.89) we have that

$$
\begin{aligned}
&E\big[N^{-1/2}\sum_{i\in\mathbb{A}_N^c}\sum_{j\in\mathbb{A}_{\ell_i}}\omega_{ij}|\Delta_{ij}|\;\big|\;\mathbb{A}_N,|\widehat{\beta}-\beta|<\eta\big] = \\
&\qquad = N^{-1/2}\sum_{i\in\mathbb{A}_N^c}\sum_{j\in\mathbb{A}_{\ell_i}}\omega_{ij}P\big[|\Delta_{ij}|=1\;\big|\;\mathbb{A}_N,|\widehat{\beta}-\beta|<\eta\big] \\
&\qquad \leqslant N^{-1/2}\sum_{i\in\mathbb{A}_N^c}\sum_{j\in\mathbb{A}_{\ell_i}}\omega_{ij}2\eta(\max_{\ell} b_\ell)\dot{M}_{hx}\dot{M}_{gg} \\
&\qquad = N^{-1/2}\sum_{i\in\mathbb{A}_N^c}\sum_{j\in\mathbb{A}_{\ell_i}}\omega_{ij}O(N^{-1/2})O(B_N^{-1}) \\
&\qquad = O(N^{-1}B_N^{-1})\sum_{i\in\mathbb{A}_N^c}\sum_{j\in\mathbb{A}_{\ell_i}}\omega_{ij} \\
&\qquad = O(B_N^{-1}), \qquad (3.2.92)
\end{aligned}
$$

since $\sum_{j\in\mathbb{A}_{\ell_i}}\omega_{ij}=1$. By A.2a, $B_N=O(N^\alpha)\to\infty$ as $N\to\infty$, which implies that (3.2.86) converges to zero in L_1. Then, conditional on $\mathbb{A}_N$ and $|\widehat{\beta}-\beta|<\eta$, (3.2.86) converges to zero in probability, and we can find $N^*_{\lambda\epsilon}$ such that for any $N>N^*_{\lambda\epsilon}$,

$$
P\big(N^{-1/2}\sum_{i\in\mathbb{A}_N^c}\sum_{j\in\mathbb{A}_{\ell_i}}\omega_{ij}|\Delta_{ij}|>\lambda\;\big|\;\mathbb{A}_N,|\widehat{\beta}-\beta|<\eta\big)<\epsilon/2. \qquad (3.2.93)
$$

By (3.2.86), $N^{1/2}\Big|\widehat{F}_L(\dot{y})-\widetilde{F}_{L\beta}(\dot{y})\Big|\leqslant N^{-1/2}\sum_{i\in\mathbb{A}_N^c}\sum_{j\in\mathbb{A}_{\ell_i}}\omega_{ij}\Big|\Delta_{ij}\Big|$. Then, the occurrence of the event $D_N=\Big\{N^{1/2}\Big|\widehat{F}_L(\dot{y})-\widetilde{F}_{L\beta}(\dot{y})\Big|>\lambda\Big\}$ implies the occurrence of $\Big\{N^{-1/2}\sum_{i\in\mathbb{A}_N^c}\sum_{j\in\mathbb{A}_{\ell_i}}\omega_{ij}\Big|\Delta_{ij}\Big|>\lambda\Big\}$. In other words,

$$
D_N=\Big\{N^{1/2}\Big|\widehat{F}_L(\dot{y})-\widetilde{F}_{L\beta}(\dot{y})\Big|>\lambda\Big\}\subseteq\Big\{N^{-1/2}\sum_{i\in\mathbb{A}_N^c}\sum_{j\in\mathbb{A}_{\ell_i}}\omega_{ij}\Big|\Delta_{ij}\Big|>\lambda\Big\},
$$

and

$$
\mathrm{P}\big(D_N\;\big|\;\mathbb{A}_N,|\widehat{\beta}-\beta|<\eta\big)\leqslant\mathrm{P}\big(N^{-1/2}\sum_{i\in\mathbb{A}_N^c}\sum_{j\in\mathbb{A}_{\ell_i}}\omega_{ij}\Big|\Delta_{ij}\Big|>\lambda\;\big|\;\mathbb{A}_N,|\widehat{\beta}-\beta|<\eta\big)<\epsilon/2 \qquad (3.2.94)
$$

for all $N>N^*_{\lambda\epsilon}$ by (3.2.93).

For all $N > \max(N_{\eta\epsilon}, N^*_{\lambda\epsilon})$, we have that both (3.2.84) and (3.2.94) hold. Then for any $N > \max(N_{\eta\epsilon}, N^*_{\lambda\epsilon})$

$$P\big(N^{1/2}|\widehat{F}_L(\dot{y}) - \widetilde{F}_{L\beta}(\dot{y})| > \lambda \bigm| \mathbb{A}_N\big) = P\big(D_N \bigm| \mathbb{A}_N\big) < \epsilon.$$

Hence,

$$N^{-1/2}\Big|\widehat{F}_L(\dot{y}) - \widetilde{F}_{L\beta}(\dot{y})\Big| \to 0$$

in probability.

Part *(b)*. We can write each of the terms of the sequence

$$\Big\{V\big(\widetilde{F}_{L\beta}(\dot{y}) - F_N(\dot{y}) \bigm| \mathbb{A}_N\big)\Big\}^{-1/2}\Big\{\widehat{F}_L(\dot{y}) - F_N(\dot{y})\Big\}$$

as

$$\begin{aligned}\Big\{V\big(\widetilde{F}_{L\beta}(\dot{y}) - F_N(\dot{y}) \bigm| \mathbb{A}_N\big)\Big\}^{-1/2}\Big\{\widehat{F}_L(\dot{y}) - F_N(\dot{y})\Big\} = \\ = \Big\{V\big(\widetilde{F}_{L\beta}(\dot{y}) - F_N(\dot{y}) \bigm| \mathbb{A}_N\big)\Big\}^{-1/2}\Big\{\widehat{F}_L(\dot{y}) - \widetilde{F}_{L\beta}(\dot{y})\Big\} + \\ + \Big\{V\big(\widetilde{F}_{L\beta}(\dot{y}) - F_N(\dot{y}) \bigm| \mathbb{A}_N\big)\Big\}^{-1/2}\Big\{\widetilde{F}_{L\beta}(\dot{y}) - F_N(\dot{y})\Big\}.\end{aligned} \tag{3.2.95}$$

In part *(a)* we showed that $N^{1/2}\Big\{\widehat{F}_L(\dot{y}) - \widetilde{F}_{L\beta}(\dot{y})\Big\}$ converges to zero in probability as $N \to \infty$. By (3.2.64) we have that $V\big(\widetilde{F}_{L\beta}(\dot{y}) - F_N(\dot{y}) \bigm| \mathbb{A}_N\big)$ is $O(N^{-1})$. Then the first term on the right hand side of (3.2.95),

$$\Big\{V\big(\widetilde{F}_{L\beta}(\dot{y}) - F_N(\dot{y}) \bigm| \mathbb{A}_N\big)\Big\}^{-1/2}\Big\{\widehat{F}_L(\dot{y}) - \widetilde{F}_{L\beta}(\dot{y})\Big\}$$

converges to zero in probability. By Slutsky's theorem, we have that

$$\Big\{V\big(\widetilde{F}_{L\beta}(\dot{y}) - F_N(\dot{y}) \bigm| \mathbb{A}_N\big)\Big\}^{-1/2}\Big\{\widehat{F}_L(\dot{y}) - F_N(\dot{y})\Big\}$$

and

$$\Big\{V\big(\widetilde{F}_{L\beta}(\dot{y}) - F_N(\dot{y}) \bigm| \mathbb{A}_N\big)\Big\}^{-1/2}\Big\{\widetilde{F}_{L\beta}(\dot{y}) - F_N(\dot{y})\Big\}$$

have the same asymptotic distribution given by part *(b)* of Theorem 3.2.4. ▲

3.3 Variance Estimation. Confidence Interval Construction

Chambers and Dunstan (1986) present an estimator for the variance of the estimation error of $\widehat{F}_{CD}(\dot{y})$. The variance of $\widehat{F}_{CD}(\dot{y}) - F_N(\dot{y})$, shown in (2.3.8), has two terms, one term that depends on the sample units, denoted by $\mathbf{W}_r^*(\dot{y}, \beta)$, and one term that depends on the unobserved units, denoted by $\mathbf{W}_r(\dot{y}, \beta)$. The term $\mathbf{W}_r^*(\dot{y}, \beta)$ is similar to the term $V\big(T_L \bigm| \mathbb{A}_N\big)$ that appears in (3.2.61), except that $V\big(T_L \bigm| \mathbb{A}_N\big)$ does not contain the variation due to the estimation of the parameter β. The terms $\mathbf{W}_r(\dot{y}, \beta)$ and $V(T_N \bigm| \mathbb{A}_N)$ are equal up to a constant, $V(T_N \bigm| \mathbb{A}_N) = (1 - nN^{-1})^2 \mathbf{W}_r(\dot{y}, \beta)$. .

Rao, Kovar and Mantel (1990) present a variance estimator for $\widehat{F}_{RKMdm}(\dot{y})$ that is the variance estimator of a difference estimator. However, Rao, Kovar and Mantel (1990) do not give an estimator for the variance of $\widehat{F}_{RKMdm}(\dot{y}) - F_N(\dot{y})$.

In this section we will present estimators for the variance of the estimation error $\widehat{F}_L(\dot{y}) - F_N(\dot{y})$ and the variance of the estimator $\widehat{F}_L(\dot{y})$ as an estimator of the superpopulation distribution function. The estimator of the variance of $\widehat{F}_L(\dot{y}) - F_N(\dot{y})$ is based on the variance of $\widehat{F}_L(\dot{y}) - F_N(\dot{y})$ given in Theorem 3.2.4. For the variance of $\widehat{F}_L(\dot{y}) - F(\dot{y})$ we present two estimators that are based on the Jackknife resampling method. Another estimator of the variance of $\widehat{F}_L(\dot{y})$, based on the distribution of the $\ddot{Z}_j$ defined in assumption A.6c of Theorem 3.2.4, is also suggested.

In Theorem 3.2.5 we showed that $\widetilde{F}_{L\beta}(\dot{y})$ and $\widehat{F}_L(\dot{y})$ have the same limiting distribution. The effect of estimating β is of smaller order, $O(N^{-1}B_N^{-1})$, than the order, $O(N^{-1})$, of the variance of estimators $\widetilde{F}_{L\beta}(\dot{y})$ and $\widehat{F}_L(\dot{y})$. We will consider the variance of $\widetilde{F}_{L\beta}(\dot{y}) - F(\dot{y})$ and the variance of $\widetilde{F}_{L\beta}(\dot{y}) - F_N(\dot{y})$ as approximations to the variance of $\widehat{F}_L(\dot{y}) - F(\dot{y})$ and the variance of $\widehat{F}_L(\dot{y}) - F_N(\dot{y})$, respectively. Furthermore, since by A.10d, $|\widehat{\beta} - \beta| = O_p(N^{-1/2})$, we will replace β by $\widehat{\beta}$ to estimate the variances of $\widetilde{F}_{L\beta}(\dot{y})$

and $\widetilde{F}_{L\beta}(\dot{y}) - F_N(\dot{y})$.

3.3.1 Estimation of the variance of $\widetilde{F}_{L\beta}(\dot{y}) - F_N(\dot{y})$

By Theorem 3.2.4 and (3.2.59), the conditional variance of $\widetilde{F}_{L\beta}(\dot{y}) - F_N(\dot{y})$ is

$$V\big(\widetilde{F}_{L\beta}(\dot{y}) - F_N(\dot{y}) \bigm| \mathbb{A}_N\big) = V\big(T_L \bigm| \mathbb{A}_N\big) + V\big(T_N \bigm| \mathbb{A}_N\big), \tag{3.3.1}$$

where

$$\begin{aligned} V\big(T_L \bigm| \mathbb{A}_N\big) = N^{-2}\sum_{\ell=1}^{B}\sum_{j\in\mathbb{A}_\ell}\sum_{i_1,i_2\in\mathbb{U}_\ell-\mathbb{A}_\ell} & \omega_{i_1 j}\omega_{i_2 j} \times \\ & \times \big[G_Y\big(\min[\ddot{y}_{i_1 j}, \ddot{y}_{i_2 j}]; x_j\big) - G_Y\big(\ddot{y}_{i_1 j}; x_j\big)G_Y\big(\ddot{y}_{i_2 j}; x_j\big)\big], \end{aligned}$$

and

$$V\big(T_N \bigm| \mathbb{A}_N\big) = N^{-2}\sum_{i\in\mathbb{A}_N^c} G_Y(\dot{y}; x_i)[1 - G_Y(\dot{y}; x_i)].$$

The term T_L is the part of $\widetilde{F}_{L\beta}(\dot{y}) - F_N(\dot{y})$ that depends on the sample units. The term T_N is the part of the estimation error that depends on the nonsample units. The following Theorem suggests a model consistent estimator for the analytical variance, given in (3.3.1), of $\widetilde{F}_{L\beta}(\dot{y}) - F_N(\dot{y})$.

Theorem 3.3.1 *Assume* A.1a *through* A.4a *from Theorem 3.2.2. Assume that the value of α in A.2a satisfies $0 < \alpha < 1$. Let the model 3.2.45 hold, and assume* A.5c *through* A.9c *from Theorem 3.2.4. Then,*

$$\lim_{N\to\infty} P\Big(\big|\widetilde{V}_{ee}(\dot{y}) - V\big(\widetilde{F}_{L\beta}(\dot{y}) - F_N(\dot{y}) \bigm| \mathbb{A}_N\big)\big| > \epsilon \bigm| \mathbb{A}_N\Big) = 0,$$

where

$$\widetilde{V}_{ee}(\dot{y}) = \widetilde{V}\Big(T_L \bigm| \mathbb{A}_N\Big) + \widetilde{V}\Big(T_N \bigm| \mathbb{A}_N\Big), \tag{3.3.2}$$

$\widetilde{V}\Big(T_L \bigm| \mathbb{A}_N\Big)$ *is defined in (3.2.71) and* $\widetilde{V}\Big(T_N \bigm| \mathbb{A}_N\Big)$ *is defined in (3.2.72).*

Proof. By part *(a)* of Corollary 3.2.1 we have that $\widetilde{G}_Y(\dot{y}; x_i)$ is model consistent for $G_Y(\dot{y}; x_i)$, where $\widetilde{G}_Y(\dot{y}; x_i)$ is defined in (3.2.67). Moreover, by parts *(b)* and *(c)* of Corollary 3.2.1, $n\widetilde{V}\Big(T_L \mid \mathbb{A}_N\Big)$ is model consistent for $nV\big(T_L \mid \mathbb{A}_N\big)$ and $n\widetilde{V}\Big(T_N \mid \mathbb{A}_N\Big)$ is model consistent for $nV\big(T_N \mid \mathbb{A}_N\big)$. Therefore, from (3.3.1) an (3.3.2), it follows that (3.3.2) is model consistent for $V\big(\widetilde{F}_{L\beta}(\dot{y}) - F_N(\dot{y}) \mid \mathbb{A}_N\big)$. ▲

Note that estimator (3.3.2) allows us to evaluate the contribution of each component of the estimation error, the part due to the sample, and the part due to the unobserved units. In Theorem 3.3.2 we show that $\widetilde{V}_{ee}(\dot{y})$ can be used to construct tests of hypotheses and confidence intervals for $\widetilde{F}_{L\beta}(\dot{y}) - F_N(\dot{y})$.

Theorem 3.3.2 *Assume* A.1a *through* A.4a *from Theorem 3.2.2. Let the model 3.2.45 hold, and assume* A.5c *through* A.9c *from Theorem 3.2.4. Also assume that the value of α in A.2a satisfies* $0.5 < \alpha < 1$. *Then,*

$$\Big\{\widetilde{V}_{ee}(\dot{y})\Big\}^{-1/2}\Big\{\widetilde{F}_{L\beta}(\dot{y}) - F_N(\dot{y})\Big\} \tag{3.3.3}$$

converges in distribution to a $N(0,1)$*, where* $\widetilde{V}_{ee}(\dot{y})$ *is defined in (3.3.2).*

Proof. By Theorem 3.2.4,

$$\Big\{V\big(\widetilde{F}_{L\beta}(\dot{y}) - F_N(\dot{y}) \mid \mathbb{A}_N\big)\Big\}^{-1/2}\Big\{\widetilde{F}_{L\beta}(\dot{y}) - F_N(\dot{y})\Big\}$$

converges in distribution to a $N(0,1)$. By Theorem 3.3.1, $n\widetilde{V}_{ee}(\dot{y})$ converges in probability to $nV\big(\widetilde{F}_{L\beta}(\dot{y}) - F_N(\dot{y}) \mid \mathbb{A}_N\big)$. Then,

$$\begin{aligned}\Big\{V\big(\widetilde{F}_{L\beta}(\dot{y}) - F_N(\dot{y}) \mid \mathbb{A}_N\big)\Big\}^{-1/2}&\Big\{\widetilde{F}_{L\beta}(\dot{y}) - F_N(\dot{y})\Big\} = \\ &= \Big\{\widetilde{V}_{ee}(\dot{y})\Big\}^{-1/2}\Big\{\widetilde{F}_{L\beta}(\dot{y}) - F_N(\dot{y})\Big\} + \\ &\quad + \bigg(\Big\{\widetilde{V}_{ee}(\dot{y})\Big\}^{-1/2} - \Big\{V\big(\widetilde{F}_{L\beta}(\dot{y}) - F_N(\dot{y}) \mid \mathbb{A}_N\big)\Big\}^{-1/2}\bigg)\Big\{\widetilde{F}_{L\beta}(\dot{y}) - F_N(\dot{y})\Big\}.\end{aligned} \tag{3.3.4}$$

The last term in the right hand side of (3.3.4) converges to zero in probability. Thus, by Slutsky's Theorem,

$$\Big\{\widetilde{V}_{ee}(\dot{y})\Big\}^{-1/2}\Big\{\widetilde{F}_{L\beta}(\dot{y}) - F_N(\dot{y})\Big\}$$

and

$$\Big\{V\big(\widetilde{F}_{L\beta}(\dot{y}) - F_N(\dot{y}) \bigm| \mathbb{A}_N\big)\Big\}^{-1/2}\Big\{\widetilde{F}_{L\beta}(\dot{y}) - F_N(\dot{y})\Big\}$$

have the same limiting distribution, given by Theorem 3.2.4. ▲

By Theorem 3.3.2, we can use $\widetilde{V}_{ee}(\dot{y})$ and (3.3.3) to construct confidence sets and do hypothesis testing for the finite population distribution function $F_N(\dot{y})$.

3.3.2 Estimation of the variance of $\widetilde{F}_{L\beta}(\dot{y}) - F(\dot{y})$

The estimator $\widehat{F}_L(\dot{y})$ is a weighted mean of indicator funtions as shown in (3.1.9). Similarly, estimator $\widetilde{F}_{L\beta}(\dot{y})$, where the true β is used in (3.1.6), is also a weighted mean of indicator functions, namely,

$$\widetilde{F}_{L\beta}(\dot{y}) = N^{-1}\Big[\sum_{j\in\mathbb{A}_N} I(Y_j \leqslant \dot{y}) + \sum_{i\in\mathbb{A}_N^c}\sum_{j\in\mathbb{A}_{\ell_i}} \omega_{ij} I\big(h_j^{-1}[Y_j - x_j\beta] \leqslant h_i^{-1}[\dot{y} - x_i\beta]\big)\Big].$$

Rearranging terms, we can write $\widetilde{F}_{L\beta}(\dot{y})$ as

$$\begin{aligned}\widetilde{F}_{L\beta}(\dot{y}) &= N^{-1}\sum_{j\in\mathbb{A}_N}\Big[I(Y_j \leqslant \dot{y}) + \sum_{i\in\mathbb{U}_{\ell_j}-\mathbb{A}_{\ell_j}} \omega_{ij} I\big(h_j^{-1}[Y_j - x_j\beta] \leqslant h_i^{-1}[\dot{y} - x_i\beta]\big)\Big]\\ &= N^{-1}\sum_{j\in\mathbb{A}_N}\Big[I(Y_j \leqslant \dot{y}) + \ddot{Z}_j\Big],\end{aligned}$$

where

$$\begin{aligned}\ddot{Z}_j &= \sum_{i\in\mathbb{U}_{\ell_j}-\mathbb{A}_{\ell_j}} \omega_{ij} I\big(h_j^{-1}[Y_j - x_j\beta] \leqslant h_i^{-1}[\dot{y} - x_i\beta]\big)\\ &= \sum_{i\in\mathbb{U}_{\ell_j}-\mathbb{A}_{\ell_j}} \omega_{ij} I\big(U_j \leqslant \dot{u}_i\big)\\ &= \sum_{i\in\mathbb{U}_{\ell_j}-\mathbb{A}_{\ell_j}} \omega_{ij} I\big(Y_j \leqslant \ddot{y}_{ij}\big)\end{aligned}$$

is defined in A.6c in Theorem 3.2.4, $U_j = h_j^{-1}[Y_j - x_j\beta]$, $\dot{u}_i = h_i^{-1}[\dot{y} - x_i\beta]$, and ℓ_j is the bin that contains unit j.

The analytical variance of $\widetilde{F}_{L\beta}(\dot{y}) - F(\dot{y})$ can be computed and estimated by using an estimator similar to the one decribed in Theorem 3.3.2. In this section we will focus on the construction of alternative estimators.

The local-residuals estimator is based on the model

$$Y_i = x_i\beta + h(x_i)U_i, \tag{3.3.5}$$

where the U_i are independent and identically distributed random variables with mean zero and distribution function $G(u)$. Under model (3.3.5) the $\ddot{Z}_j$ are conditionally independent given $\mathbb{A}_N$. In addition to being independent, for each bin ℓ, the k random variables $\ddot{Z}_j$ with $j \in \mathbb{A}_\ell$ are identically distributed.

Let $I_J = I\big(Y_j \leqslant \dot{y}\big)$. Then, I_j are independent, but not identically distributed, even when the units are in the same bin. The expectation and variance of $I\big(Y_j \leqslant \dot{y}\big)$ are $G\big(h_j^{-1}[\dot{y} - x_j\beta]\big)$ and $G\big(h_j^{-1}[\dot{y} - x_j\beta]\big)\big[1 - G\big(h_j^{-1}[\dot{y} - x_j\beta]\big)\big]$, respectively. Under the assumptions of Theorem 3.2.4, for j_1 and j_2 in the same bin ℓ, we have that

$$\lim_{N\to\infty} \big|E\big[I\big(Y_{j_1} \leqslant \dot{y}\big) \mid \mathbb{A}_N\big] - E\big[I\big(Y_{j_2} \leqslant \dot{y}\big) \mid \mathbb{A}_N\big]\big| = 0, \tag{3.3.6}$$

and,

$$\lim_{N\to\infty} \big|V\big[I\big(Y_{j_1} \leqslant \dot{y}\big) \mid \mathbb{A}_N\big] - V\big[I\big(Y_{j_2} \leqslant \dot{y}\big) \mid \mathbb{A}_N\big]\big| = 0. \tag{3.3.7}$$

Results (3.3.6) and (3.3.7) suggest that we can approximate the model variance of $\widetilde{F}_{L\beta}(\dot{y})$ by the variance of the n independent variables

$$\mathcal{L}_j = I\big(Y_j \leqslant \dot{y}\big) + \ddot{Z}_j, \tag{3.3.8}$$

for $j \in \mathbb{A}_\ell$ and $\ell = 1, \ldots, B_N$. Asymptotically the $\mathcal{L}_j$ for all j in the same bin, have the same model mean and model variance. Because the $\mathcal{L}_j$ are independent,

$$\begin{aligned} V\big(\widetilde{F}_{L\beta}(\dot{y}) \bigm| \mathbb{A}_N\big) &= V\big(N^{-1} \sum_{j \in \mathbb{A}_N} \mathcal{L}_j \bigm| \mathbb{A}_N\big) \\ &= V\big(N^{-1} \sum_{\ell=1}^{B_N} \sum_{j \in \mathbb{A}_\ell} \mathcal{L}_j \bigm| \mathbb{A}_N\big) \\ &= N^{-2} \sum_{\ell=1}^{B_N} V\big(\sum_{j \in \mathbb{A}_\ell} \mathcal{L}_j \bigm| \mathbb{A}_N\big) \\ &= N^{-2} \sum_{\ell=1}^{B_N} k_N \sigma_\ell^2, \end{aligned} \tag{3.3.9}$$

where $\sigma_\ell^2 = k_N^{-1} \sum_{j \in \mathbb{A}_\ell} V\big(\mathcal{L}_j \bigm| \mathbb{A}_N\big)$. Note that $V\big(\mathcal{L}_j \bigm| \mathbb{A}_N\big)$ for $j \in \mathbb{A}_\ell$ are not equal to σ_ℓ^2, but the quantities $\{V\big(\mathcal{L}_j \bigm| \mathbb{A}_N\big) - \sigma_\ell^2\}$ converge to zero as $N \to \infty$. We propose two estimators of $V\big(\widetilde{F}_{L\beta}(\dot{y}) \bigm| \mathbb{A}_N\big)$:

(1) an estimator based on the sample variance of the $\mathcal{L}_j$,

(2) a Jackknife estimator constructed by iteratively deleting one unit from the sample at a time and recomputing the local-residuals estimator with the $n-1$ remaining units.

The estimator based on the sample variance of the $\mathcal{L}_j$ is,

$$\widetilde{V}_{\mathcal{L}}(\dot{y}) = N^{-2} \sum_{\ell=1}^{B_N} (k-1)^{-1} \sum_{j \in \mathbb{A}_\ell} (\mathcal{L}_j - \bar{\mathcal{L}}_\ell)^2, \tag{3.3.10}$$

where $\bar{\mathcal{L}}_\ell = k^{-1} \sum_{j \in \mathbb{A}_\ell} \mathcal{L}_j$, and the sample variances $\widehat{\sigma}_\ell^2 = (k-1)^{-1} \sum_{j \in \mathbb{A}_\ell} (\mathcal{L}_j - \bar{\mathcal{L}}_\ell)^2$ are asymptotically unbiased for σ_ℓ^2, for $\ell = 1, \ldots, B$. The Jackknife based estimator is

$$\widetilde{V}_{JK}(\dot{y}) = n^{-1} \sum_{\ell=1}^{B_N} \sum_{\alpha \in \mathbb{A}_\ell} \left(\widetilde{F}^*_{L\beta(-\alpha)}(\dot{y}) - \widetilde{F}_{L\beta}(\dot{y})\right)^2, \tag{3.3.11}$$

where $\widetilde{F}^*_{L\beta(-\alpha)}(\dot{y})$ is the local-residuals estimator computed from the reduced sample $\mathbb{A}_N - \{\alpha\}$. The reduced sample $\mathbb{A}_N - \{\alpha\}$ is the set of indices remaining when unit α

is removed from bin ℓ_α in the original sample $\mathbb{A}_N$. The estimator computed from the reduced sample is

$$\widetilde{F}^*_{L\beta(-\alpha)}(\dot{y}) = N^{-1}\Big\{\sum_{\ell\neq\ell_\alpha}\sum_{j\in\mathbb{A}_\ell}\mathcal{L}_j + \sum_{j\in\mathbb{A}^*_{\ell_\alpha}}\mathcal{L}^*_j\Big\}, \tag{3.3.12}$$

where $\mathbb{A}^*_{\ell_\alpha} = \mathbb{A}_\ell - \{\alpha\}$,

$$\mathcal{L}^*_j = I(Y_j \leqslant \dot{y}) + \sum_{i\in\mathbb{U}_{\ell_\alpha}-\mathbb{A}^*_{\ell_\alpha}} \omega^*_{ij} I\big(h_j^{-1}[Y_j - x_j\beta] \leqslant h_i^{-1}[\dot{y} - x_i\beta]\big),$$

and, $\omega^*_{ij} = \pi_j^{-1}\Big[\sum_{j'\in\mathbb{A}^*_{\ell_\alpha}}\pi_{j'}^{-1}\Big]^{-1}$ are the adjusted weights when unit α is deleted.

In practice, we have to estimate β to compute each $\widetilde{F}^*_{L\beta(-\alpha)}(\dot{y})$. We computed two versions of the Jackknife estimator: (1) one version that uses the $\widehat{\beta}$ and the $\widehat{y}_{ij}$ computed from the sample $\mathbb{A}_N$, and (2) another version that recomputes $\widehat{\beta}$ and $\widehat{y}_{ij}$ for each of the n reduced samples $\mathbb{A}_N - \{\alpha\}$.

3.4 Estimation of the Superpopulation Distribution Function

The finite population distribution function has N jumps of magnitude N^{-1} at the points y_i for $i \in \mathbb{U}_N$, provided that the y_i are different. Once the sample is selected and the y_j are observed for $j \in \mathbb{A}_N$, we know where n of the N jumps are located, provided that the y_j are different. From a superpopulation perspective, the y_j for $j \in \mathbb{A}_N$ are a particular realization of the random variables Y_j. Recall that the superpopulation distribution function is defined in (2.1.5) as

$$\begin{aligned} F(\dot{y}) = \mathrm{P}(Y \leqslant \dot{y}) &= N^{-1} \sum_{i \in \mathbb{U}} \mathrm{P}(Y_i \leqslant \dot{y} \mid \mathbf{x} = \mathbf{x}_i) \\ &= N^{-1} \sum_{i \in \mathbb{U}} E\{I(Y_i \leqslant \dot{y}) \mid \mathbf{x} = \mathbf{x}_i\}. \end{aligned}$$

We define an estimator of the superpopulation distribution function as

$$\widehat{F}_L^{fi}(\dot{y}) = N^{-1} \sum_{i \in \mathbb{U}} \widehat{G}_{L\ell_i}(h_i^{-1}[\dot{y} - x_i\widehat{\beta}]), \tag{3.4.1}$$

where $\widehat{G}_{L\ell_i}$ and $\widehat{\beta}$ are defined in (3.1.6) and (3.1.5), respectively. The full-imputation local-residuals estimator of (3.4.1) can be written as

$$\widehat{F}_L^{fi}(\dot{y}) = N^{-1} \Big[\sum_{j \in \mathbb{A}} \widehat{G}_{L\ell_j}(h_j^{-1}[\dot{y} - x_j\widehat{\beta}]) + \sum_{i \in \mathbb{A}^c} \widehat{G}_{L\ell_i}(h_i^{-1}[\dot{y} - x_i\widehat{\beta}]) \Big]. \tag{3.4.2}$$

The difference between the local-residuals estimator (3.1.4),

$$\widehat{F}_L(\dot{y}) = N^{-1} \Big[\sum_{j \in \mathbb{A}} I(Y_j \leqslant \dot{y}) + \sum_{i \in \mathbb{A}^c} \widehat{G}_{L\ell_i}(h_i^{-1}[\dot{y} - x_i\widehat{\beta}]) \Big],$$

and the estimator (3.4.2) is that the distribution function of the residuals is also estimated for the sample units in (3.4.1), while in estimator (3.1.4) the quantities $I(Y_j \leqslant \dot{y})$ are taken for the units in the sample.

We will consider the distribution of estimator (3.4.1) under the superpopulation model used in Theorem 3.2.4 for β known and for β estimated from the data.

3.4.1 Case E: $Y_i \sim G_Y(\dot{y}; x_i)$, β known

We proved in Theorem 3.2.4 that the local-residuals estimator (3.1.4) is robust against misspecifications of the mean and variance functions in the superpopulation model. We will study the properties of the full-imputation local-residuals estimator $\widehat{F}_L^{fi}(\dot{y})$ as an estimator of the superpopulation distribution function. In Theorem 3.4.1 we will show model consistency and limiting normality of estimator (3.4.1).

Theorem 3.4.1 *Let $\{\mathbb{A}_N\}$ be a sequence of samples selected from the sequence of finite populations $\{\mathbb{U}_N\}$. Assume that the sample $\mathbb{A}_N$ is divided into B_N groups, each of size k_N, as described in (3.1.3). Assume a superpopulation model where the Y_i are independent and*

$$P(Y_i \leqslant y \mid x_i) = G_Y(y; x_i) \tag{3.4.3}$$

for $i \in \mathbb{U}_N$. Let $h_i = h(x_i)$, where $h(\cdot)$ is the function used in constructing estimator (3.4.1). Assume that there exists an m_h such that $0 < m_h \leqslant h(x_i) < \infty$ for $i \in \mathbb{U}_N$. Let $\dot{y}$ be a fixed point. Assume A.1a *through* A.4a *from Theorem 3.2.2,* A.7c *through* A.9c *from Theorem 3.2.4, and*

A.6e *For any N, $\left\{N^{-1}\sum_{j\in\mathbb{A}_N} V(\check{Z}_j \mid \mathbb{A}_N)\right\}$ is positive, where*

$$V(\check{Z}_j \mid \mathbb{A}_N) = \sum_{i_1,i_2\in\mathbb{U}_\ell} \omega_{i_1j}\omega_{i_2j}\left[G_Y\left(\min[\ddot{y}_{i_1j}, \ddot{y}_{i_2j}]; x_j\right) - G_Y\left(\ddot{y}_{i_1j}; x_j\right)G_Y\left(\ddot{y}_{i_2j}; x_j\right)\right],$$

$$\check{Z}_j = \sum_{i\in\mathbb{U}_\ell} \omega_{ij} I\left(Y_j \leqslant \ddot{y}_{ij}\right), j \in \mathbb{A}_N,$$

$$\ddot{y}_{ij} = \dot{y} + h_j(h_i^{-1} - h_j^{-1})\dot{y} + h_j(h_j^{-1}x_j - h_i^{-1}x_i)\beta,$$

ℓ is the bin that contains unit j and $\mathbb{U}_\ell$ is the set of indices in bin ℓ.

Let $\widetilde{F}_{L\beta}^{fi}(\dot{y})$ be the estimator (3.4.1) with the true β used in (3.1.6). Then, *(a) The estimator* $\widetilde{F}^{fi}_{L\beta}(\dot{y})$ *satisfies*

$$\lim_{N\to\infty} P\big(\big|\widetilde{F}^{fi}_{L\beta}(\dot{y}) - F(\dot{y})\big| > \epsilon \bigm| \mathbb{A}_N\big) = 0$$

for all $\epsilon > 0$.

(b) If the α *in A.2a is greater than 0.5, then the sequence*

$$\Big\{V\big(\widetilde{F}^{fi}_{L\beta}(\dot{y}) \bigm| \mathbb{A}_N\big)\Big\}^{-1/2}\Big\{\widetilde{F}^{fi}_{L\beta}(\dot{y}) - F(\dot{y})\Big\} \tag{3.4.4}$$

converges in distribution to a $N(0,1)$ *random variable, where*

$$\begin{aligned} V\big(\widetilde{F}^{fi}_{L\beta}(\dot{y}) \bigm| \mathbb{A}_N\big) = N^{-2}\sum_{\ell=1}^{B}\sum_{j\in\mathbb{A}_\ell}\sum_{i_1,i_2\in\mathbb{U}_\ell} \omega_{i_1 j}\omega_{i_2 j} \times \\ \times \big[G_Y\big(\min[\ddot{y}_{i_1 j}, \ddot{y}_{i_2 j}]; x_j\big) - G_Y\big(\ddot{y}_{i_1 j}; x_j\big)G_Y\big(\ddot{y}_{i_2 j}; x_j\big)\big]. \end{aligned} \tag{3.4.5}$$

Proof. Part *(a)*. Estimator (3.4.1) can be written as

$$\begin{aligned} \widetilde{F}^{fi}_{L\beta}(\dot{y}) &= N^{-1}\sum_{i\in\mathbb{U}_N}\sum_{j\in\mathbb{A}_{\ell_i}} \omega_{ij} I\big(h_j^{-1}[Y_j - x_j\beta] \leqslant h_i^{-1}[\dot{y} - x_i\beta]\big) \\ &= N^{-1}\sum_{i\in\mathbb{U}_N}\sum_{j\in\mathbb{A}_{\ell_i}} \omega_{ij} I\big(Y_j \leqslant \dot{y} + \ddot{\delta}_{ij}\big) \end{aligned} \tag{3.4.6}$$

by (3.2.48), where $\ddot{\delta}_{ij} = h_j(h_i^{-1} - h_j^{-1})\dot{y} + h_j(h_i^{-1}x_i - h_j^{-1}x_j)\beta$ is defined in (3.2.50). The superpopulation distribution function can be written as

$$\begin{aligned} F(\dot{y}) = \mathrm{P}(Y \leqslant \dot{y}) &= N^{-1}\sum_{i\in\mathbb{U}_N} \mathrm{P}(Y_i \leqslant \dot{y} \bigm| \mathbf{x} = \mathbf{x}_i) \\ &= N^{-1}\sum_{i\in\mathbb{U}_N} G_Y(\dot{y}; x_i) \\ &= N^{-1}\sum_{\ell=1}^{B_N}\sum_{i\in\mathbb{U}_\ell} G_Y(\dot{y}; x_i) \\ &= N^{-1}\sum_{\ell=1}^{B_N}\sum_{i\in\mathbb{U}_\ell}\sum_{j\in\mathbb{A}_\ell} \omega_{ij} G_Y(\dot{y}; x_i) \\ &= N^{-1}\sum_{\ell=1}^{B_N}\sum_{j\in\mathbb{A}_\ell}\sum_{i\in\mathbb{U}_\ell} \omega_{ij} G_Y(\dot{y}; x_i), \end{aligned} \tag{3.4.7}$$

because the $\sum_{j \in \mathbb{A}_\ell} \omega_{ij} = 1$. Combining (3.4.6) and (3.4.7), the estimation error can be expressed as

$$\widetilde{F}_{L\beta}^{fi}(\dot{y}) - F(\dot{y}) = N^{-1} \sum_{\ell=1}^{B_N} \sum_{j \in \mathbb{A}_\ell} \sum_{i \in \mathbb{U}_\ell} \omega_{ij} \Big[I\big(Y_j \leqslant \dot{y} + \ddot{\delta}_{ij}\big) - G_Y(\dot{y}; x_i) \Big]. \tag{3.4.8}$$

The model expectation of (3.4.8) is

$$\begin{aligned} E\Big[\widetilde{F}_{L\beta}^{fi}(\dot{y}) - F(\dot{y}) \mid \mathbb{A}_N\Big] &= E\Big[N^{-1} \sum_{\ell=1}^{B_N} \sum_{j \in \mathbb{A}_\ell} \sum_{i \in \mathbb{U}_\ell} \omega_{ij} \{ I\big(Y_j \leqslant \dot{y} + \ddot{\delta}_{ij}\big) - G_Y(\dot{y}; x_i) \} \mid \mathbb{A}_N \Big] \\ &= N^{-1} \sum_{\ell=1}^{B_N} \sum_{j \in \mathbb{A}_\ell} \sum_{i \in \mathbb{U}_\ell} \omega_{ij} \Big[G_Y(\dot{y} + \ddot{\delta}_{ij}; x_j) - G_Y(\dot{y}; x_i) \Big]. \end{aligned}$$

In Theorem 3.2.4 we proved that under assumptions A.7c, A.8c and A.9c,

$$\max_{\ell=1,\ldots,B_N} \max_{i \in \mathbb{U}_\ell,\ j \in \mathbb{U}_\ell} \Big[G_Y(\dot{y} + \ddot{\delta}_{ij}; x_j) - G_Y(\dot{y}; x_i) \Big] = O(B_N^{-1})$$

as shown in (3.2.57). Then, we have that

$$\begin{aligned} \Big| E\Big[\widetilde{F}_{L\beta}^{fi}(\dot{y}) - F(\dot{y}) \mid \mathbb{A}_N\Big] \Big| &= \Big| N^{-1} \sum_{\ell=1}^{B_N} \sum_{j \in \mathbb{A}_\ell} \sum_{i \in \mathbb{U}_\ell} \omega_{ij} \Big[G_Y(\dot{y} + \ddot{\delta}_{ij}; x_j) - G_Y(\dot{y}; x_i) \Big] \Big| \\ &\leqslant N^{-1} \sum_{\ell=1}^{B_N} \sum_{j \in \mathbb{A}_\ell} \sum_{i \in \mathbb{U}_\ell} \omega_{ij} \Big| G_Y(\dot{y} + \ddot{\delta}_{ij}; x_j) - G_Y(\dot{y}; x_i) \Big| \\ &= N^{-1} \sum_{\ell=1}^{B_N} \sum_{j \in \mathbb{A}_\ell} \sum_{i \in \mathbb{U}_\ell} \omega_{ij} O(B_N^{-1}) \\ &= O(B_N^{-1}). \end{aligned} \tag{3.4.9}$$

Model unbiasedness of estimator (3.4.1) follows, since by A.2a we have that $B_N \to \infty$ as $N \to \infty$. Result (3.4.9) shows that model unbiasedness of (3.4.1) holds even when both the mean and variance function of Y given x are misspecified.

The model variance of $\widetilde{F}^{fi}_{L\beta}(\dot{y})$ is

$$V\big(\widetilde{F}^{fi}_{L\beta}(\dot{y}) \mid \mathbb{A}_N\big) = V\big(N^{-1}\sum_{\ell=1}^{B_N}\sum_{j\in\mathbb{A}_\ell}\sum_{i\in\mathbb{U}_\ell}\omega_{ij}I\big(Y_j \leqslant \dot{y}+\ddot{\delta}_{ij}\big) \mid \mathbb{A}_N\big)$$
$$= N^{-2}\sum_{\ell=1}^{B_N}\sum_{j\in\mathbb{A}_\ell} V\Big(\sum_{i\in\mathbb{U}_\ell}\omega_{ij}I\big(Y_j \leqslant \dot{y}+\ddot{\delta}_{ij}\big) \mid \mathbb{A}_N\Big), \tag{3.4.10}$$

because the Y_j are independent under model (3.4.3). Using the $\check{Z}_j$ defined in A.6e we can write (3.4.10) as

$$V\big(\widetilde{F}^{fi}_{L\beta}(\dot{y}) \mid \mathbb{A}_N\big) = N^{-2}\sum_{\ell=1}^{B_N}\sum_{j\in\mathbb{A}_\ell} V\Big(\check{Z}_j \mid \mathbb{A}_N\Big)$$
$$= N^{-2}\sum_{j\in\mathbb{A}_N} V\Big(\check{Z}_j \mid \mathbb{A}_N\Big), \tag{3.4.11}$$

where

$$V\Big(\check{Z}_j \mid \mathbb{A}_N\Big) = V\Big(\sum_{i\in\mathbb{U}_{\ell_j}}\omega_{ij}I\big(Y_j \leqslant \dot{y}+\ddot{\delta}_{ij}\big)\Big)$$
$$= V\Big(\sum_{i\in\mathbb{U}_{\ell_j}}\omega_{ij}I\big(Y_j \leqslant \ddot{y}_{ij}\big)\Big)$$

by (3.2.49), where ℓ_j is the index of the bin that contains unit j. Then,

$$V\Big(\check{Z}_j \mid \mathbb{A}_N\Big) = V\Big(\sum_{i\in\mathbb{U}_{\ell_j}}\omega_{ij}I\big(Y_j \leqslant \ddot{y}_{ij}\big)\Big)$$
$$= \sum_{i_1,i_2\in\mathbb{U}_{\ell_j}}\omega_{i_1j}\omega_{i_2j}\mathrm{Cov}\big[I\big(Y_j \leqslant \ddot{y}_{i_1j}\big), I\big(Y_j \leqslant \ddot{y}_{i_2j}\big)\big]$$
$$= \sum_{i_1,i_2\in\mathbb{U}_{\ell_j}}\omega_{i_1j}\omega_{i_2j}\big[G_Y\big(\min[\ddot{y}_{i_1j},\ddot{y}_{i_2j}];x_j\big) - G_Y\big(\ddot{y}_{i_1j};x_j\big)G_Y\big(\ddot{y}_{i_2j};x_j\big)\big]. \tag{3.4.12}$$

Then, combining (3.4.10) and (3.4.12), we have that

$$V\Big(\widetilde{F}^{fi}_{L\beta}(\dot{y}) \mid \mathbb{A}_N\Big) = N^{-2}\sum_{\ell=1}^{B_N}\sum_{j\in\mathbb{A}_\ell}\sum_{i_1,i_2\in\mathbb{U}_\ell}\omega_{i_1j}\omega_{i_2j}\times$$
$$\times\big[G_Y\big(\min[\ddot{y}_{i_1j},\ddot{y}_{i_2j}];x_j\big) - G_Y\big(\ddot{y}_{i_1j};x_j\big)G_Y\big(\ddot{y}_{i_2j};x_j\big)\big]. \tag{3.4.13}$$

The covariances that appear in (3.4.13) are bounded, in absolute value, by 1/4. Then,

$$
\begin{aligned}
V\Big(\widetilde{F}^{fi}_{L\beta}(\dot{y}) \mid \mathbb{A}_N\Big) &= \Big| N^{-2}\sum_{\ell=1}^{B_N}\sum_{j\in\mathbb{A}_\ell}\sum_{i_1,i_2\in\mathbb{U}_\ell} \omega_{i_1 j}\omega_{i_2 j} \times \\
&\qquad \times \big[G_Y\big(\min[\ddot{y}_{i_1 j},\ddot{y}_{i_2 j}];x_j\big) - G_Y\big(\ddot{y}_{i_1 j};x_j\big)G_Y\big(\ddot{y}_{i_2 j};x_j\big)\big]\Big| \\
&\leqslant N^{-2}\sum_{\ell=1}^{B_N}\sum_{j\in\mathbb{A}_\ell}\sum_{i_1,i_2\in\mathbb{U}_\ell} \omega_{i_1 j}\omega_{i_2 j} \times \\
&\qquad \times \Big|\big[G_Y\big(\min[\ddot{y}_{i_1 j},\ddot{y}_{i_2 j}];x_j\big) - G_Y\big(\ddot{y}_{i_1 j};x_j\big)G_Y\big(\ddot{y}_{i_2 j};x_j\big)\big]\Big| \\
&\leqslant N^{-2}\sum_{\ell=1}^{B_N}\sum_{j\in\mathbb{A}_\ell}\sum_{i_1,i_2\in\mathbb{U}_\ell} \omega_{i_1 j}\omega_{i_2 j}\ 4^{-1} \\
&= 4^{-1}N^{-2}\sum_{\ell=1}^{B_N}\sum_{j\in\mathbb{A}_\ell}\Big(\sum_{i\in\mathbb{U}_\ell}\omega_{ij}\Big)^2 \\
&= 4^{-1}N^{-2}\sum_{\ell=1}^{B_N}\sum_{j\in\mathbb{A}_\ell} O(1) \\
&= N^{-1}O(1) = O(N^{-1}) \qquad (3.4.14)
\end{aligned}
$$

because $\big(\sum_{i\in\mathbb{U}_\ell}\omega_{ij}\big)$ is $O(1)$ as shown in (3.2.16). Thus, we have that $\widetilde{F}^{fi}_{L\beta}(\dot{y})$ is asymptotically model unbiased, by (3.4.9), and that the model variance of $\widetilde{F}^{fi}_{L\beta}(\dot{y})$ goes to zero as $N \to \infty$, by (3.4.14). Then, $\widetilde{F}^{fi}_{L\beta}(\dot{y})$ is model consistent for the superpopulation distribution function.

Part (b). We proved in (3.4.13) that the model variance of $\widetilde{F}^{fi}_{L\beta}(\dot{y})$ is given by (3.4.5). We must show that (3.4.4) converges in distribution to a standard normal. The estimator $\widetilde{F}^{fi}_{L\beta}(\dot{y})$ is a weighted sum of indicator functions, as expressed in (3.4.6). All the moments of the indicator functions exist, then, A.6e is sufficient for the Lyapounov condition for the sum $N^{-1}\sum_{j\in\mathbb{A}_N}\check{Z}_j$. Then, since α is assumed to be greater than 0.5,

$$
\begin{aligned}
&\Big\{V\big(\widetilde{F}^{fi}_{L\beta}(\dot{y}) \mid \mathbb{A}_N\big)\Big\}^{-1/2}\Big\{\widetilde{F}^{fi}_{L\beta}(\dot{y}) - F(\dot{y})\Big\} = \\
&\qquad = \Big\{V\big(N^{-1}\sum_{j\in\mathbb{A}_N}\check{Z}_j \mid \mathbb{A}_N\big)\Big\}^{-1/2}\Big\{N^{-1}\sum_{j\in\mathbb{A}_N}\check{Z}_j - F(\dot{y})\Big\}
\end{aligned}
$$

converges in distribution to a standard normal as $N \to \infty$.

3.4.2 Case F: $Y_i \sim G_Y(\dot{y}; x_i)$; β estimated

In Section 3.2.4 we proved that the results for estimator $\widetilde{F}_{L\beta}(\dot{y})$ given in Theorem 3.2.4 hold when β is estimated from the data. We will prove in Theorem 3.4.2 that the results for estimator $\widetilde{F}^{fi}_{L\beta}(\dot{y})$ given in Theorem 3.4.1 also hold when β is estimated. Theorem 3.4.2 essentially reproduces Theorem 3.2.5 for the full-imputation local-residuals estimator.

Theorem 3.4.2 *Let $\{\mathbb{A}_N\}$ be a sequence of samples selected from the sequence of finite populations $\{\mathbb{U}_N\}$. Assume that the sample $\mathbb{A}_N$ is divided into B_N groups, each of size k_N, as described in (3.1.3). Assume a superpopulation model where the Y_i are independent and*

$$P(Y_i \leqslant y \mid x_i) = G_Y(y; x_i) \tag{3.4.15}$$

for $i \in \mathbb{U}_N$. The set $\{x_1, \dots, x_N\}$ is assumed fixed and known. Let $h_i = h(x_i)$, where $h(\cdot)$ is the function used in constructing estimator (3.4.1). Assume that there exists an m_h such that $0 < m_h \leqslant h(x_i) < \infty$ for $i \in \mathbb{U}_N$. Let $\dot{y}$ be a fixed point. Assume A.1a *through* A.4a *from Theorem 3.2.2,* A.5c *through* A.7c *from Theorem 3.2.4, and* A.8d *through* A.10d *from Theorem 3.2.5.*

Let $\widehat{F}^{fi}_L(\dot{y})$ be the estimator (3.4.1) and let $\widetilde{F}^{fi}_{L\beta}(\dot{y})$ be the estimator (3.4.1) with the true β used in (3.1.6). Then,

(a) The sequence

$$N^{1/2}\Big\{\widehat{F}^{fi}_L(\dot{y}) - \widetilde{F}^{fi}_{L\beta}(\dot{y})\Big\}$$

converges to zero in probability.

(b) If the α in A.2a is greater than 0.5, then the sequence

$$\Big\{V\big(\widetilde{F}^{fi}_{L\beta}(\dot{y}) \mid \mathbb{A}_N\big)\Big\}^{-1/2}\Big\{\widehat{F}^{fi}_L(\dot{y}) - F(\dot{y})\Big\}$$

converges in distribution to a $N(0,1)$ *random variable, where* $V\big(\widetilde{F}^{fi}_{L\beta}(\dot{y}) \mid \mathbb{A}_N\big)$ *is given in Theorem 3.4.1.*

Proof. Part *(a)*. The estimation error $\Big\{\widehat{F}^{fi}_{L}(\dot{y}) - F(\dot{y})\Big\}$ can be decomposed into two parts,

$$\Big\{\widehat{F}^{fi}_{L}(\dot{y}) - F(\dot{y})\Big\} = \Big\{\widehat{F}^{fi}_{L}(\dot{y}) - \widetilde{F}^{fi}_{L\beta}(\dot{y})\Big\} + \Big\{\widetilde{F}^{fi}_{L\beta}(\dot{y}) - F(\dot{y})\Big\}. \tag{3.4.16}$$

For $N^{1/2}\big[\widehat{F}^{fi}_{L}(\dot{y}) - \widetilde{F}^{fi}_{L\beta}(\dot{y})\big]$ to converge to zero in probability, we need to show that for any $\lambda > 0$ and $\epsilon > 0$, there exists $N_{\lambda\epsilon}$ such that $N > N_{\lambda\epsilon}$ implies that

$$P\big(N^{1/2}|\widehat{F}^{fi}_{L}(\dot{y}) - \widetilde{F}^{fi}_{L\beta}(\dot{y})| > \lambda \mid \mathbb{A}_N\big) < \epsilon. \tag{3.4.17}$$

Let $D_N = \Big\{N^{1/2}|\widehat{F}^{fi}_{L}(\dot{y}) - \widetilde{F}^{fi}_{L\beta}(\dot{y})| > \lambda\Big\}$. By assumption A.10d, for any $\epsilon > 0$ we can find $\eta = O(N^{-1/2})$ and $N_{\eta\epsilon}$ such that for $N > N_{\eta\epsilon}$,

$$P\big(|\widehat{\beta} - \beta| \geqslant \eta\big) < \epsilon/2.$$

Then for all $N > N_{\eta\epsilon}$,

$$\begin{aligned} P\big(D_N \mid \mathbb{A}_N\big) &= P\big(|\widehat{\beta}-\beta| \geqslant \eta\big)P\big(D_N \mid \mathbb{A}_N, |\widehat{\beta}-\beta| \geqslant \eta\big) + \\ &\quad P\big(|\widehat{\beta}-\beta| < \eta\big)P\big(D_N \mid \mathbb{A}_N, |\widehat{\beta}-\beta| < \eta\big) \\ &< \epsilon/2 + \\ &\quad P\big(|\widehat{\beta}-\beta| < \eta\big)P\big(D_N \mid \mathbb{A}_N, |\widehat{\beta}-\beta| < \eta\big) \\ &\leqslant \epsilon/2 + P\big(D_N \mid \mathbb{A}_N, |\widehat{\beta}-\beta| < \eta\big). \end{aligned} \tag{3.4.18}$$

We extend the notation for $\ddot{\delta}_{ij}$ defined in (3.2.50) to

$$\ddot{\delta}_{ij(b)} = h_j(h_i^{-1} - h_j^{-1})\dot{y} + h_j(h_j^{-1}x_j - h_i^{-1}x_i)b, \tag{3.4.19}$$

to make explicit whether $\ddot{\delta}_{ij}$ is computed using β or $\widehat{\beta}$. Let

$$\Delta_{ij} = I\big(Y_j \leqslant \dot{y} + \ddot{\delta}_{ij(\widehat{\beta})}\big) - I\big(Y_j \leqslant \dot{y} + \ddot{\delta}_{ij(\beta)}\big).$$

By (3.4.6), we have that

$$\begin{aligned} N^{1/2}\Big|\widehat{F}_L^{fi}(\dot{y}) - \widetilde{F}_{L\beta}^{fi}(\dot{y})\Big| &= N^{1/2}\Big|N^{-1}\sum_{i\in\mathbb{U}_N}\sum_{j\in\mathbb{A}_{\ell_i}} \omega_{ij} I\big(Y_j \leqslant \dot{y} + \ddot{\delta}_{ij(\widehat{\beta})}\big) - \\ &\qquad - N^{-1}\sum_{i\in\mathbb{U}_N}\sum_{j\in\mathbb{A}_{\ell_i}} \omega_{ij} I\big(Y_j \leqslant \dot{y} + \ddot{\delta}_{ij(\beta)}\big)\Big| \\ &\leqslant N^{-1/2}\sum_{i\in\mathbb{U}_N}\sum_{j\in\mathbb{A}_{\ell_i}} \omega_{ij}\Big|I\big(Y_j \leqslant \dot{y} + \ddot{\delta}_{ij(\widehat{\beta})}\big) - I\big(Y_j \leqslant \dot{y} + \ddot{\delta}_{ij(\beta)}\big)\Big| \\ &= N^{-1/2}\sum_{i\in\mathbb{U}_N}\sum_{j\in\mathbb{A}_{\ell_i}} \omega_{ij}\Big|\Delta_{ij}\Big|. \end{aligned} \tag{3.4.20}$$

We will prove that (3.4.20) converges to zero in L_1 conditional on $\mathbb{A}_N$ and $|\widehat{\beta}-\beta| < \eta$, that is,

$$E\Big[N^{-1/2}\sum_{i\in\mathbb{U}_N}\sum_{j\in\mathbb{A}_{\ell_i}} \omega_{ij}|\Delta_{ij}| \;\big|\; \mathbb{A}_N, |\widehat{\beta}-\beta| < \eta\Big] \longrightarrow 0 \tag{3.4.21}$$

as $N \to \infty$. Note that, conditional on $\mathbb{A}_N$ and $|\widehat{\beta}-\beta| < \eta$,

- $|\Delta_{ij}|$ can only take the values 0 or 1,
- $|\Delta_{ij}| = 1$ only when

  $\dot{y} + \ddot{\delta}_{ij(\widehat{\beta})} < Y_j \leqslant \dot{y} + \ddot{\delta}_{ij(\beta)}$ or

  $\dot{y} + \ddot{\delta}_{ij(\beta)} < Y_j \leqslant \dot{y} + \ddot{\delta}_{ij(\widehat{\beta})}$,
- $\ddot{\delta}_{ij(b)}$ is a monotone function of b,
- $\ddot{\delta}_{ij(\widehat{\beta})}$ is then restricted to be between

$$m_\eta = \min\big(\ddot{\delta}_{ij(\beta+\eta)}, \ddot{\delta}_{ij(\beta-\eta)}\big) \text{ and } M_\eta = \max\big(\ddot{\delta}_{ij(\beta+\eta)}, \ddot{\delta}_{ij(\beta-\eta)}\big),$$

- by (3.2.50), the distance between m_η and M_η is

$$\begin{aligned} M_\eta - m_\eta &= |\ddot{\delta}_{ij(\beta+\eta)} - \ddot{\delta}_{ij(\beta-\eta)}| \\ &= |(\beta+\eta)-(\beta-\eta)| \times |h_j(h_j^{-1}x_j - h_i^{-1}x_i)| \\ &= 2\eta|h_j(h_j^{-1}x_j - h_i^{-1}x_i)|. \end{aligned} \tag{3.4.22}$$

The expected value of (3.4.20) given $|\widehat{\beta}-\beta| < \eta$ and $\mathbb{A}_N$ is

$$\begin{aligned} &E\big[N^{-1/2}\sum_{i\in\mathbb{U}_N}\sum_{j\in\mathbb{A}_{\ell_i}} \omega_{ij}|\Delta_{ij}| \bigm| \mathbb{A}_N, |\widehat{\beta}-\beta|<\eta\big] = \\ &\qquad = N^{-1/2}\sum_{i\in\mathbb{U}_N}\sum_{j\in\mathbb{A}_{\ell_i}} \omega_{ij} E\big(|\Delta_{ij}| \bigm| |\widehat{\beta}-\beta|<\eta\big) \\ &\qquad = N^{-1/2}\sum_{i\in\mathbb{U}_N}\sum_{j\in\mathbb{A}_{\ell_i}} \omega_{ij} P\big(|\Delta_{ij}| = 1 \bigm| |\widehat{\beta}-\beta|<\eta\big), \end{aligned} \tag{3.4.23}$$

with

$$\begin{aligned} P\big(|\Delta_{ij}| = 1 \bigm| |\widehat{\beta}-\beta|<\eta\big) &= P\big(\dot{y}+m_\eta < Y_j \leqslant \dot{y}+\ddot{\delta}_{ij(\beta)}\big) + \\ & P\big(\dot{y}+\ddot{\delta}_{ij(\beta)} < Y_j \leqslant \dot{y}+M_\eta\big) \\ &= G_Y(\dot{y}+M_\eta;x_j) - G_Y(\dot{y}+m_\eta;x_j). \end{aligned} \tag{3.4.24}$$

An upper bound for (3.4.24) is

$$\begin{aligned} G_Y(\dot{y}+M_\eta;x_j) - G_Y(\dot{y}+m_\eta;x_j) &\leqslant (M_\eta - m_\eta)\dot{M}_{gg} && \text{by A.7d} \\ &= 2\eta|h_j(h_j^{-1}x_j - h_i^{-1}x_i)|\dot{M}_{gg} && \text{by (3.4.22)} \\ &\leqslant 2\eta|x_i - x_j|\dot{M}_{hx}\dot{M}_{gg} && \text{by A.8d} \\ &\leqslant 2\eta(\max_\ell b_\ell)\dot{M}_{hx}\dot{M}_{gg}. \end{aligned} \tag{3.4.25}$$

The order of (3.4.25) depends on the order of η and $\max_\ell b_\ell$, since for a fixed $\dot{y}$ both $\dot{M}_{hx}$ and $\dot{M}_{gg}$ are constants. By A.10d, $\eta = O(N^{-1/2})$ and, by A.9d, $\max_\ell b_\ell = O(B_N^{-1})$.

Substituting (3.4.25) into (3.4.23) we have that

$$\begin{aligned}
E\big[N^{-1/2}\sum_{i\in\mathbb{U}_N}\sum_{j\in\mathbb{A}_{\ell_i}}\omega_{ij}|\Delta_{ij}| \;\big|\; \mathbb{A}_N, |\widehat{\beta}-\beta|<\eta\big] = & \\
&= N^{-1/2}\sum_{i\in\mathbb{U}_N}\sum_{j\in\mathbb{A}_{\ell_i}}\omega_{ij}P\big[|\Delta_{ij}|=1 \;\big|\; \mathbb{A}_N, |\widehat{\beta}-\beta|<\eta\big] \\
&= N^{-1/2}\sum_{i\in\mathbb{U}_N}\sum_{j\in\mathbb{A}_{\ell_i}}\omega_{ij}2\eta(\max_{\ell} b_\ell)\dot{M}_{hx}\dot{M}_{gg} \\
&= N^{-1/2}\sum_{i\in\mathbb{U}_N}\sum_{j\in\mathbb{A}_{\ell_i}}\omega_{ij}O(N^{-1/2})O(B_N^{-1}) \\
&= O(N^{-1}B_N^{-1})\sum_{i\in\mathbb{U}_N}\sum_{j\in\mathbb{A}_{\ell_i}}\omega_{ij} \\
&= O(B_N^{-1}), \qquad (3.4.26)
\end{aligned}$$

since $\sum_{j\in\mathbb{A}_{\ell_i}}\omega_{ij}=1$. By A.2a, $B_N=O(N^{\alpha})\to\infty$ as $N\to\infty$, which implies that (3.4.20) converges to zero in L_1. Then, conditional on $\mathbb{A}_N$ and $|\widehat{\beta}-\beta|<\eta$, (3.4.20) converges to zero in probability, and we can find $N^*_{\lambda\epsilon}$ such that for any $N>N^*_{\lambda\epsilon}$,

$$P\big(N^{-1/2}\sum_{i\in\mathbb{U}_N}\sum_{j\in\mathbb{A}_{\ell_i}}\omega_{ij}|\Delta_{ij}|>\lambda \;\big|\; \mathbb{A}_N, |\widehat{\beta}-\beta|<\eta\big)<\epsilon/2. \qquad (3.4.27)$$

By (3.4.20), $N^{1/2}\Big|\widehat{F}_L^{fi}(\dot{y})-\widetilde{F}_{L\beta}^{fi}(\dot{y})\Big|\leqslant N^{-1/2}\sum_{i\in\mathbb{U}_N}\sum_{j\in\mathbb{A}_{\ell_i}}\omega_{ij}\Big|\Delta_{ij}\Big|$. Then, the occurrence of the event $D_N=\Big\{N^{1/2}\Big|\widehat{F}_L^{fi}(\dot{y})-\widetilde{F}_{L\beta}^{fi}(\dot{y})\Big|>\lambda\Big\}$ implies the occurrence of $\Big\{N^{-1/2}\sum_{i\in\mathbb{U}_N}\sum_{j\in\mathbb{A}_{\ell_i}}\omega_{ij}\Big|\Delta_{ij}\Big|>\lambda\Big\}$. That is,

$$D_N=\Big\{N^{1/2}\Big|\widehat{F}_L^{fi}(\dot{y})-\widetilde{F}_{L\beta}^{fi}(\dot{y})\Big|>\lambda\Big\}\subseteq\Big\{N^{-1/2}\sum_{i\in\mathbb{U}_N}\sum_{j\in\mathbb{A}_{\ell_i}}\omega_{ij}\Big|\Delta_{ij}\Big|>\lambda\Big\},$$

and

$$\mathrm{P}\big(D_N \;\big|\; \mathbb{A}_N, |\widehat{\beta}-\beta|<\eta\big)\leqslant\mathrm{P}\big(N^{-1/2}\sum_{i\in\mathbb{U}_N}\sum_{j\in\mathbb{A}_{\ell_i}}\omega_{ij}\Big|\Delta_{ij}\Big|>\lambda \;\big|\; \mathbb{A}_N, |\widehat{\beta}-\beta|<\eta\big)<\epsilon/2 \qquad (3.4.28)$$

for $N>N^*_{\lambda\epsilon}$ by (3.4.27).

For $N > \max(N_{\eta\epsilon}, N^*_{\lambda\epsilon})$ then, we have that both (3.4.18) and (3.4.28) hold. Then for any $N > \max(N_{\eta\epsilon}, N^*_{\lambda\epsilon})$

$$P\big(N^{1/2}|\widehat{F}^{fi}_{L}(\dot{y}) - \widetilde{F}^{fi}_{L\beta}(\dot{y})| > \lambda \;\big|\; \mathbb{A}_N\big) = P\big(D_N \;\big|\; \mathbb{A}_N\big) < \epsilon.$$

Hence,

$$N^{-1/2}\Big|\widehat{F}^{fi}_{L}(\dot{y}) - \widetilde{F}^{fi}_{L\beta}(\dot{y})\Big| \to 0$$

in probability.

Part *(b)*. We can write each of the terms of the sequence

$$\Big\{V\big(\widetilde{F}^{fi}_{L\beta}(\dot{y}) \mid \mathbb{A}_N\big)\Big\}^{-1/2}\Big\{\widehat{F}^{fi}_{L}(\dot{y}) - F(\dot{y})\Big\}$$

as

$$\begin{aligned}\Big\{V\big(\widetilde{F}^{fi}_{L\beta}(\dot{y}) \mid \mathbb{A}_N\big)\Big\}^{-1/2}\Big\{\widehat{F}^{fi}_{L}(\dot{y}) - F(\dot{y})\Big\} &= \\ &= \Big\{V\big(\widetilde{F}^{fi}_{L\beta}(\dot{y}) \mid \mathbb{A}_N\big)\Big\}^{-1/2}\Big\{\widehat{F}^{fi}_{L}(\dot{y}) - \widetilde{F}^{fi}_{L\beta}(\dot{y})\Big\} + \\ &+ \Big\{V\big(\widetilde{F}^{fi}_{L\beta}(\dot{y}) \mid \mathbb{A}_N\big)\Big\}^{-1/2}\Big\{\widetilde{F}^{fi}_{L\beta}(\dot{y}) - F(\dot{y})\Big\}. \qquad (3.4.29)\end{aligned}$$

In part *(a)* we showed that $N^{1/2}\Big\{\widehat{F}^{fi}_{L}(\dot{y}) - \widetilde{F}^{fi}_{L\beta}(\dot{y})\Big\}$ converges to zero in probability as $N \to \infty$. By (3.4.14) we have that $V\big(\widetilde{F}^{fi}_{L\beta}(\dot{y}) \;\big|\; \mathbb{A}_N\big)$ is $O(N^{-1})$. Then the first term on the right hand side of (3.4.29),

$$\Big\{V\big(\widetilde{F}^{fi}_{L\beta}(\dot{y}) \mid \mathbb{A}_N\big)\Big\}^{-1/2}\Big\{\widehat{F}^{fi}_{L}(\dot{y}) - \widetilde{F}^{fi}_{L\beta}(\dot{y})\Big\}$$

converges to zero in probability. By Slutsky's theorem, we have that

$$\Big\{V\big(\widetilde{F}^{fi}_{L\beta}(\dot{y}) \mid \mathbb{A}_N\big)\Big\}^{-1/2}\Big\{\widehat{F}^{fi}_{L}(\dot{y}) - F(\dot{y})\Big\}$$

and

$$\Big\{V\big(\widetilde{F}^{fi}_{L\beta}(\dot{y}) \mid \mathbb{A}_N\big)\Big\}^{-1/2}\Big\{\widetilde{F}^{fi}_{L\beta}(\dot{y}) - F(\dot{y})\Big\}$$

have the same asymptotic distribution given by part *(b)* of Theorem 3.4.1. ▲

4 MONTE CARLO RESULTS

A Monte Carlo simulation was conducted to study the performance of the local-residuals estimator and the performance of the variance estimators introduced in Section 3.3. The superpopulation models used for generating the data and the model used in the construction of the local-residuals estimator are presented in Section 4.1. The results from the Monte Carlo simulation and a description of the methodology used in the reported Monte Carlo estimates is presented in Section 4.2. Comments about the performance of the local-residuals estimator are included in Section 4.3.

4.1 Superpopulation models

Three superpopulation models are considered. This set of models is by no means intended to be an exhaustive list of real situations, but represents different types of situations that we may encounter. The models are:

Model 1: "Correct" model. Data are generated using the model specified in the construction of the Chambers and Dunstan estimator with the distribution of the x-values skewed. For $i = 1, \ldots, N$:

- x_i is generated from a Chi-square distribution with 3 degrees of freedom,
- u_i is generated from a standard normal distribution,

- the value for y_i is computed as $y_i = \max(0.01,\ 2 + x_i + u_i)$. In most finite populations selected, all of the y values are strictly larger than 0.01.

Model 2: "Heteroscedastic" model. Data are generated using the model with increasing variance introduced by Hansen, Madow and Tepping (1983). For $i = 1, \dots, N$:

- x_i is generated from a Gamma distribution with density

$$f(x) = .04x \exp(-x/5),$$

- y_i given x_i is generated from a Gamma distribution with density

$$f(y \mid x) = [b^c \Gamma(c)]^{-1} y^{c-1} \exp(-y/b),$$

  where $b = 1.25x^{3/2}(8+5x)^{-1}$ and $c = .04x^{-3/2}(8+5x)^2$.

- Model 2 can be written as

$$Y_i = .4 + .25x + .25x^{3/4} U_i,$$

  where the U_i are independent identically distributed random variables with expected value zero and variance equal to one.

Model 3: Model with quadratic mean. Data are generated using a quadratic relation between y and x. For $i = 1, \dots, N$:

- x_i is generated from a Uniform(0,10) distribution,
- u_i is generated from a Uniform(-0.5, 0.5) distribution,
- y_i is set equal to $y_i = 5 + 0.2(x_i - 5)^2 + u_i$.

The selected and nonselected points for a simple random sample of size 60 from a finite population of size 600 for Models 1, 2 and 3 are presented in Figures 4.1, 4.2

and 4.3 respectively. The fitted regression line for Model 1, from which residuals are computed, is also included in the figures.

The average R^2 for samples of size 60 from Models 1, 2 and 3 are 0.861, 0.226 and 0.025 respectively, where $R^2 = \big\{\sum_{j\in\mathbb{A}}(\widehat{y}_j - \bar{y})^2\big\}\big\{\sum_{j\in\mathbb{A}}(y_j - \bar{y})^2\big\}^{-1}$, $\bar{y} = n^{-1}\sum_{j\in\mathbb{A}} y_j$, and $\widehat{y}_j$ is computed using the ordinary least squares estimators for β_0 and β_1 as $\widehat{y}_j = \widehat{\beta}_0 + \widehat{\beta}_1 x_j$.

4.2 Methodology

4.2.1 Sample sizes

Two finite population sizes, $N = 600$ and $N = 1200$, are considered in the Monte Carlo study. A single set of auxiliary variables $x_1, \dots, x_N$ was generated for each model and used in all Monte Carlo iterations. In each Monte Carlo iteration a new set of $y_1, \dots, y_N$ is generated and a simple random sample without replacement of n units is selected. The sample sizes considered are $n = 60$ for the population of $N = 600$ and $n = 120$ for the population of size $N = 1200$.

4.2.2 Selection of the number of bins

In this section we will consider the problem of selecting the number of bins, B, used to construct the local-residuals estimator. Intuitively, if we believe that the model with a single conditional density adequately represents the data, we would select one bin. Conversely, if we want to be conservative against model misspecification, a larger value of B_N should be selected.

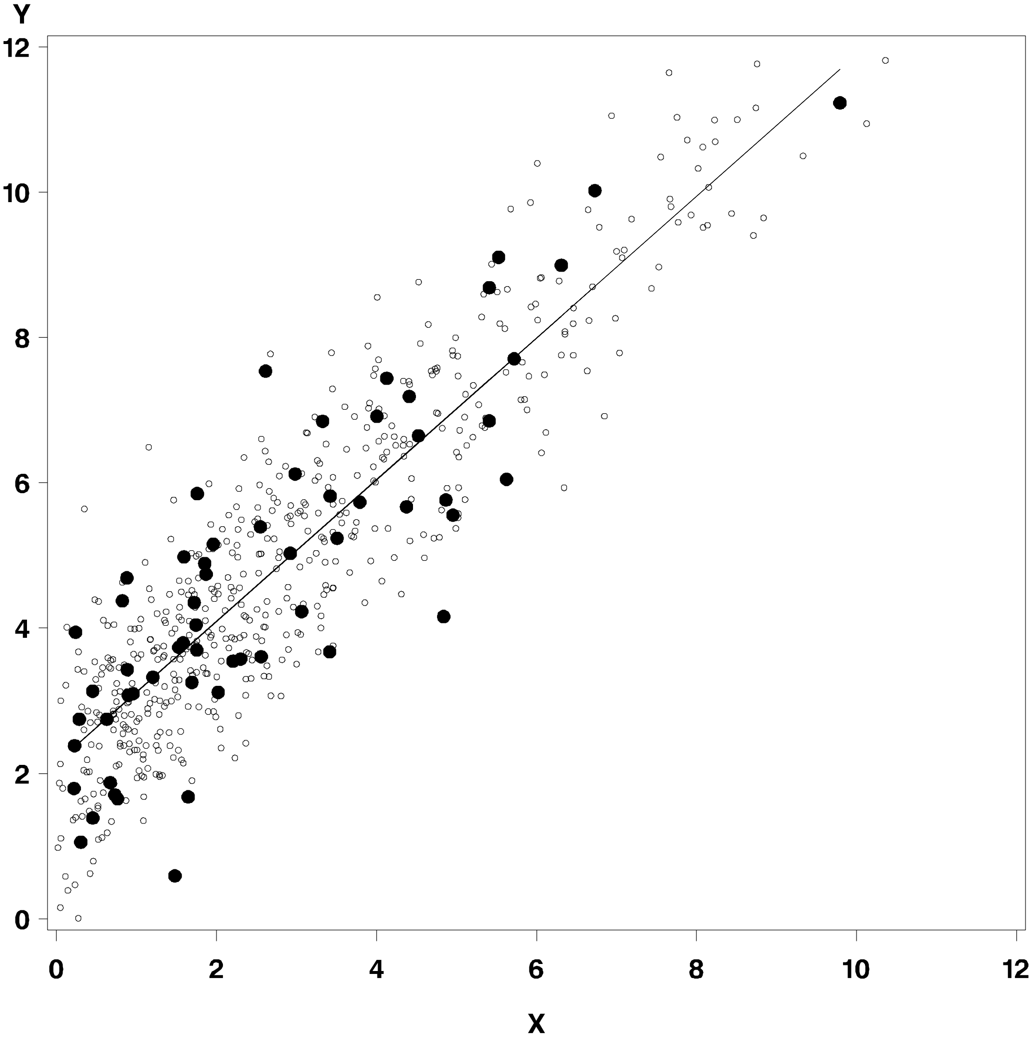

Figure 4.1 Plot of y against x and estimated regression line for a sample of size 60 from a population of size 600 generated by Model 1. Sample=●, nonsample=○

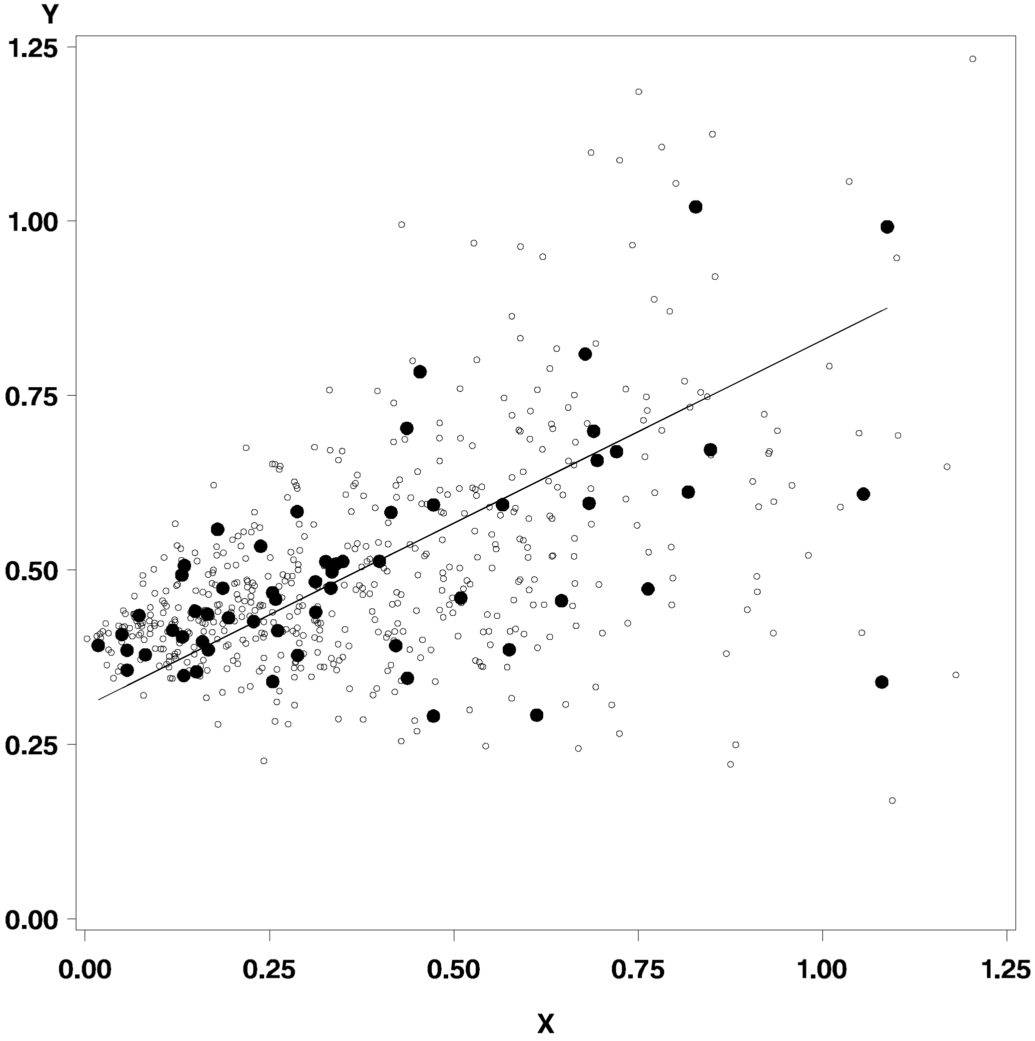

Figure 4.2 Plot of y against x and estimated regression line for a sample of size 60 from a population of size 600 generated by Model 2. Sample=●, nonsample=○

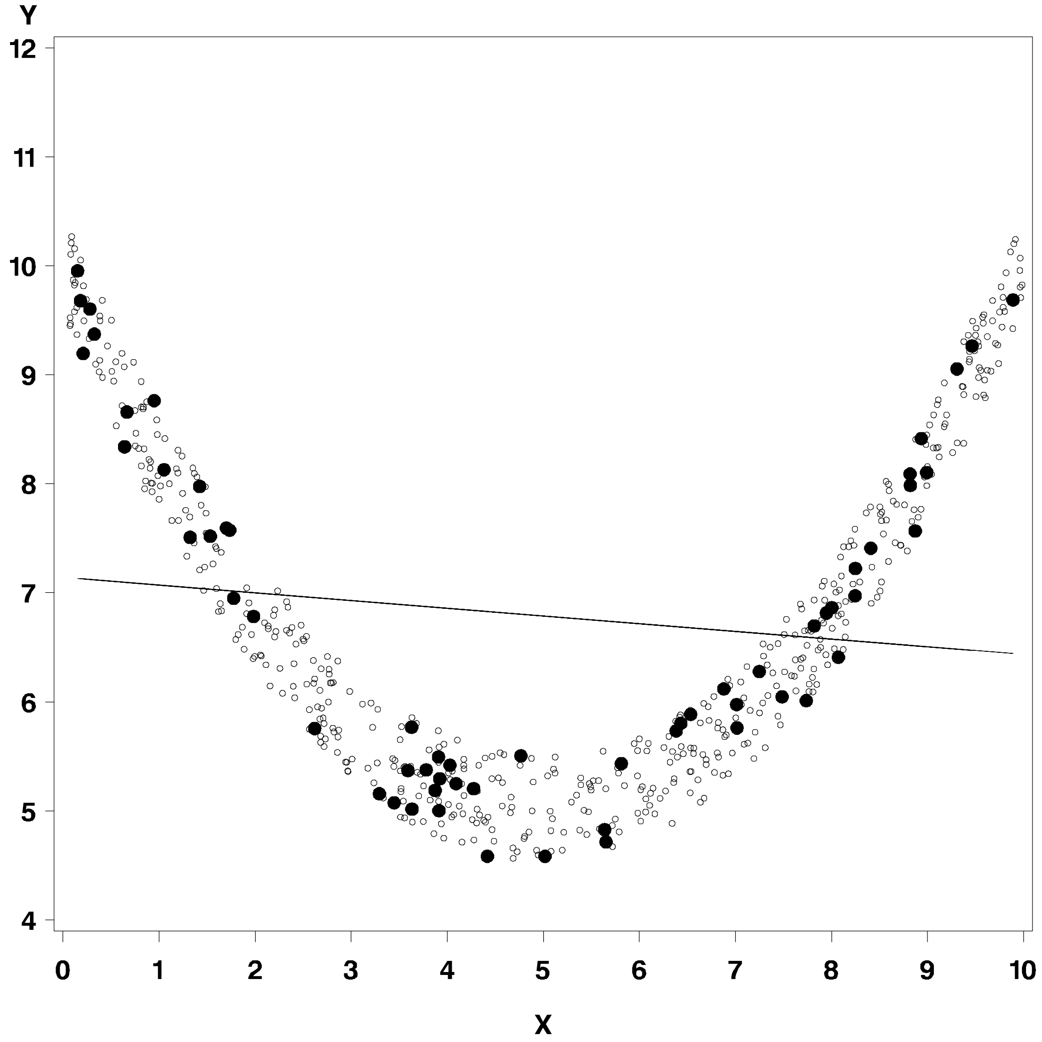

Figure 4.3 Plot of y against x and estimated regression line for a sample of size 60 from a population of size 600 generated by Model 3. Sample=●, nonsample=○

Let $\widehat{\beta}$ be the least squares estimator of β defined in (3.1.5), and let

$$\widehat{u}_j = y_j - x_j\widehat{\beta}, \tag{4.2.1}$$

be the observed residuals for $j \in \mathbb{A}_N$. Under the Chambers and Dunstan model,

$$U_j = Y_j - x_j\beta \tag{4.2.2}$$

are independent and identically distributed random variables. The local-residuals estimator is constructed under the assumption that if x_i is "close" to x_j, then the distribution of U_i is "close" to that of U_j. In both cases, the observed residuals (4.2.1) are used to approximate the distribution function of U_j.

We use crossvalidation to determine the number of bins to use in constructing the local-residuals estimator. Crossvalidation for a sample of size k uses the k possible samples of size $k-1$ to predict the omitted element. Let unit α from bin ℓ_α be the omitted element. For unit α, let

$$\widehat{G}_{\ell_\alpha[-\alpha]}(\dot{u}) = \sum_{j\in\mathbb{A}^*_{\ell_\alpha}} \omega^*_{j[-\alpha]} I(\widehat{u}_j \leqslant \dot{u}) \tag{4.2.3}$$

be the local estimator of the distribution function of the residuals evaluated at $\dot{u}$, where $\mathbb{A}^*_{\ell_\alpha} = \mathbb{A}_{\ell_\alpha} - \alpha$ is the reduced sample in bin ℓ_α after removing unit α, and the weights $\omega^*_{j[-\alpha]} = \pi_j^{-1}\left[\sum_{j'\in\mathbb{A}^*_{\ell_\alpha}} \pi_{j'}^{-1}\right]^{-1}$ are the adjusted sampling weights for the remaining units in bin $\mathbb{A}_{\ell_\alpha}$. We construct a measure of how good $\widehat{G}_{\ell_\alpha[-\alpha]}(\dot{u})$ is as a predictor of $\mathrm{P}(U_\alpha \leqslant \dot{u})$ in order to choose an "optimal" number of bins. To do this, we select 20 values of $\dot{u}$ and evaluate both $\widehat{G}_{\ell_\alpha[-\alpha]}(\dot{u})$ and $I(\widehat{u}_\alpha \leqslant \dot{u})$ at these 20 values. Let $\widehat{u}_{(j_o)}$ for $j_o = 1, \ldots, n$ be the sorted values of $\widehat{u}_j$ for $j \in \mathbb{A}_N$. Let $\dot{u}_\gamma = \widehat{u}_{\lfloor \gamma n/21 \rfloor}$ for $\gamma = 1, \ldots, 20$, be the selected values of $\dot{u}$, where the function $\lfloor x \rfloor$ is the largest integer that is less than or equal to x.

Divisions of the sample into $B = 1, 2, \ldots, B_c$ bins are constructed, as described in (3.1.3), where B_c is the maximum number of bins considered. The value of B_c was initially set at 20. For Model 3, the value of B_c was changed to $B_c = 30$ for samples of size $n = 60$ and to $B_c = 35$ for samples of size $n = 120$. Then, for each B, we compute

$$C(B) = \sum_{\alpha \in \mathbb{A}_N} \sum_{\gamma=1}^{20} \Big(I(\widehat{u}_\alpha \leqslant \dot{u}_\gamma) - \widehat{G}_{\ell_{\alpha[-\alpha]}}(\dot{u}_\gamma) \Big)^2 \Big(\widehat{G}(\dot{u}_\gamma) \big[1 - \widehat{G}(\dot{u}_\gamma)\big] \Big)^{-1}, \tag{4.2.4}$$

where $\widehat{G}(\dot{u}_\gamma) = \big\{ \sum_{j \in \mathbb{A}_N} \pi_j^{-1} \big\}^{-1} \sum_{j \in \mathbb{A}_N} \pi_j^{-1} I(\widehat{u}_j \leqslant \dot{u}_\gamma)$. The criterion is an approximation to the mean integrated square error defined in (2.4.5). Let B_{min} be the value of B that minimizes $C(B)$. The number $B = B_{min}$ of bins is used to construct the local-residuals estimator. In some cases the value of n/B is not an integer. The n sample units are assigned to bins as follows. Let $x_{(j_o)}$ for $j_o = 1, \ldots, n$ be the sorted values of x_j for $j \in \mathbb{A}_N$. For $j_o = 1, \ldots, n$, unit j_o is assigned to bin $\ell = 1 + \lfloor (j_o - 1)(B/n) \rfloor$, where the function $\lfloor x \rfloor$ is the largest integer that is less than or equal to x.

4.2.3 Calculation of Monte Carlo means and variances

The reported estimates are means, and functions of means, from M Monte Carlo iterations. Let $a_{(t)}$ and $b_{(t)}$ be two quantities computed in Monte Carlo iteration t for $t = 1, \ldots, M$. The four types of estimates are:

a. the mean of $a_{(t)}$, $a_{(\cdot)} = M^{-1} \sum_{t=1}^{M} a_{(t)}$,

b. the square root of a mean, $(a_{(\cdot)})^{1/2}$,

c. the ratio between two means $r = a_{(\cdot)}(b_{(\cdot)})^{-1}$,

d. the square root of a ratio $r^{1/2}$.

The formulas used to approximate the variances of the estimators of items a. through d. are presented in Table 4.1.

Table 4.1 Formulas for the Monte Carlo estimators and Monte Carlo variances for the quantities presented in Tables 4.5 through 4.32

Item reported	Estimator	Estimated variance of estimator
a. Mean	$a_{(\cdot)} = M^{-1}\sum_{t=1}^{M} a_{(t)}$	$M^{-2}\sum_{t=1}^{M}(a_{(t)} - a_{(\cdot)})^2$
b. Square root of mean	$(a_{(\cdot)})^{1/2}$	$4^{-1}(a_{(\cdot)})^{-1}M^{-2}\sum_{t=1}^{M}(a_{(t)} - a_{(\cdot)})^2$
c. Ratio of means	$r = a_{(\cdot)}(b_{(\cdot)})^{-1}$	$(b_{(\cdot)})^{-2}M^{-2}\sum_{t=1}^{M}(a_{(t)} - rb_{(t)})^2$
d. Square root of ratio	$r^{1/2}$	$4^{-1}r^{-1}(b_{(\cdot)})^{-2}M^{-2}\sum_{t=1}^{M}(a_{(t)} - rb_{(t)})^2$

4.3 Monte Carlo results

The estimators presented in Tables 4.5 through 4.32 are: local-residuals estimator (3.1.4), Chambers and Dunstan estimator (2.3.4), Rao, Kovar and Mantel estimator (2.3.13), a poststratified estimator defined below in (4.3.1), and the Horvitz-Thompson estimator (2.2.1).

For each Monte Carlo iteration, the finite population distribution function and the estimators mentioned above were calculated at seven points $\dot{y}$ that represent the 5th, 10th, 25th, 50th, 75th, 90th and 95th percentiles of the superpopulation distribution function $F(\dot{y})$ defined in (2.1.5). The superpopulation quantiles were estimated by generating 5000 finite populations from the corresponding models and computing the corresponding sample quantiles from the resulting $N \times 5000$ values of y.

Two local-residuals estimators are considered, one based on the B_{min} bins selected by the crossvalidation procedure that minimizes (4.2.4), identified as Local-residuals (c-val) in Tables 4.5 through 4.20, and another based on a fixed number of bins. Let $\widehat{F}_L^{min}(\dot{y})$ denote the local-residuals estimator computed with B_{min} bins and $\widehat{F}_L(\dot{y})$ denote the local-residuals estimator computed with a fixed number of bins. For the population of size 600 we considered $B = 6$ bins, and for the population of size 1200 we considered $B = 10$ bins. The number of elements per bin, k_N, is equal to: $k_{600} = 10$ and $k_{1200} = 12$. According to the assumption of $\alpha > 0.5$ for the value of α in A.2a, we increased the number of bins more than the number of elements per bin when considering the larger population with $N = 1200$.

The poststratified estimator that appears in Tables 4.5 through 4.32 is constructed as the average of two poststratified estimators: $\widehat{F}^{[1]}_{ps}(\dot{y})$ and $\widehat{F}^{[2]}_{ps}(\dot{y})$. The first poststratified estimator, $\widehat{F}^{[1]}_{ps}(\dot{y})$, is constructed as follows: divide the population of size N in B_N poststrata of equal size, $N^{[1]}_h = NB_N^{-1}$ for $h = 1, \dots, B_N$. The number of poststrata is the same as the number of bins used for the local-residuals estimator computed with a fixed number of bins. For the population of size 600, $B_{600} = 6$ and for the population of size 1200, $B_{1200} = 10$. The sample is assigned to the strata and the stratum sample sizes, n_h, are computed for $h = 1, \dots, B_N$. Then, $\widehat{F}^{[1]}_{ps}(\dot{y})$ is defined as

$$\widehat{F}^{[1]}_{ps}(\dot{y}) = N^{-1} \sum_{h=1}^{B_N} N^{[1]}_h n_h^{-1} \sum_{j \in \mathbb{A}_h} I\big(y_j \leqslant \dot{y}\big),$$

where $\mathbb{A}_h$ is the part of the sample that falls into stratum h. To compute the second poststratified estimator, $\widehat{F}^{[2]}_{ps}(\dot{y})$, the population is divided into $B_N + 1$ poststrata. The stratum sizes, $N^{[2]}_h$, for the first and last strata are equal to $2^{-1}NB_N^{-1}$, and the $N^{[2]}_h$ for the remaining strata are equal to $N^{[2]}_h = NB_N^{-1}$ for $h = 2, \dots, B_N$. Then,

$$\widehat{F}^{[2]}_{ps}(\dot{y}) = N^{-1} \sum_{h=1}^{B_N+1} N^{[2]}_h n_h^{-1} \sum_{j \in \mathbb{A}_h} I\big(y_j \leqslant \dot{y}\big).$$

The poststratified estimator included in the tables is

$$\widehat{F}_{ps}(\dot{y}) = 2^{-1}\big[\widehat{F}^{[2]}_{ps}(\dot{y}) + \widehat{F}^{[2]}_{ps}(\dot{y})\big]. \tag{4.3.1}$$

For either $\widehat{F}^{[1]}_{ps}(\dot{y})$ or $\widehat{F}^{[2]}_{ps}(\dot{y})$, strata are collapsed when one or more of the n_h are zero. If the first or the last stratum is empty, the empty stratum is collapsed with the contiguous stratum. If one of the middle strata is empty, say stratum h, the corresponding N_h is divided by 2 and $2^{-1}N_h$ units are added to strata $h-1$ and $h+1$. If more than one stratum is empty, then the N units are reclassified into 2 poststrata with $2^{-1}N$ units each.

Table 4.2 presents the average number of bins selected by the crossvalidation procedure described in Section 4.2.2 for Model 1, Model 2 and Model 3, and for sample sizes of 60 and 120. An estimate of the distribution of B_{min} for Model 1, Model 2 and Model 3 is presented in Tables 4.3 and 4.4 for the populations of size 600 and 1200 respectively. The columns corresponding to Model 1 show that almost 90% of the time the procedure selects $B_{min} = 1$ for samples of size 60 and 120. Note that when $B_{min} = 1$, the local-residuals estimator and the Chambers and Dunstan estimator are equal under simple random sampling. For Model 2, the crossvalidation procedure tends to select values of B_{min} larger than the ones selected for Model 1. The values of B_{min} are larger for the sample size of 120 than for the sample size of 60. For Model 3, it is clear from Table 4.3 and 4.4 that the procedure selects larger values of B_{min} than for other models. For Model 3, with $B_c = 35$ and a sample size of 120, the average B_{min} is 19.91, with a standard error of 0.10.

Table 4.2 Average number of bins selected by the crossvalidation procedure for alternative models and sample sizes of 60 and 120. $B_c = 20$ for Model 1 and Model 2; $B_c = 30$ for Model 3, n=60; $B_c = 35$ for Model 3, n=120; 10000 iterations for Model 1 and Model 2, 2500 iterations for Model 3

Sample size	Model 1	Model 2	Model 3
$n = 60$	1.310	2.397	13.571
	(0.011)	(0.018)	(0.084)
$n = 120$	1.345	3.171	19.912
	(0.011)	(0.021)	(0.104)

Tables 4.5 through 4.10 present the estimated bias in the estimated finite population distribution function of the estimators considered. In the following discussion, we use $\widehat{F}(\dot{y})$ to denote any of the six estimators presented in the tables: $\widehat{F}_L^{min}(\dot{y})$ defined in (3.1.4) with the number of bins selected by the procedure described in Section 4.2.2, $\widehat{F}_L(\dot{y})$ defined in (3.1.4), $\widehat{F}_{CD}(\dot{y})$ defined in (2.3.4), $\widehat{F}_{RKMdm}(\dot{y})$ defined in (2.3.13), $\widehat{F}_{ps}(\dot{y})$ defined in (4.3.1) and $\widehat{F}_{HT}(\dot{y})$ defined in (2.2.1). Let $\widehat{F}_{(t)}(\dot{y})$ and $F_{N(t)}(\dot{y})$ be the values of the estimator $\widehat{F}(\dot{y})$ and of the finite population distribution function at the point $\dot{y}$ in iteration t for $t = 1, \ldots, M$. The estimated bias of an estimator is computed as $M^{-1}\sum_{t=1}^{M}\left\{\widehat{F}_{(t)}(\dot{y}) - F_{N(t)}(\dot{y})\right\}$. Tables 4.5 and 4.8 present the bias for Model 1 for $n = 60$ and $n = 120$ respectively. Tables 4.5 and 4.8 show that when the model is correctly specified none of the estimators has noticeable bias for the parameter values investigated. Although none of the biases represented in Tables 4.5 and 4.8 exceeds 0.2 percent points, more than 5% of the entries are significantly different from zero for a test of level 5%. We discuss several explanations for the possible bias.

The local-residuals estimator for a fixed number of bins and the Chambers and Dunstan estimator are model unbiased when computed with the true β. In Tables 4.5 and 4.8 the local-residuals estimator and the Chambers and Dunstan estimator are computed using $\widehat{\beta}$ defined in (3.1.5). The Rao, Kovar and Mantel estimator is asymptotically unbiased, but for a sample of size 60 may be biased. Note that none of the entries in Table 4.8 is significantly different from zero for the local-residuals estimators, for the Chambers and Dunstan estimator and for the Rao, Kovar and Mantel estimator. The bias in the poststratified estimator may be due to the collapsing algorithm described above. We could have used the method of collapsing strata presented in Fuller (1966) to obtain an unbiased estimator.

The Horvitz-Thompson estimator is conditionally biased given the sample indexes, but unbiased when averaging over all possible simple random samples. Although in

Table 4.5 there are two out of the seven entries for the Horvitz-Thompson estimator that are significantly different from zero, none of the entries in the next five tables is significant at a 5% level. Thus, out of 42 entries, we have approximately 5% that are significantly different from zero.

Tables 4.6 and 4.9 give the bias of the estimators for Model 2, for sample sizes of 60 and 120, respectively. The Chambers and Dunstan estimator $\widehat{F}_{CD}(\dot{y})$ is the most affected by misspecification of the variance function, followed by the local-residuals estimator computed with B_{min} bins. For a sample size of 60, the Chambers and Dunstan estimator is overestimating the finite population distribution function by as much as 3.987 percent points for the 5th superpopulation quantile. The bias in the Chambers and Dunstan estimator does not decrease when the sample size increases to 120. For a sample of size 120 the finite population distribution function is overestimated by 3.941 percent points for the 5th superpopulation quantile.

The bias for the local-residuals estimator computed with B_{min} bins does decrease when the sample size increases, due to the fact that the number of bins selected by the crossvalidation procedure are larger for samples of size 120 than for samples of size 60. The bias in the local-residuals estimator with six bins is less than half a percent point for the seven quantiles considered when the sample size is 60. When the sample size is 120, the biases of the local-residuals estimator computed with a fixed number of bins for the quantiles investigated are less than 0.3%. The reduction in bias is proportional to the increase in sample size as the bins lengths decrease with the sample size.

The three other estimators considered, $\widehat{F}_{RKMdm}(\dot{y})$, $\widehat{F}_{ps}(\dot{y})$ and $\widehat{F}_{HT}(\dot{y})$, are robust to heterogeneous variances. For both sample sizes, 60 and 120, the biases in the estimators $\widehat{F}_{RKMdm}(\dot{y})$, $\widehat{F}_{ps}(\dot{y})$ and $\widehat{F}_{HT}(\dot{y})$ are essentially zero. Thus, under misspecification of the model variance function, the local-residuals estimator with fixed number of bins and the

$\widehat{F}_{RKMdm}(\dot{y})$ exhibit superior performance with respect to the bias criterion to that of the Chambers and Dunstan estimator. Also, the biases in the local-residuals estimator computed with B_{min} bins for Model 2, although somewhat larger, are still significantly smaller than those of the Chambers and Dunstan estimator.

Tables 4.7 and 4.10 give the bias of the estimators under Model 3, that is, when the mean function of the model has been misspecified. In this case, the poststratified estimator and the Horvitz-Thompson estimator have negligible bias. The biases of the other four estimators are functions of the quantiles and the sample size, with the local-residuals estimators with fixed B, in general, having smaller bias than the three estimators $\widehat{F}_L^{min}(\dot{y})$, $\widehat{F}_{CD}(\dot{y})$ and $\widehat{F}_{RKMdm}(\dot{y})$. For lower quantiles the local-residuals estimator with fixed B has negligible bias, whereas the biases of the local-residuals with B_{min} bins estimator, Chambers and Dunstan estimator and Rao, Kovar and Mantel estimator are significant. For upper quantiles, all estimators have comparable biases.

Tables 4.11 through 4.14 show the contribution of the bias to the mean square error: $\left\{E\left[\widehat{F}(\dot{y})\right] - F_N(\dot{y})\right\}^2 \left\{E\left[\widehat{F}(\dot{y}) - F_N(\dot{y})\right]^2\right\}^{-1}$, for Models 2 and 3 for the local-residuals estimators $\widehat{F}_L^{min}(\dot{y})$ and $\widehat{F}_L(\dot{y})$, and for the Chambers and Dunstan estimator. For the Chambers and Dunstan estimator and for the local-residuals estimator $\widehat{F}_L^{min}(\dot{y})$ the bias makes an important contribution to the mean square error. The bias contribution to the mean square error for the local-residuals estimator with fixed B is negligible when either the mean or the variance have been misspecified. The bias in the local-residuals estimator with B_{min} bins makes a significant contribution to the mean square error when the variance function has been misspecified, but the contributions are much smaller in magnitude than those of the Chambers and Dunstan estimator. For Model 2 and for Model 3, the bias contribution to the mean square error for the Chambers and Dunstan estimator does not decrease when the sample size increases.

Tables 4.15 through 4.20 show the square root of the mean square error for the Horvitz-Thompson estimator and the ratios of the root mean square errors of the other estimators to the root mean square error of the Horvitz-Thompson estimator for Models 1, 2 and 3 and sample sizes of 60 and 120. When the model is correctly specified, Tables 4.15 and 4.18, the Chambers and Dunstan estimator has the smallest mean square errors for the seven quantiles investigated for both sample sizes. The local-residuals estimator with B_{min} bins is the second best with respect to the mean square error criterion, followed by the local-residuals estimator with fixed B. The root mean square error of estimator $\widehat{F}_L(\dot{y})$ for the superpopulation median is about 60% of the root mean square error of the Horvitz-Thompson estimator for samples of size 60 or 120.

For Model 2, Tables 4.16 and 4.19 indicate that the local-residuals estimator with fixed B, in general, has smaller root mean square error than the other estimators. The Chambers and Dunstan estimator is the estimator most affected by the variance misspecification of Model 2 for the sample size of 60. The local-residuals estimator with B_{min} bins performs uniformly better than the Chambers and Dunstan estimator.

For Model 3 and for a sample size of 60, Table 4.17 shows that both local-residuals estimators perform better than $\widehat{F}_{CD}(\dot{y})$, $\widehat{F}_{RKMdm}(\dot{y})$ and $\widehat{F}_{HT}(\dot{y})$, and the performance of the local-residuals estimators is similar to the performance of $\widehat{F}_{ps}(\dot{y})$. For Model 3 and for a sample size of 120, the performances of the estimators $\widehat{F}_L^{min}(\dot{y})$, $\widehat{F}_L(\dot{y})$, $\widehat{F}_{RKMdm}(\dot{y})$ and $\widehat{F}_{ps}(\dot{y})$ are similar. The local-residuals estimator with fixed B has uniformly smaller root mean square error than the other estimators for the seven superpopulation quantiles considered.

We can summarize the results from Tables 4.5 through 4.10, referring to the bias of the estimators, and the results from Tables 4.15 through 4.20, referring to the root mean square error, as follows:

- the local-residuals estimator with fixed B is robust in terms of bias against departures from the superpopulation model used to construct the estimator,
- the local-residuals estimator with B_{min} bins is less sensitive to misspecification than the Chambers and Dunstan estimator,
- the procedure that selects the number of bins seems to select too few bins for Model 2, and too many bins for Model 3,
- the local-residuals estimator with fixed B has in general smaller mean square error than the Rao, Kovar and Mantel estimator, the poststratified estimator and the Horvitz-Thompson estimator,
- the performance of both local-residuals estimators is superior to the performance of the Chambers and Dunstan estimator when the model is incorrectly specified.

We study the estimation of the variance for the local-residuals estimator with a fixed number of bins. The remaining tables, Table 4.21 through Table 4.32, are related to the performance of the variance estimators studied in Section 3.3 for the model variance of $\widehat{F}_L(\dot{y}) - F_N(\dot{y})$ and the model variance of $\widehat{F}_L(\dot{y})$ as an estimator of the superpopulation distribution function. The three estimators considered are $\widetilde{V}_{ee}(\dot{y})$ defined in (3.3.2), $\widetilde{V}_{\mathcal{L}}(\dot{y})$ defined in (3.3.10), and $\widetilde{V}_{JK}(\dot{y})$ defined in (3.3.11). Estimator $\widetilde{V}_{ee}(\dot{y})$ is an estimator of the variance of the finite population estimation error of the local-residuals estimator, $\widehat{F}_L(\dot{y}) - F_N(\dot{y})$. Estimators $\widetilde{V}_{\mathcal{L}}(\dot{y})$ and $\widetilde{V}_{JK}(\dot{y})$ are used to estimate the variance of the error of the local-residuals estimator $\widehat{F}_L(\dot{y})$ as an estimator of the superpopulation distribution function, $\widehat{F}_L(\dot{y}) - F(\dot{y})$, where $\widehat{F}_L(\dot{y})$ is defined in (3.1.4) and $F(\dot{y})$ is defined in (2.1.5). When the variance estimators are computed using the estimated $\widehat{\beta}$, we use the notation $\widehat{V}_{ee}(\dot{y})$, $\widehat{V}_{\mathcal{L}}(\dot{y})$, and $\widehat{V}_{JK}(\dot{y})$, for the three estimators. The two versions of the jackknife estimator (3.3.11) that we mentioned in Section (3.3) are:

- $\widehat{V}_{JKwo}(\dot{y})$ for the version that uses the $\widehat{\beta}$ and the $\widehat{y}_{ij}$ computed from the sample $\mathbb{A}_N$,
- $\widehat{V}_{JK}(\dot{y})$ for the version that recomputes $\widehat{\beta}$ and $\widehat{y}_{ij}$ for each of the n reduced samples $\mathbb{A}_N - \{\alpha\}$.

From Tables 4.21 through 4.26 one can see that for all combination of models and sample sizes considered, the average of estimator $\widehat{V}_{ee}(\dot{y})$ and the average of estimator $\widehat{V}_{\mathcal{L}}(\dot{y})$ are, in general, very similar to the variance of $\widehat{F}_L(\dot{y}) - F_N(\dot{y})$ and to the variance of $\widehat{F}_L(\dot{y}) - F(\dot{y})$, respectively. The mean of the estimator $\widehat{V}_{JK}(\dot{y})$ is always greater than the mean of the estimator $\widehat{V}_{JKwo}(\dot{y})$, as may be expected from the fact that $\widehat{V}_{JK}(\dot{y})$ introduces additional variability due to the estimation of β and recalculation of the $\widehat{y}_{ij}$. Both Jackknife variances are greater, on average, than the variance computed using $\widehat{V}_{\mathcal{L}}(\dot{y})$.

In Tables 4.27 through 4.32 we present results on the estimated probability that a 95% confidence interval constructed with one of the four variance estimators will contain the percentile of the finite population distribution function. Since the sampling fraction is 0.10 in both populations, the variance of $\widehat{F}_L(\dot{y}) - F_N(\dot{y})$ is about 0.9 of the variance of $\widehat{F}_L(\dot{y}) - F(\dot{y})$. The three estimators of the model variance of $\widehat{F}_L(\dot{y}) - F(\dot{y})$ can be modified to obtain estimators of the model variance of $\widehat{F}_L(\dot{y}) - F_N(\dot{y})$ by multiplying $\widehat{V}_{\mathcal{L}}(\dot{y})$, $\widehat{V}_{JKwo}(\dot{y})$ and $\widehat{V}_{JK}(\dot{y})$ by $(1 - nN^{-1})$. The standardized estimators are likely to converge faster to the limiting normal variable in the middle part of the distribution than in the tails of the distribution. Thus, in finite samples, the confidence intervals for the finite population distribution function evaluated at quantiles towards the tails of the distribution constructed using the limiting normal theory are likely to have inferior coverage probabilities to those of the confidence intervals for the finite population distribution function for quantiles near the middle part of the distribution.

The coverage probabilities of the confidence intervals constructed with $\widehat{V}_{ee}(\dot{y})$ and with $0.9\widehat{V}_{\mathcal{L}}(\dot{y})$ are very similar in the six tables that represent the three models and two sample sizes for the seven quantiles considered. Since $\widehat{V}_{\mathcal{L}}(\dot{y})$ is much faster to compute than $\widehat{V}_{ee}(\dot{y})$, we can approximate $\widehat{V}_{ee}(\dot{y})$ by $(1-nN^{-1})\widehat{V}_{\mathcal{L}}(\dot{y})$ to reduce computing time. Due to the larger variances obtained for the jackknife estimators, the confidence intervals constructed with the jackknife estimators generally have coverage probabilities larger than those of the confidence intervals constructed with $\widehat{V}_{ee}(\dot{y})$ or $(1-nN^{-1})\widehat{V}_{\mathcal{L}}(\dot{y})$. For Model 1 and Model 2, the confidence intervals have probabilities close to 0.95 of including the finite population distribution function, especially for the middle part of the distribution. For Model 3, the confidence intervals for quantiles in the middle part of the distribution constructed with the four variance estimators, $\widehat{V}_{ee}(\dot{y})$, $\widehat{V}_{\mathcal{L}}(\dot{y})$, $\widehat{V}_{JKwo}(\dot{y})$ and $\widehat{V}_{JK}(\dot{y})$, tend to cover the percentiles of the finite population distribution function more than 95% of the time.

Table 4.3 Estimated distribution of the number of bins (B) selected by the cross-validation procedure for Model 1, Model 2 and Model 3 for a sample size of $n = 60$. $B_c = 20$, 2500 iterations for Model 1 and Model 2; $B_c = 30$, 10000 iterations for Model 3. Standard errors are smaller than 0.0047 for Model 1 and Model 2 and smaller than 0.0064 for Model 3

B	Model 1	Model 2	Model 3
1	0.8783	0.3312	0.0000
2	0.0500	0.3331	0.0000
3	0.0286	0.1844	0.0000
4	0.0164	0.0626	0.0000
5	0.0106	0.0398	0.0028
6	0.0052	0.0137	0.0088
7	0.0042	0.0132	0.0260
8	0.0022	0.0074	0.0424
9	0.0018	0.0046	0.0724
10	0.0008	0.0031	0.0908
11	0.0008	0.0019	0.1144
12	0.0004	0.0015	0.1012
13	0.0002	0.0008	0.0904
14	0.0001	0.0010	0.0820
15	0.0001	0.0008	0.0892
16	0.0000	0.0005	0.0700
17	0.0001	0.0001	0.0536
18	0.0000	0.0002	0.0340
19	0.0002	0.0001	0.0320
20	0.0000	0.0000	0.0188
21	0.0000	0.0000	0.0148
22	0.0000	0.0000	0.0168
23	0.0000	0.0000	0.0144
24	0.0000	0.0000	0.0092
25	0.0000	0.0000	0.0052
26	0.0000	0.0000	0.0028
27	0.0000	0.0000	0.0024
28	0.0000	0.0000	0.0020
29	0.0000	0.0000	0.0016
30	0.0000	0.0000	0.0020

Table 4.4 Estimated distribution of the number of bins (B) selected by the cross-validation procedure for Model 1, Model 2 and Model 3 for a sample size of $n = 120$. $B_c = 20$, 2500 iterations for Model 1 and Model 2; $B_c = 30$, 10000 iterations for Model 3. Standard errors are smaller than 0.0046 for Model 1 and Model 2, and smaller than 0.0057 for Model 3

B	Model 1	Model 2	Model 3
1	0.8734	0.1398	0.0000
2	0.0457	0.3085	0.0000
3	0.0322	0.2547	0.0000
4	0.0176	0.1201	0.0000
5	0.0109	0.0830	0.0000
6	0.0071	0.0299	0.0000
7	0.0051	0.0252	0.0000
8	0.0031	0.0121	0.0000
9	0.0015	0.0077	0.0008
10	0.0011	0.0063	0.0052
11	0.0007	0.0039	0.0240
12	0.0003	0.0026	0.0212
13	0.0003	0.0017	0.0424
14	0.0004	0.0010	0.0504
15	0.0002	0.0011	0.0704
16	0.0000	0.0006	0.0700
17	0.0003	0.0007	0.0592
18	0.0000	0.0003	0.0904
19	0.0001	0.0003	0.0744
20	0.0000	0.0005	0.0836
21	0.0000	0.0000	0.0808
22	0.0000	0.0000	0.0536
23	0.0000	0.0000	0.0448
24	0.0000	0.0000	0.0384
25	0.0000	0.0000	0.0404
26	0.0000	0.0000	0.0304
27	0.0000	0.0000	0.0316
28	0.0000	0.0000	0.0156
29	0.0000	0.0000	0.0156
30	0.0000	0.0000	0.0140
31	0.0000	0.0000	0.0136
32	0.0000	0.0000	0.0092
33	0.0000	0.0000	0.0092
34	0.0000	0.0000	0.0056
35	0.0000	0.0000	0.0052

Table 4.5 Estimated bias of alternative estimators of distribution function × 100 for Model 1. Standard error in parentheses. $N = 600$, $n = 60$; $10,000$ iterations

$F(\dot{y})$	Local-residuals (c-val)	Local-residuals ($B = 6$)	Chambers-Dunstan	Rao-Kovar Mantel	Post-stratified	Horvitz-Thompson
5%	0.046 (0.017)	0.051 (0.023)	0.022 (0.016)	0.065 (0.026)	0.051 (0.027)	0.057 (0.028)
10%	0.051 (0.024)	0.062 (0.031)	0.015 (0.022)	0.075 (0.034)	0.073 (0.035)	0.066 (0.038)
25%	0.016 (0.031)	0.004 (0.039)	-0.009 (0.029)	0.016 (0.043)	-0.008 (0.044)	0.005 (0.054)
50%	0.059 (0.027)	0.096 (0.037)	0.065 (0.025)	0.114 (0.042)	0.119 (0.043)	0.138 (0.063)
75%	0.018 (0.016)	-0.024 (0.024)	0.048 (0.014)	-0.054 (0.030)	-0.059 (0.032)	-0.031 (0.054)
90%	0.026 (0.009)	0.012 (0.011)	0.027 (0.009)	-0.008 (0.018)	0.016 (0.022)	-0.019 (0.037)
95%	-0.000 (0.008)	0.014 (0.009)	-0.006 (0.007)	0.009 (0.015)	0.020 (0.022)	-0.009 (0.027)

Table 4.6 Estimated bias of alternative estimators of distribution function × 100 for Model 2. Standard error in parentheses. $N = 600$, $n = 60$; $10,000$ iterations

$F(\dot{y})$	Local-residuals (c-val)	Local-residuals $(B = 6)$	Chambers-Dunstan	Rao-Kovar Mantel	Post-stratified	Horvitz-Thompson
5%	2.345	0.391	3.987	0.016	0.016	0.004
	(0.030)	(0.027)	(0.029)	(0.030)	(0.029)	(0.028)
10%	2.144	0.414	3.481	0.007	0.002	0.001
	(0.037)	(0.037)	(0.034)	(0.041)	(0.040)	(0.039)
25%	-0.449	-0.451	-0.930	-0.093	-0.080	-0.107
	(0.046)	(0.053)	(0.041)	(0.057)	(0.058)	(0.057)
50%	-2.658	-0.351	-6.129	-0.026	-0.031	-0.041
	(0.059)	(0.059)	(0.052)	(0.061)	(0.062)	(0.065)
75%	-1.885	-0.267	-3.790	0.026	0.007	-0.004
	(0.052)	(0.050)	(0.056)	(0.052)	(0.052)	(0.056)
90%	-0.580	-0.144	-0.141	0.013	0.012	0.008
	(0.036)	(0.035)	(0.039)	(0.036)	(0.036)	(0.039)
95%	0.055	-0.039	0.729	-0.018	-0.015	-0.011
	(0.026)	(0.026)	(0.024)	(0.027)	(0.027)	(0.028)

Table 4.7 Estimated bias of alternative estimators of distribution function × 100 for Model 3. Standard error in parentheses. $N = 600$, $n = 60$; 2500 iterations for (c-val) and $10,000$ iterations for all others

$F(\dot{y})$	Local-residuals (c-val)	Local-residuals ($B = 6$)	Chambers-Dunstan	Rao-Kovar Mantel	Post-stratified	Horvitz-Thompson
5%	0.376 (0.056)	0.010 (0.025)	1.186 (0.029)	0.437 (0.030)	-0.022 (0.026)	-0.014 (0.028)
10%	0.410 (0.069)	-0.019 (0.033)	0.902 (0.036)	0.525 (0.040)	-0.041 (0.033)	-0.025 (0.038)
25%	0.238 (0.073)	-0.013 (0.037)	-0.047 (0.048)	0.662 (0.055)	-0.018 (0.036)	0.033 (0.053)
50%	0.083 (0.055)	0.129 (0.034)	0.087 (0.059)	0.508 (0.062)	0.033 (0.030)	0.087 (0.062)
75%	-0.004 (0.052)	0.804 (0.034)	0.539 (0.052)	0.340 (0.053)	0.082 (0.028)	0.073 (0.053)
90%	0.289 (0.059)	0.379 (0.033)	0.913 (0.034)	0.209 (0.037)	0.046 (0.030)	0.028 (0.037)
95%	0.210 (0.050)	0.127 (0.026)	0.244 (0.024)	0.124 (0.027)	0.002 (0.026)	-0.007 (0.028)

Table 4.8 Estimated bias of alternative estimators of distribution function × 100 for Model 1. Standard error in parentheses. $N = 1200$, $n = 120$; $10,000$ iterations

$F(\dot{y})$	Local-residuals (c-val)	Local-residuals ($B = 10$)	Chambers-Dunstan	Rao-Kovar Mantel	Post-stratified	Horvitz-Thompson
5%	0.005 (0.012)	0.008 (0.017)	-0.005 (0.011)	0.013 (0.018)	0.015 (0.019)	0.018 (0.020)
10%	0.013 (0.017)	0.036 (0.022)	-0.012 (0.015)	0.034 (0.023)	0.043 (0.024)	0.047 (0.027)
25%	-0.005 (0.021)	0.036 (0.028)	-0.042 (0.020)	0.039 (0.029)	0.031 (0.030)	0.068 (0.038)
50%	-0.003 (0.018)	-0.029 (0.027)	-0.000 (0.016)	-0.024 (0.029)	-0.027 (0.030)	0.025 (0.044)
75%	0.004 (0.012)	-0.004 (0.019)	0.031 (0.011)	0.011 (0.022)	0.012 (0.022)	0.043 (0.038)
90%	-0.001 (0.008)	-0.011 (0.011)	0.012 (0.008)	-0.015 (0.015)	-0.014 (0.016)	-0.020 (0.027)
95%	0.005 (0.006)	0.027 (0.008)	0.010 (0.006)	0.017 (0.012)	0.023 (0.013)	0.017 (0.019)

Table 4.9 Estimated bias of alternative estimators of distribution function × 100 for Model 2. Standard error in parentheses. $N = 1200$, $n = 120$; $10,000$ iterations

	Local-	Local-	Chambers-	Rao-Kovar	Post-	Horvitz-
$F(\dot{y})$	residuals	residuals	Dunstan	Mantel	stratified	Thompson
	(c-val)	($B = 10$)				
5%	1.550	0.180	3.941	0.028	0.036	0.031
	(0.023)	(0.020)	(0.020)	(0.021)	(0.021)	(0.020)
10%	1.517	0.213	3.483	-0.003	0.012	0.002
	(0.027)	(0.027)	(0.024)	(0.029)	(0.028)	(0.028)
25%	-0.593	-0.273	-1.177	0.015	0.044	0.024
	(0.034)	(0.038)	(0.028)	(0.040)	(0.040)	(0.039)
50%	-1.744	-0.109	-6.677	-0.000	-0.007	0.014
	(0.044)	(0.042)	(0.036)	(0.042)	(0.043)	(0.045)
75%	-1.022	-0.142	-3.531	-0.041	-0.049	-0.035
	(0.036)	(0.036)	(0.042)	(0.036)	(0.037)	(0.039)
90%	-0.410	-0.085	0.344	-0.010	-0.020	-0.004
	(0.025)	(0.025)	(0.026)	(0.026)	(0.026)	(0.027)
95%	-0.048	-0.023	0.930	0.003	0.000	0.009
	(0.018)	(0.018)	(0.016)	(0.019)	(0.019)	(0.020)

Table 4.10 Estimated bias of alternative estimators of distribution function × 100 for Model 3. Standard error in parentheses. $N = 1200$, $n = 120$; 2500 iterations for (c-val) and $10,000$ iterations for all others

$F(\dot{y})$	Local- residuals (c-val)	Local- residuals $(B = 10)$	Chambers- Dunstan	Rao-Kovar Mantel	Post- stratified	Horvitz- Thompson
5%	0.175	0.001	0.678	0.233	-0.018	0.001
	(0.039)	(0.018)	(0.020)	(0.021)	(0.018)	(0.020)
10%	0.221	-0.009	0.539	0.299	-0.006	-0.002
	(0.049)	(0.023)	(0.025)	(0.028)	(0.024)	(0.027)
25%	0.136	0.006	0.109	0.384	0.021	0.027
	(0.052)	(0.026)	(0.035)	(0.039)	(0.026)	(0.039)
50%	0.034	-0.031	-0.054	0.201	-0.022	-0.015
	(0.038)	(0.020)	(0.042)	(0.044)	(0.019)	(0.044)
75%	-0.071	0.112	0.098	0.124	0.007	-0.022
	(0.033)	(0.019)	(0.037)	(0.038)	(0.017)	(0.038)
90%	0.141	0.328	0.250	0.074	0.004	-0.022
	(0.033)	(0.019)	(0.025)	(0.026)	(0.015)	(0.026)
95%	0.161	0.184	0.364	0.075	0.023	0.010
	(0.031)	(0.017)	(0.017)	(0.019)	(0.015)	(0.019)

Table 4.11 Ratio of Bias square to Mean Square Error × 100 for Local-residuals estimator and Chambers-Dunstan estimator for Model 2. Standard error in parentheses. $N = 600$, $n = 60$; $10,000$ iterations

$F(\dot{y})$	Local-residuals	Local-residuals	Chambers-Dunstan
	(c-val)	$(B = 6)$	
5%	37.29	2.03	65.68
	(0.60)	(0.27)	(0.40)
10%	25.51	1.22	51.47
	(0.66)	(0.21)	(0.55)
25%	0.95	0.71	4.91
	(0.19)	(0.17)	(0.42)
50%	16.85	0.35	58.50
	(0.68)	(0.12)	(0.62)
75%	11.75	0.29	31.17
	(0.58)	(0.11)	(0.73)
90%	2.48	0.17	0.13
	(0.30)	(0.08)	(0.07)
95%	0.04	0.02	8.20
	(0.04)	(0.03)	(0.57)

Table 4.12 Ratio of Bias square to Mean Square Error × 100 for Local-residuals estimator and Chambers-Dunstan estimator for Model 3. Standard error in parentheses. $N = 600$, $n = 60$; 2500 iterations for (c-val) and $10,000$ iterations for all others

$F(\dot{y})$	Local-residuals (c-val)	Local-residuals $(B = 6)$	Chambers-Dunstan
5%	1.77 (0.51)	0.00 (0.01)	13.93 (0.55)
10%	1.40 (0.46)	0.00 (0.01)	6.00 (0.44)
25%	0.42 (0.26)	0.00 (0.01)	0.01 (0.02)
50%	0.09 (0.12)	0.14 (0.08)	0.02 (0.03)
75%	0.00 (0.01)	5.38 (0.43)	1.06 (0.20)
90%	0.96 (0.39)	1.33 (0.23)	6.79 (0.51)
95%	0.71 (0.34)	0.24 (0.10)	1.06 (0.21)

Table 4.13 Ratio of Bias square to Mean Square Error × 100 for Local-residuals estimator and Chambers-Dunstan estimator for Model 2. Standard error in parentheses. $N = 1200$, $n = 120$; $10,000$ iterations

$F(\dot{y})$	Local-residuals (c-val)	Local-residuals ($B = 10$)	Chambers-Dunstan
5%	32.17 (0.61)	0.84 (0.18)	79.44 (0.26)
10%	23.79 (0.65)	0.62 (0.15)	68.40 (0.39)
25%	2.99 (0.34)	0.50 (0.14)	14.85 (0.65)
50%	13.37 (0.61)	0.07 (0.05)	77.76 (0.34)
75%	7.32 (0.49)	0.16 (0.08)	41.88 (0.68)
90%	2.71 (0.31)	0.12 (0.07)	1.68 (0.26)
95%	0.07 (0.05)	0.02 (0.02)	25.69 (0.79)

Table 4.14 Ratio of Bias square to Mean Square Error × 100 for Local-residuals estimator and Chambers-Dunstan estimator for Model 3. Standard error in parentheses. $N = 1200$, $n = 120$; 2500 iterations for (c-val) and $10,000$ iterations for all others

$F(\dot{y})$	Local-residuals	Local-residuals	Chambers-Dunstan
	(c-val)	$(B = 10)$	
5%	0.81	0.00	10.46
	(0.35)	(0.00)	(0.52)
10%	0.81	0.00	4.29
	(0.36)	(0.01)	(0.39)
25%	0.28	0.00	0.09
	(0.21)	(0.00)	(0.06)
50%	0.03	0.02	0.02
	(0.07)	(0.03)	(0.03)
75%	0.19	0.35	0.07
	(0.17)	(0.12)	(0.05)
90%	0.72	2.81	0.97
	(0.33)	(0.32)	(0.20)
95%	1.08	1.21	4.31
	(0.41)	(0.22)	(0.41)

Table 4.15 Ratio of Root Mean Square Error (rMSE) of alternative estimators to the rMSE of Horvitz-Thompson estimator, and rMSE of Horvitz-Thompson estimator × 100, for Model 1. Standard error in parentheses. $N = 600$, $n = 60$; $10,000$ iterations

	Ratios of rMSE for alternative estimators:					rMSE ×100 of
$F(\dot{y})$	Local-residuals (c-val)	Local-residuals ($B = 6$)	Chambers-Dunstan	Rao-Kovar Mantel	Post-stratified	Horvitz-Thompson
5%	0.617 (0.006)	0.831 (0.005)	0.558 (0.004)	0.929 (0.004)	0.970 (0.005)	2.804 (0.028)
10%	0.630 (0.005)	0.809 (0.005)	0.583 (0.005)	0.889 (0.005)	0.919 (0.005)	3.812 (0.038)
25%	0.570 (0.005)	0.715 (0.005)	0.541 (0.004)	0.781 (0.005)	0.808 (0.005)	5.444 (0.054)
50%	0.432 (0.004)	0.584 (0.005)	0.395 (0.004)	0.661 (0.005)	0.678 (0.005)	6.291 (0.063)
75%	0.299 (0.003)	0.442 (0.004)	0.269 (0.002)	0.566 (0.005)	0.592 (0.005)	5.383 (0.054)
90%	0.240 (0.002)	0.284 (0.003)	0.233 (0.002)	0.491 (0.004)	0.601 (0.005)	3.713 (0.037)
95%	0.278 (0.003)	0.320 (0.003)	0.272 (0.003)	0.565 (0.005)	0.803 (0.005)	2.718 (0.027)

Table 4.16 Ratio of Root Mean Square Error (rMSE) of alternative estimators to the rMSE of Horvitz-Thompson estimator, and rMSE of Horvitz-Thompson estimator × 100, for Model 2. Standard error in parentheses. $N = 600$, $n = 60$; $10,000$ iterations

	Ratios of rMSE for alternative estimators:					rMSE ×100 of
	Local-	Local-	Chambers-	Rao-Kovar	Post-	Horvitz-
$F(\dot{y})$	residuals	residuals	Dunstan	Mantel	stratified	Thompson
	(c-val)	($B = 6$)				
5%	1.357	0.967	1.738	1.057	1.029	2.830
	(0.011)	(0.005)	(0.014)	(0.003)	(0.004)	(0.028)
10%	1.098	0.969	1.255	1.050	1.030	3.867
	(0.008)	(0.005)	(0.010)	(0.003)	(0.003)	(0.039)
25%	0.816	0.944	0.743	1.008	1.026	5.652
	(0.004)	(0.004)	(0.005)	(0.003)	(0.003)	(0.057)
50%	0.998	0.914	1.236	0.944	0.957	6.483
	(0.006)	(0.005)	(0.009)	(0.004)	(0.004)	(0.065)
75%	0.984	0.890	1.215	0.930	0.931	5.589
	(0.006)	(0.005)	(0.009)	(0.004)	(0.004)	(0.056)
90%	0.955	0.904	1.006	0.945	0.946	3.857
	(0.006)	(0.005)	(0.007)	(0.004)	(0.005)	(0.039)
95%	0.938	0.916	0.911	0.966	0.977	2.794
	(0.006)	(0.006)	(0.006)	(0.004)	(0.005)	(0.028)

Table 4.17 Ratio of Root Mean Square Error (rMSE) of alternative estimators to the rMSE of Horvitz-Thompson estimator, and rMSE of Horvitz-Thompson estimator × 100, for Model 3. Standard error in parentheses. $N = 600$, $n = 60$; 2500 iterations for (c-val) and $10,000$ iterations for all others

	Ratios of rMSE for alternative estimators:					rMSE ×100 of
$F(\dot{y})$	Local-residuals (c-val)	Local-residuals ($B = 6$)	Chambers-Dunstan	Rao-Kovar Mantel	Post-stratified	Horvitz-Thompson
5%	1.010 (0.016)	0.903 (0.005)	1.133 (0.010)	1.085 (0.005)	0.943 (0.005)	2.803 (0.028)
10%	0.917 (0.014)	0.865 (0.005)	0.973 (0.006)	1.066 (0.004)	0.884 (0.005)	3.786 (0.038)
25%	0.685 (0.012)	0.690 (0.005)	0.894 (0.003)	1.039 (0.003)	0.669 (0.005)	5.346 (0.053)
50%	0.444 (0.008)	0.550 (0.005)	0.954 (0.003)	1.012 (0.002)	0.486 (0.004)	6.160 (0.062)
75%	0.483 (0.010)	0.649 (0.005)	0.982 (0.003)	1.003 (0.001)	0.534 (0.004)	5.336 (0.053)
90%	0.790 (0.012)	0.883 (0.006)	0.939 (0.004)	0.995 (0.001)	0.795 (0.005)	3.731 (0.037)
95%	0.904 (0.014)	0.937 (0.005)	0.858 (0.004)	0.995 (0.001)	0.948 (0.005)	2.758 (0.028)

Table 4.18 Ratio of Root Mean Square Error (rMSE) of alternative estimators to the rMSE of Horvitz-Thompson estimator, and rMSE of Horvitz-Thompson estimator × 100, for Model 1. Standard error in parentheses. $N = 1200$, $n = 120$; $10,000$ iterations

	Ratios of rMSE for alternative estimators:					rMSE ×100 of
$F(\dot{y})$	Local-residuals (c-val)	Local-residuals ($B = 12$)	Chambers-Dunstan	Rao-Kovar Mantel	Post-stratified	Horvitz-Thompson
5%	0.609 (0.007)	0.873 (0.005)	0.554 (0.004)	0.928 (0.004)	0.960 (0.005)	1.968 (0.020)
10%	0.628 (0.007)	0.836 (0.005)	0.580 (0.005)	0.874 (0.004)	0.903 (0.005)	2.664 (0.027)
25%	0.557 (0.006)	0.730 (0.005)	0.526 (0.004)	0.761 (0.005)	0.780 (0.005)	3.836 (0.038)
50%	0.420 (0.005)	0.613 (0.005)	0.372 (0.003)	0.654 (0.005)	0.670 (0.005)	4.401 (0.044)
75%	0.311 (0.003)	0.500 (0.004)	0.279 (0.003)	0.571 (0.005)	0.590 (0.005)	3.804 (0.038)
90%	0.304 (0.003)	0.413 (0.004)	0.291 (0.003)	0.564 (0.005)	0.595 (0.005)	2.665 (0.027)
95%	0.309 (0.003)	0.399 (0.004)	0.299 (0.003)	0.604 (0.005)	0.699 (0.005)	1.927 (0.019)

Table 4.19 Ratio of Root Mean Square Error (rMSE) of alternative estimators to the rMSE of Horvitz-Thompson estimator, and rMSE of Horvitz-Thompson estimator × 100, for Model 2. Standard error in parentheses. $N = 1200$, $n = 120$; $10,000$ iterations

	Ratios of rMSE for alternative estimators:					rMSE ×100 of
$F(\dot{y})$	Local-residuals (c-val)	Local-residuals ($B = 12$)	Chambers-Dunstan	Rao-Kovar Mantel	Post-stratified	Horvitz-Thompson
5%	1.359 (0.015)	0.980 (0.004)	2.199 (0.017)	1.055 (0.003)	1.024 (0.003)	2.011 (0.020)
10%	1.130 (0.011)	0.986 (0.004)	1.530 (0.012)	1.051 (0.003)	1.026 (0.003)	2.752 (0.028)
25%	0.875 (0.009)	0.980 (0.004)	0.778 (0.006)	1.011 (0.003)	1.024 (0.003)	3.923 (0.039)
50%	1.068 (0.010)	0.934 (0.005)	1.695 (0.013)	0.943 (0.004)	0.960 (0.004)	4.467 (0.045)
75%	0.966 (0.010)	0.913 (0.005)	1.395 (0.010)	0.927 (0.004)	0.936 (0.004)	3.912 (0.039)
90%	0.916 (0.009)	0.921 (0.005)	0.975 (0.006)	0.949 (0.003)	0.952 (0.004)	2.719 (0.027)
95%	0.924 (0.009)	0.936 (0.005)	0.932 (0.006)	0.966 (0.003)	0.975 (0.005)	1.970 (0.020)

Table 4.20 Ratio of Root Mean Square Error (rMSE) of alternative estimators to the rMSE of Horvitz-Thompson estimator, and rMSE of Horvitz-Thompson estimator × 100, for Model 3. Standard error in parentheses. $N = 1200$, $n = 120$; 2500 iterations for (c-val) and $10,000$ iterations for all others

	Ratios of rMSE for alternative estimators:					rMSE ×100 of
	Local-	Local-	Chambers-	Rao-Kovar	Post-	Horvitz-
$F(\dot{y})$	residuals	residuals	Dunstan	Mantel	stratified	Thompson
	(c-val)	($B = 12$)				
5%	0.979	0.913	1.060	1.052	0.932	1.977
	(0.020)	(0.005)	(0.008)	(0.003)	(0.005)	(0.020)
10%	0.906	0.866	0.963	1.042	0.871	2.700
	(0.018)	(0.005)	(0.005)	(0.003)	(0.005)	(0.027)
25%	0.671	0.678	0.919	1.026	0.672	3.850
	(0.013)	(0.005)	(0.003)	(0.002)	(0.005)	(0.039)
50%	0.436	0.463	0.965	1.007	0.430	4.396
	(0.009)	(0.004)	(0.002)	(0.001)	(0.004)	(0.044)
75%	0.429	0.492	0.973	1.000	0.433	3.826
	(0.009)	(0.004)	(0.003)	(0.001)	(0.004)	(0.038)
90%	0.632	0.743	0.962	0.997	0.586	2.640
	(0.013)	(0.006)	(0.003)	(0.001)	(0.004)	(0.026)
95%	0.811	0.878	0.918	0.993	0.812	1.908
	(0.016)	(0.006)	(0.005)	(0.001)	(0.005)	(0.019)

Table 4.21 Square root of Monte Carlo estimated variance ×100 (*) and square root of Monte Carlo average of variance estimators × 100 (**) for Model 1. Standard error in parentheses. $N = 600$, $n = 60$; 1000 iterations. Local-residuals estimator calculated with fixed number of bins ($B = 6$)

	(*)	(**)	(*)	(**)	(**)	(**)
$F(\dot{y})$	$\widehat{F}_L(\dot{y}) - F_N(\dot{y})$	$\widehat{V}_{ee}(\dot{y})$	$\widehat{F}_L(\dot{y}) - F(\dot{y})$	$\widehat{V}_{\mathcal{L}}(\dot{y})$	$\widehat{V}_{JKwo}(\dot{y})$	$\widehat{V}_{JK}(\dot{y})$
5%	2.252	2.244	2.332	2.326	2.431	2.445
	(0.024)	(0.008)	(0.024)	(0.009)	(0.009)	(0.009)
10%	2.975	2.965	3.094	3.079	3.218	3.235
	(0.031)	(0.007)	(0.032)	(0.007)	(0.008)	(0.008)
25%	3.710	3.745	3.875	3.891	4.066	4.080
	(0.036)	(0.006)	(0.038)	(0.007)	(0.007)	(0.007)
50%	3.535	3.551	3.657	3.680	3.845	3.855
	(0.036)	(0.007)	(0.037)	(0.007)	(0.008)	(0.008)
75%	2.319	2.374	2.358	2.430	2.536	2.516
	(0.023)	(0.007)	(0.024)	(0.007)	(0.007)	(0.007)
90%	1.124	1.109	1.067	1.093	1.125	1.121
	(0.012)	(0.004)	(0.012)	(0.004)	(0.004)	(0.005)
95%	0.909	0.837	0.872	0.813	0.829	0.917
	(0.009)	(0.003)	(0.009)	(0.003)	(0.003)	(0.004)

Table 4.22 Square root of Monte Carlo estimated variance ×100 (*) and square root of Monte Carlo average of variance estimators × 100 (**) for Model 2. Standard error in parentheses. $N = 600$, $n = 60$; 1000 iterations. Local-residuals estimator calculated with fixed number of bins ($B = 6$)

	(*)	(**)	(*)	(**)	(**)	(**)
$F(\dot{y})$	$\widehat{F}_L(\dot{y}) - F_N(\dot{y})$	$\widehat{V}_{ee}(\dot{y})$	$\widehat{F}_L(\dot{y}) - F(\dot{y})$	$\widehat{V}_{\mathcal{L}}(\dot{y})$	$\widehat{V}_{JKwo}(\dot{y})$	$\widehat{V}_{JK}(\dot{y})$
5%	2.585	2.649	2.674	2.746	2.874	2.888
	(0.027)	(0.011)	(0.028)	(0.011)	(0.012)	(0.012)
10%	3.562	3.612	3.714	3.759	3.933	3.956
	(0.036)	(0.010)	(0.038)	(0.010)	(0.011)	(0.012)
25%	5.096	5.075	5.299	5.298	5.543	5.609
	(0.052)	(0.007)	(0.054)	(0.008)	(0.008)	(0.009)
50%	5.634	5.591	5.877	5.845	6.109	6.205
	(0.056)	(0.006)	(0.057)	(0.006)	(0.007)	(0.008)
75%	4.740	4.669	4.927	4.883	5.103	5.214
	(0.049)	(0.007)	(0.051)	(0.007)	(0.008)	(0.009)
90%	3.358	3.225	3.528	3.369	3.520	3.679
	(0.034)	(0.009)	(0.036)	(0.009)	(0.009)	(0.010)
95%	2.497	2.351	2.592	2.453	2.560	2.700
	(0.026)	(0.009)	(0.027)	(0.010)	(0.010)	(0.011)

Table 4.23 Square root of Monte Carlo estimated variance ×100 (*) and square root of Monte Carlo average of variance estimators × 100 (**) for Model 3. Standard error in parentheses. $N = 600$, $n = 60$; 1000 iterations. Local-residuals estimator calculated with fixed number of bins ($B = 6$)

	(*)	(**)	(*)	(**)	(**)	(**)
$F(\dot{y})$	$\widehat{F}_L(\dot{y}) - F_N(\dot{y})$	$\widehat{V}_{ee}(\dot{y})$	$\widehat{F}_L(\dot{y}) - F(\dot{y})$	$\widehat{V}_{\mathcal{L}}(\dot{y})$	$\widehat{V}_{JKwo}(\dot{y})$	$\widehat{V}_{JK}(\dot{y})$
5%	2.445	2.455	2.555	2.564	2.682	3.010
	(0.027)	(0.010)	(0.027)	(0.010)	(0.010)	(0.015)
10%	3.187	3.228	3.327	3.375	3.529	3.876
	(0.033)	(0.008)	(0.034)	(0.009)	(0.009)	(0.014)
25%	3.691	3.798	3.828	3.968	4.149	4.536
	(0.037)	(0.009)	(0.039)	(0.009)	(0.010)	(0.014)
50%	3.323	3.556	3.390	3.713	3.882	4.078
	(0.036)	(0.010)	(0.037)	(0.011)	(0.011)	(0.012)
75%	3.319	3.578	3.364	3.739	3.910	4.010
	(0.038)	(0.011)	(0.038)	(0.011)	(0.012)	(0.012)
90%	3.270	3.259	3.314	3.419	3.573	3.651
	(0.034)	(0.009)	(0.034)	(0.009)	(0.010)	(0.011)
95%	2.569	2.523	2.601	2.643	2.763	2.934
	(0.029)	(0.010)	(0.029)	(0.011)	(0.011)	(0.014)

Table 4.24 Square root of Monte Carlo estimated variance ×100 (*) and square root of Monte Carlo average of variance estimators × 100 (**) for Model 1. Standard error in parentheses. $N = 1200$, $n = 120$; 1000 iterations. Local-residuals estimator calculated with fixed number of bins ($B = 10$)

	(*)	(**)	(*)	(**)	(**)	(**)
$F(\dot{y})$	$\widehat{F}_L(\dot{y}) - F_N(\dot{y})$	$\widehat{V}_{ee}(\dot{y})$	$\widehat{F}_L(\dot{y}) - F(\dot{y})$	$\widehat{V}_{\mathcal{L}}(\dot{y})$	$\widehat{V}_{JKwo}(\dot{y})$	$\widehat{V}_{JK}(\dot{y})$
5%	1.679	1.660	1.747	1.747	1.800	1.802
	(0.040)	(0.010)	(0.041)	(0.010)	(0.011)	(0.011)
10%	2.123	2.172	2.210	2.286	2.356	2.358
	(0.046)	(0.009)	(0.049)	(0.009)	(0.009)	(0.009)
25%	2.680	2.721	2.763	2.864	2.951	2.952
	(0.060)	(0.008)	(0.060)	(0.008)	(0.008)	(0.008)
50%	2.623	2.570	2.798	2.702	2.784	2.784
	(0.063)	(0.009)	(0.069)	(0.009)	(0.009)	(0.009)
75%	1.920	1.846	1.972	1.929	1.987	1.988
	(0.041)	(0.008)	(0.044)	(0.009)	(0.009)	(0.009)
90%	1.075	1.119	1.087	1.148	1.181	1.167
	(0.025)	(0.006)	(0.026)	(0.007)	(0.007)	(0.007)
95%	0.767	0.781	0.751	0.793	0.811	0.821
	(0.017)	(0.005)	(0.016)	(0.006)	(0.006)	(0.006)

Table 4.25 Square root of Monte Carlo estimated variance ×100 (*) and square root of Monte Carlo average of variance estimators × 100 (**) for Model 2. Standard error in parentheses. $N = 1200$, $n = 120$; 1000 iterations. Local-residuals estimator calculated with fixed number of bins ($B = 10$)

	(*)	(**)	(*)	(**)	(**)	(**)
$F(\dot{y})$	$\widehat{F}_L(\dot{y}) - F_N(\dot{y})$	$\widehat{V}_{ee}(\dot{y})$	$\widehat{F}_L(\dot{y}) - F(\dot{y})$	$\widehat{V}_{\mathcal{L}}(\dot{y})$	$\widehat{V}_{JKwo}(\dot{y})$	$\widehat{V}_{JK}(\dot{y})$
5%	1.858	1.877	1.962	1.977	2.038	2.034
	(0.044)	(0.012)	(0.046)	(0.013)	(0.013)	(0.014)
10%	2.571	2.585	2.716	2.726	2.810	2.803
	(0.057)	(0.011)	(0.060)	(0.012)	(0.012)	(0.013)
25%	3.841	3.679	4.008	3.886	4.007	4.020
	(0.088)	(0.008)	(0.095)	(0.009)	(0.009)	(0.010)
50%	4.199	4.010	4.357	4.239	4.369	4.381
	(0.094)	(0.006)	(0.098)	(0.007)	(0.007)	(0.007)
75%	3.474	3.381	3.659	3.577	3.686	3.712
	(0.078)	(0.008)	(0.084)	(0.008)	(0.008)	(0.009)
90%	2.382	2.363	2.473	2.498	2.574	2.619
	(0.054)	(0.010)	(0.054)	(0.010)	(0.010)	(0.011)
95%	1.800	1.743	1.868	1.842	1.897	1.934
	(0.043)	(0.011)	(0.045)	(0.011)	(0.012)	(0.012)

Table 4.26 Square root of Monte Carlo estimated variance ×100 (*) and square root of Monte Carlo average of variance estimators × 100 (**) for Model 3. Standard error in parentheses. N = 1200, n = 120; 1000 iterations. Local-residuals estimator calculated with fixed number of bins ($B = 10$)

	$(*)$	$(**)$	$(*)$	$(**)$	$(**)$	$(**)$
$F(\dot{y})$	$\widehat{F}_L(\dot{y}) - F_N(\dot{y})$	$\widehat{V}_{ee}(\dot{y})$	$\widehat{F}_L(\dot{y}) - F(\dot{y})$	$\widehat{V}_{\mathcal{L}}(\dot{y})$	$\widehat{V}_{JKwo}(\dot{y})$	$\widehat{V}_{JK}(\dot{y})$
5%	1.758 (0.041)	1.723 (0.011)	1.834 (0.042)	1.822 (0.011)	1.878 (0.011)	1.969 (0.014)
10%	2.307 (0.051)	2.243 (0.009)	2.429 (0.054)	2.373 (0.009)	2.446 (0.010)	2.559 (0.014)
25%	2.454 (0.055)	2.520 (0.010)	2.520 (0.057)	2.665 (0.010)	2.747 (0.010)	2.873 (0.016)
50%	1.922 (0.041)	2.078 (0.012)	1.996 (0.043)	2.195 (0.012)	2.263 (0.012)	2.334 (0.014)
75%	1.870 (0.046)	1.948 (0.013)	1.930 (0.047)	2.054 (0.014)	2.117 (0.014)	2.179 (0.015)
90%	1.857 (0.046)	1.970 (0.011)	1.901 (0.045)	2.085 (0.011)	2.149 (0.012)	2.173 (0.012)
95%	1.627 (0.041)	1.657 (0.010)	1.667 (0.041)	1.756 (0.011)	1.809 (0.011)	1.863 (0.013)

Table 4.27 Coverage probability for a 95% confidence interval of $F_N(\dot{y})$ based on alternative variance estimators × 100 for Model 1. Standard error in parentheses. $N = 600$, $n = 60$; 1000 iterations. Local-residuals estimator calculated with fixed number of bins ($B = 6$)

	Variance estimator used for the confidence interval			
$F(\dot{y})$	$\widehat{V}_{ee}(\dot{y})$	$0.9\widehat{V}_{\mathcal{L}}(\dot{y})$	$0.9\widehat{V}_{JKwo}(\dot{y})$	$0.9\widehat{V}_{JK}(\dot{y})$
5%	89.24 (0.44)	88.86 (0.44)	89.52 (0.43)	89.68 (0.43)
10%	92.96 (0.36)	92.64 (0.37)	93.38 (0.35)	93.48 (0.35)
25%	94.92 (0.31)	94.70 (0.32)	95.40 (0.30)	95.50 (0.29)
50%	94.28 (0.33)	94.08 (0.33)	94.78 (0.31)	94.92 (0.31)
75%	93.84 (0.34)	93.18 (0.36)	93.82 (0.34)	93.78 (0.34)
90%	92.64 (0.37)	90.72 (0.41)	91.48 (0.39)	91.22 (0.40)
95%	89.60 (0.43)	86.76 (0.48)	86.70 (0.48)	89.56 (0.43)

Table 4.28 Coverage probability for a 95% confidence interval of $F_N(\dot{y})$ based on alternative variance estimators × 100 for Model 2. Standard error in parentheses. $N = 600$, $n = 60$; 1000 iterations. Local-residuals estimator calculated with fixed number of bins ($B = 6$)

	Variance estimator used for the confidence interval			
$F(\dot{y})$	$\widehat{V}_{ee}(\dot{y})$	$0.9\widehat{V}_{\mathcal{L}}(\dot{y})$	$0.9\widehat{V}_{JKwo}(\dot{y})$	$0.9\widehat{V}_{JK}(\dot{y})$
5%	90.46 (0.42)	90.06 (0.42)	90.88 (0.41)	91.10 (0.40)
10%	93.18 (0.36)	92.80 (0.37)	93.50 (0.35)	93.36 (0.35)
25%	93.44 (0.35)	93.12 (0.36)	94.18 (0.33)	94.24 (0.33)
50%	94.72 (0.32)	94.58 (0.32)	95.56 (0.29)	95.86 (0.28)
75%	93.34 (0.35)	93.28 (0.35)	94.22 (0.33)	94.56 (0.32)
90%	91.12 (0.40)	90.90 (0.41)	92.02 (0.38)	93.16 (0.36)
95%	88.06 (0.46)	87.80 (0.46)	88.72 (0.45)	89.40 (0.44)

Table 4.29 Coverage probability for a 95% confidence interval of $F_N(\dot{y})$ based on alternative variance estimators × 100 for Model 3. Standard error in parentheses. $N = 600$, $n = 60$; 1000 iterations. Local-residuals estimator calculated with fixed number of bins ($B = 6$)

	Variance estimator used for the confidence interval			
$F(\dot{y})$	$\widehat{V}_{ee}(\dot{y})$	$0.9\widehat{V}_{\mathcal{L}}(\dot{y})$	$0.9\widehat{V}_{JKwo}(\dot{y})$	$0.9\widehat{V}_{JK}(\dot{y})$
5%	89.34 (0.44)	89.04 (0.44)	90.18 (0.42)	92.28 (0.38)
10%	92.58 (0.37)	92.44 (0.37)	93.30 (0.35)	94.92 (0.31)
25%	94.62 (0.32)	94.40 (0.33)	95.30 (0.30)	96.56 (0.26)
50%	95.78 (0.28)	95.60 (0.29)	96.52 (0.26)	96.96 (0.24)
75%	95.22 (0.30)	95.18 (0.30)	95.96 (0.28)	96.22 (0.27)
90%	91.68 (0.39)	91.66 (0.39)	92.76 (0.37)	93.10 (0.36)
95%	87.28 (0.47)	87.24 (0.47)	88.30 (0.45)	89.42 (0.43)

Table 4.30 Coverage probability for a 95% confidence interval of $F_N(\dot{y})$ based on alternative variance estimators × 100 for Model 1. Standard error in parentheses. $N = 1200$, $n = 120$; 1000 iterations. Local-residuals estimator calculated with fixed number of bins ($B = 10$)

	Variance estimator used for the confidence interval			
$F(\dot{y})$	$\widehat{V}_{ee}(\dot{y})$	$0.9\widehat{V}_{\mathcal{L}}(\dot{y})$	$0.9\widehat{V}_{JKwo}(\dot{y})$	$0.9\widehat{V}_{JK}(\dot{y})$
5%	92.70 (0.82)	92.70 (0.82)	93.30 (0.79)	93.30 (0.79)
10%	94.40 (0.73)	94.40 (0.73)	95.10 (0.68)	95.10 (0.68)
25%	95.20 (0.68)	95.40 (0.66)	95.90 (0.63)	95.90 (0.63)
50%	94.30 (0.73)	94.30 (0.73)	94.70 (0.71)	94.70 (0.71)
75%	93.40 (0.79)	93.20 (0.80)	94.10 (0.75)	94.10 (0.75)
90%	95.10 (0.68)	94.00 (0.75)	95.00 (0.69)	94.90 (0.70)
95%	93.90 (0.76)	92.50 (0.83)	92.90 (0.81)	93.50 (0.78)

Table 4.31 Coverage probability for a 95% confidence interval of $F_N(\dot{y})$ based on alternative variance estimators × 100 for Model 2. Standard error in parentheses. $N = 1200$, $n = 120$; 1000 iterations. Local-residuals estimator calculated with fixed number of bins ($B = 10$)

	Variance estimator used for the confidence interval			
$F(\dot{y})$	$\widehat{V}_{ee}(\dot{y})$	$0.9\widehat{V}_{\mathcal{L}}(\dot{y})$	$0.9\widehat{V}_{JKwo}(\dot{y})$	$0.9\widehat{V}_{JK}(\dot{y})$
5%	92.60 (0.83)	92.50 (0.83)	92.90 (0.81)	92.90 (0.81)
10%	94.70 (0.71)	94.60 (0.71)	95.40 (0.66)	95.20 (0.68)
25%	93.80 (0.76)	93.90 (0.76)	95.00 (0.69)	95.30 (0.67)
50%	94.70 (0.71)	94.80 (0.70)	95.50 (0.66)	95.40 (0.66)
75%	93.40 (0.79)	93.50 (0.78)	94.20 (0.74)	94.20 (0.74)
90%	93.70 (0.77)	93.90 (0.76)	94.50 (0.72)	95.00 (0.69)
95%	91.40 (0.89)	91.60 (0.88)	92.10 (0.85)	92.50 (0.83)

Table 4.32 Coverage probability for a 95% confidence interval of $F_N(\dot{y})$ based on alternative variance estimators × 100 for Model 3. Standard error in parentheses. $N = 1200$, $n = 120$; 1000 iterations. Local-residuals estimator calculated with fixed number of bins ($B = 10$)

	Variance estimator used for the confidence interval			
$F(\dot{y})$	$\widehat{V}_{ee}(\dot{y})$	$0.9\widehat{V}_{\mathcal{L}}(\dot{y})$	$0.9\widehat{V}_{JKwo}(\dot{y})$	$0.9\widehat{V}_{JK}(\dot{y})$
5%	91.20 (0.90)	91.30 (0.89)	92.50 (0.83)	93.40 (0.79)
10%	93.20 (0.80)	93.30 (0.79)	93.90 (0.76)	94.70 (0.71)
25%	95.20 (0.68)	95.30 (0.67)	95.90 (0.63)	96.50 (0.58)
50%	96.00 (0.62)	96.10 (0.61)	96.50 (0.58)	96.60 (0.57)
75%	95.40 (0.66)	95.40 (0.66)	95.90 (0.63)	96.60 (0.57)
90%	94.80 (0.70)	95.20 (0.68)	95.70 (0.64)	95.90 (0.63)
95%	91.50 (0.88)	91.70 (0.87)	92.30 (0.84)	92.40 (0.84)

5 CONCLUSIONS

There are some general comments that we can make about the local-residuals estimator of the finite population distribution function.

The local-residuals estimator was designed to overcome the sensitivity to model misspecification of the Chambers and Dunstan estimator, and to retain the good performance of the model based Chambers and Dunstan method when the model is correctly specified. The robustness of the local-residuals estimator resides in the construction of small bins where it seems reasonable to assume a common distribution function for the residuals. The variance estimators for the local-residuals estimator exhibit robustness against model misspecification similar to the robustness of the local-residuals estimator.

The local-residuals estimator is a nondecreasing function of $\dot{y}$ with limit of zero when $\dot{y}$ goes to $-\infty$ and limit equal to one as $\dot{y} \to +\infty$. The three estimators proposed by Rao, Kovar and Mantel (1990) fail to meet this property.

The $k(N-n)$ imputed values $\widehat{y}_{ij}$ used in the local-residuals estimator need to be computed only once. As shown in (3.1.9), the distribution function can be computed for as many $\dot{y}$ values as desired by using the y_j from the sample and the $\widehat{y}_{ij}$ imputed values, without recomputing the regression line or reusing the auxiliary information. Variance estimators can also be computed from the y_j from the sample and the $\widehat{y}_{ij}$ imputed values. This may be important in practice to save computation time when there is a large number of auxiliary variables or a large number of points $\dot{y}$ where we want to compute the distribution function.

The results of model consistency and limiting normal distribution for the local-residuals estimator obtained in Theorem 3.2.3 and Theorem 3.2.4 require only that the number of bins increase as the sample size increases. Thus the theory holds for bins containing a fixed number of elements. Variance estimators that are consistent under the same conditions were developed. One variance estimator, that proposed in Theorem 3.3.2, requires the number of elements per bin to increase as N increases.

When the number of bins increases faster than the number of elements per bin, by Theorem 3.2.4, the local-residuals estimator has a limiting normal distribution. On the other hand, when the number of bins is chosen on the basis of the sample, one may be able to choose a number of bins to get efficiency close to that of the Chambers and Dunstan estimator under a correctly specified model.

BIBLIOGRAPHY

Abbitt, P. J., Goyeneche, J. J., and Schumi, J. A. (1998). An approach to estimating clay profile distributions (to appear). In *Proceedings of the Section on Survey Research Methods*. American Statistical Association.

Cassel, C.-M., Särndal, C.-E., and Wretman, J. H. (1993). *Foundations of inference in survey sampling*. John Wiley & Sons, New York.

Chambers, R. L., Dorfman, A. H., and Wehrly, E. (1993). Bias robust estimation in finite populations using nonparametric calibration. *Journal of the American Statistical Association*, 88:268–277.

Chambers, R. L. and Dunstan, R. (1986). Estimating distribution functions from survey data. *Biometrika*, 73:597–604.

Dorfman, A. H. (1993). A comparison of design-based and model-based estimators of the finite population distribution function. *Australian Journal of Statistics*, 35:29–41.

Dorfman, A. H. and Hall, P. (1993). Estimators of the finite population distribution using nonparametric regression. *Annals of Statistics*, 21:1452–1475.

Dunstan, R. and Chambers, R. L. (1989). Estimating distribution functions from survey data with limited benchmark information. *Australian Journal of Statistics*, 31:1–11.

Francisco, C. A. and Fuller, W. A. (1991). Quantile estimation with a complex survey design. *Annals of Statistics*, 19:454–469.

Fuller, W. A. (1966). Estimation employing post strata. *Journal of the American Statistical Association*, 61:1172–1183.

Hansen, M. H., Madow, W. G., and Tepping, B. J. (1983). An evaluation of model-dependent and probability-sampling inferences in sample surveys. *Journal of the American Statistical Association*, 78:776–793.

Isaki, C. T. and Fuller, W. A. (1982). Survey design under the regression superpopulation model. *Journal of the American Statistical Association*, 77:89–96.

Kuk, A. Y. C. (1993). A kernel method for estimationg finite population distribution functions using auxiliary information. *Biometrika*, 80:385–392.

Kuo, L. (1988). Classical and prediction approaches to estimating distribution functions from survey data. In *Proceedings of the Section on Survey Research Methods*, pages 280–285. American Statistical Association.

Nascimento-Silva, P. L. D. and Skinner, C. J. (1995). Estimating distribution functions with auxiliary information using poststratification. *Journal of Official Statistics*, 11:277–294.

Randles, R. H. (1982). On the asymptotic normality of statistics with estimated parameters. *Annals of Statistics*, 10:462–474.

Rao, J. N. K., Kovar, J. G., and Mantel, H. J. (1990). On estimating distribution functions and quantiles from survey data using auxiliary information. *Biometrika*, 77:365–375.

Rao, J. N. K. and Liu, J. (1992). On estimating distribution functions from sample survey data using supplementary information at the estimation stage. In Saleh, A. K. M. E., editor, *Nonparametric Statistics and Related Topics*, pages 399–407. Elsevier Science Publishers, Amsterdam.

Silverman, B. W. (1985). *Density estimation for statistics and data analysis.* Chapman and Hall, London.

Wand, M. P. and Jones, M. C. (1995). *Kernel Smoothing.* Chapman and Hall, London.

Wang, S. and Dorfman, A. H. (1996). A new estimator for the finite population distribution function. *Biometrika*, 83:639–652.

Wey, I.-T. (1966). *Estimation of the mean using the rank statistics of an auxiliary variable.* PhD dissertation, Iowa State University.