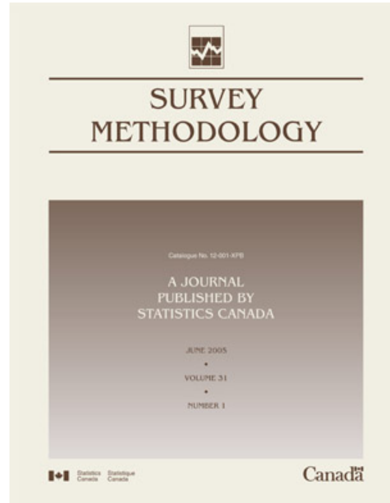
Catalogue no. 12-001-XIE

Survey Methodology

June 2002



Statistics Canada Statistique Canada

Canada

How to obtain more information

Specific inquiries about this product and related statistics or services should be directed to: Business Survey Methods Division, Statistics Canada, Ottawa, Ontario, K1A 0T6 (telephone: 1-800-263-1136).

For information on the wide range of data available from Statistics Canada, you can contact us by calling one of our toll-free numbers. You can also contact us by e-mail or by visiting our website.

National inquiries line	1-800-263-1136
National telecommunications device for the hearing impaired	1-800-363-7629
Depository Services Program inquiries	1-800-700-1033
Fax line for Depository Services Program	1-800-889-9734
E-mail inquiries	infostats@statcan.ca
Website	www.statcan.ca

Information to access the product

This product, catalogue no. 12-001-XIE, is available for free. To obtain a single issue, visit our website at www.statcan.ca and select Publications.

Standards of service to the public

Statistics Canada is committed to serving its clients in a prompt, reliable and courteous manner and in the official language of their choice. To this end, the Agency has developed standards of service that its employees observe in serving its clients. To obtain a copy of these service standards, please contact Statistics Canada toll free at 1 800 263-1136. The service standards are also published on www.statcan.ca under About us > Providing services to Canadians.

Statistics Canada

Business Survey Methods Division

Survey Methodology

June 2002

Published by authority of the Minister responsible for Statistics Canada

October 2006

Catalogue no. 12-001-XIE
ISSN 1492-0921

Frequency: semi-annual

Ottawa

Cette publication est disponible en français sur demande (nº 12-001-XIF au catalogue).

Note of appreciation

Canada owes the success of its statistical system to a long-standing partnership between Statistics Canada, the citizens of Canada, its businesses, governments and other institutions. Accurate and timely statistical information could not be produced without their continued cooperation and goodwill.

Survey Methodology, June 2002
Vol. 28, No. 1, pp. 5-23
Statistics Canada, Catalogue No. 12-001-XIE

Regression Estimation for Survey Samples

Wayne A. Fuller [1]

Abstract

Regression and regression related procedures have become common in survey estimation. We review the basic properties of regression estimators, discuss implementation of regression estimation, and investigate variance estimation for regression estimators. The role of models in constructing regression estimators and the use of regression in nonresponse adjustment are explored.

Key Words: Auxiliary information; Calibration; Least squares; Design consistency; Linear prediction.

1. Introduction

Design and estimation in survey sampling involve the use of information about the study population to construct efficient procedures. While design and estimation are intimately related, with estimators depending on the design, the two topics are often treated somewhat separately in the survey sampling literature. We follow tradition first studying estimation treating the design as given. The estimation task is to combine the available information about the population, with the sample data to produce good representations of characteristics of interest.

Regression estimation is one of the important procedures that use population information or information from a larger sample, to construct estimators with good efficiency. The information, sometimes called *auxiliary information*, may have been used in the design or may not have been available at the design stage. In surveys of the human population, the information often comes from official sources such as the national census. Similar sources may provide information for other types of surveys. For example, in a survey of land use the total surface area, the area owned by the national government, and the area in permanent water bodies may be available from national data archives.

Three distinct situations can be identified with respect to the nature of the auxiliary information that is available. In the first, the values of the auxiliary vector $\mathbf{x}$ are known for each element in the population at the time of sample selection. In this case the auxiliary variable can be used in designing the sample selection procedure.

In the second situation all values of the vector $\mathbf{x}$ are known, but a particular value cannot be associated with a particular element until the sample is observed. In this case, the auxiliary information cannot be used in design, but a wide range of estimation options are available once the observations are available. For example, the population census may give the age-sex distribution of the population, but a list of individuals and their characteristics is not available to non governmental institutions selecting samples.

In the third situation, only the population mean of $\mathbf{x}$ is known, or known for a large sample. In this case, the auxiliary information cannot be used in design and the estimation options are limited. For example the U.S. Department of Agriculture might release an estimate of the total number of animals of a particular type on farms on a particular date. Our discussion concentrates on this situation.

Two estimation situations can also be identified. In one, a single variable and a parameter, or a very small number of parameters, is under consideration. The analyst is willing to invest a great deal of effort in the analysis, has a well formulated population model, and is prepared to support the estimation procedure on the basis of the reasonableness of the model. In the second situation, a large number of analyses of a large number of variables is anticipated. No single model is judged adequate for all variables. The prototypical example of the second situation is the case in which a data set is prepared by the survey sampler to be analyzed by others. Because the person preparing the data set does not have knowledge of the analysis variables, emphasis is placed on the use of estimators that can be defended with minimal recourse to models.

Regression estimators fall in the class of linear estimators. Linear estimators have a particular advantage in survey sampling because once the weights are calculated they are appropriate for any analysis variable. Several properties of estimators will be examined in our discussion. Given a model, we accept the classical goal of minimizing the mean square error in a class of estimators. That class may be the class of linear estimators that are unbiased under the model, but the class may be further restricted.

Estimators that are scale and location invariant can be used in general settings. Mickey (1959) suggested that the term regression estimator be restricted to linear estimators that are location and scale invariant. While we may not adhere strictly to this definition, we support the distinction

1. Wayne Fuller, Emeritus Distinguished Professor, Iowa State University, 221 Snedecor Hall, Ames, IA 50011-1210, U.S.A.

between estimators that are location and scale invariant and those that are not. We consider location invariance to be important for sampling designs where the unit of interest for analysis is also the sampling unit. For cluster and two stage designs in which weights are constructed for primary sampling unit totals, location invariance is less important.

Models play an important role in the construction of regression estimators. It is desirable that the estimators retain good properties if the model specification is not exact. Therefore properties conditional on the realized finite population, as well as properties under the model, are important.

Linear estimators that reproduce the known means of the auxiliary variables are said to be calibrated. This is a desirable property in that, for example, the marginals of tables with an auxiliary variable as an analysis variable agree with known totals. If the auxiliary variable is of no analytic interest, then calibration is less important.

2. Background

The earliest references to the use of regression in survey sampling include Jessen (1942) and Cochran (1942). Regression in similar contexts would certainly have been used earlier and Cochran (1977, page 189) mentions a regression on leaf area by Watson (1937). It is interesting that Jessen's use of regression was essentially composite estimation where regression was used to improve estimates for two time points given samples at each point with some common elements in the two samples. Cochran (1942) gave the basic theory for regression in survey sampling relying heavily on linear model theory. He showed that the linear model did not need to hold in order for the regression estimator to perform well. He derived an expression for the $O(n^{-1})$ bias and an $O(n^{-2})$ approximation for the variance. He also showed that for the model with regression passing through the origin and error variances proportional to x, the ratio estimator is the generalized least squares estimator.

Regression estimation attracted theoretical interest in the 1950's, often in the form of studies of the bias. See Mickey (1959). Brewer (1963) is an early reference that considers linear estimation using a superpopulation model to determine an optimal procedure. He was concerned with finding the optimal design for the ratio estimator and discussed the possible conflict between an optimal design under the model and a design that is less model dependent. See also Brewer (1979). Royall (1970) argued for the use of models, that the conditional properties that are important are those conditional on the auxiliary information in the sample, and that the design should be chosen to optimize those properties. Royall and his coworkers, *e.g.*, Royall and Cumberland (1981), studied the conditional properties of regression estimators, conditional on the realized sample of auxiliary variables.

A great deal of research was conducted in the 1970's and 1980's on the general nature of the regression estimator in survey samples and on the degree to which the model prediction approach can be reconciled with the design perspective. Fuller (1973, 1975) gave the large sample properties of a vector of regression coefficients computed from a survey sample. Isaki (1970) studied regression estimators and the results were published in expanded versions in Isaki and Fuller (1982) and Fuller and Isaki (1981). It was shown that a regression estimator constructed under a model is design consistent for the population mean if the model contains certain variables. Cassel, Särndal and Wretman (1976) considered both model and design principles in estimator construction and suggested the term "generalized regression estimator" for design consistent estimators of the total of the form

$$\hat{T}_{y,\,\mathrm{GREG}} = \hat{T}_{y,\,\mathrm{HT}} + (T_{x,\,N} - \hat{T}_{x,\,\mathrm{HT}})\,\hat{\boldsymbol{\beta}},$$

where $\hat{T}_{y,\,\mathrm{HT}}$ and $\hat{T}_{x,\,\mathrm{HT}}$ are the Horvitz-Thompson estimators of the totals of y and x, respectively, $T_{x,\,N}$ is the know population total of x and $\hat{\boldsymbol{\beta}}$ is an estimated regression coefficient. Särndal (1980), Wright (1983), and Särndal and Wright (1984) discussed classes of regression estimators. The text by Särndal, Swensson and Wretman (1992) contains an extensive discussion of regression estimation and Mukhopadhyay (1993) is a review.

It was the 1970's before the use of regression for general purpose, multiple characteristic, surveys appeared and it was the 1990's before the use of regression weighting could be called widespread. An early use of regression weights was at Doane Agricultural Services Inc., now Doane Marketing Research. During 1971−1972 a readership study of farmers was conducted under the direction of Mr. John Wilkin in which 6,920 farmers responded. Weights for the respondents were constructed using regression procedures, where the controls came from the U.S. Agricultural Census and from Department of Agriculture sources. Doane provided financial support to Iowa State University to develop a regression weight generation program. To guarantee positive weights in the Doane study, observations with small weights were grouped and assigned a common weight. Grouping continued until the common weight was positive. Later computer programs used modifications of the Huang and Fuller (1978) procedure to guarantee positive weights. Doane has used regression weights for their syndicated market research studies since 1972.

Regression estimation was first used at Statistics Canada in 1988 for the Canadian Labour Force Survey. In 1992 regression estimation was used by the 1991 *Canadian Census of Population* to ensure that the weighted sum of variables collected via the long form (a one in five systematic sample of all households in Canada) was equal to known household and population totals as collected in the 1991 Census. See Bankier, Rathwell and Majkowski (1992) and Bankier, Houle and Luc (1997). The regression estimator is also the key component of the Generalized

Estimation System (GES) developed at Statistics Canada and used in numerous business and social surveys since its release in 1992. The methodology is described in Estevao, Hidiroglou and Särndal (1995). See also Hidiroglou, Särndal and Binder (1995). Regression estimation is now used to construct composite estimators for the Canadian Labour Force Survey. See Singh, Kennedy and Wu (2001), Gambino, Kennedy and Singh (2001) and Fuller and Rao (2001).

Bethlehem and Keller (1987) report on the use of regression estimation at the Netherlands Central Bureau of Statistics (now Statistics Netherlands) in a program called LIN WEIGHT. Nieuwenbrock, Renssen and Hofman (2000) describe the software package Bascula, that has replaced LIN WEIGHT. Deville, Särndal and Sautory (1993) describe a computer program CALMAR developed at Institut National de la Statistique et des Etudes Economiques (I.N.S.E.E.) that computes weights of the regression type with options for different objective functions. A program developed at Statistics Sweden and called CLAN97 is documented in Anderson and Nordberg (1998). Folsom and Singh (2000) discuss a procedure developed at the Research Triangle Institute.

3. The Classical Linear Model

The classical linear model is the foundation for survey regression estimation, but the survey situation requires certain adaptations. To introduce regression estimation for survey samples, we review the classical linear model. Assume

$$\begin{aligned} y_i &= \mathbf{x}_i \boldsymbol{\beta} + e_i, \quad i = 1, 2, ..., n, \\ e_i &\sim \mathbf{NI}(0, \sigma_e^2), \end{aligned} \tag{3.1}$$

where e_i is independent of the k–dimensional row vectors $\mathbf{x}_j$ for all i and j, and $\boldsymbol{\beta}$ is the unknown parameter column vector. We will also use matrix representations for the sample quantities. Thus, for a sample of n elements,

$$\mathbf{X}' = (\mathbf{x}_1', \mathbf{x}_2', ..., \mathbf{x}_n') \text{ and } \mathbf{y}' = (y_1, y_2, ..., y_n).$$

Given a sample of size n and treating the $\mathbf{x}_i$ as fixed, the best (minimum mean squared error) estimator of $\boldsymbol{\beta}$ is

$$\hat{\boldsymbol{\beta}} = \left(\sum_{i \in A} \mathbf{x}_i' \mathbf{x}_i \right)^{-1} \sum_{i \in A} \mathbf{x}_i' y_i = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y}, \tag{3.2}$$

where A is the set of indexes of the sample elements and we assume, as we will throughout, that the matrix to be inverted is nonsingular. If the e_i are not normally distributed, $\hat{\boldsymbol{\beta}}$ is the estimator with smallest variance in the class of linear unbiased estimators. The estimator of a linear combination of the coefficients, say $\theta_\alpha = \boldsymbol{\Sigma}_{j=1}^k \alpha_j \beta_j$, can be written as

$$\hat{\theta}_\alpha = \sum_{i \in A} w_{\alpha i} y_i$$

where the weights, $w_{\alpha i}$, minimize the Lagrangean

$$\sum_{i \in A} w_{\alpha i}^2 + \sum_{j=1}^k \lambda_j \left(\sum_{i \in A} w_{\alpha i} x_{ij} - \alpha_j \right)$$

and the λ_j are Lagrange multipliers. The variance of $\hat{\theta}_\alpha$ is

$$V\{\hat{\theta}_\alpha\} = V\left\{ \sum_{i \in A} w_{\alpha i} e_i \right\} = \sum_{i \in A} w_{\alpha i}^2 \sigma_e^2$$

because the weights are functions of the $\mathbf{x}_i$ and not of y_i.

The covariance matrix of $\hat{\boldsymbol{\beta}}$ is

$$\begin{aligned} V\{\hat{\boldsymbol{\beta}}\} &= \left(\sum_{i \in A} \mathbf{x}_i' \mathbf{x}_i \right)^{-1} V\left\{ \sum_{i \in A} \mathbf{b}_i' \right\} \left(\sum_{i \in A} \mathbf{x}_i' \mathbf{x}_i \right)^{-1} \\ &= V\left\{ \sum_{i \in A} \mathbf{c}_i \right\} \end{aligned} \tag{3.3}$$

where $\mathbf{b}_i' = \mathbf{x}_i' e_i$ and $\mathbf{c}_i = (\mathbf{X}' X)^{-1} \mathbf{x}_i' e_i$. Because e_i is independent of $\mathbf{x}_j$ for all i and j,

$$V\left\{ \sum_{i \in A} \mathbf{b}_i' \right\} = \sum_{i \in A} V\{\mathbf{b}_i'\} = \sum_{i \in A} \mathbf{x}_i' \mathbf{x}_i \sigma_e^2$$

and we obtain the familiar expression,

$$V\{\hat{\boldsymbol{\beta}}\} = \left(\sum_{i \in A} \mathbf{x}_i' \mathbf{x}_i \right)^{-1} \sigma_e^2.$$

The usual unbiased estimator of the covariance matrix of $\hat{\boldsymbol{\beta}}$ is obtained by replacing σ_e^2 with the unbiased estimator of σ_e^2 obtained as the mean square of the residuals, $\hat{e}_i = y_i - \mathbf{x}_i \hat{\boldsymbol{\beta}}$. An estimator of the covariance matrix that estimates $V\{\boldsymbol{\Sigma}_{i \in A} \mathbf{b}_i'\}$ directly is

$$\begin{aligned} \tilde{V}_b\{\hat{\boldsymbol{\beta}}\} &= \left(\sum_{i \in A} \mathbf{x}_i' \mathbf{x}_i \right)^{-1} \sum_{i \in A} \hat{\mathbf{b}}_i' \hat{\mathbf{b}}_i \left(\sum_{i \in A} \mathbf{x}_i' \mathbf{x}_i \right)^{-1} \\ &= \sum_{i \in A} \hat{\mathbf{c}}_i' \hat{\mathbf{c}}_i, \end{aligned} \tag{3.4}$$

where $\hat{\mathbf{b}}_i' = \mathbf{x}_i' \hat{e}_i$ and $\hat{\mathbf{c}}_i = (\mathbf{X}' X)^{-1} \mathbf{x}_i' \hat{e}_i$. In the same way

$$\hat{V}_b\{\hat{\theta}_\alpha\} = \sum_{i \in A} w_{\alpha i}^2 \hat{e}_i^2 \tag{3.5}$$

is a linear combination of the elements of (3.4) and is a consistent estimator of $V\{\hat{\theta}_\alpha\}$. The estimator (3.4) is a consistent estimator of $V\{\hat{\boldsymbol{\beta}}\}$ when the covariance matrix of the e_i is a diagonal matrix with bounded elements. Thus it is a more robust estimator. However, the estimator (3.4) is biased downward because the variance of $\hat{e}_i$ is usually less than the variance of e_i. Two methods are available for reducing the bias. The first is to make a degrees-of-freedom adjustment by multiplying $\tilde{V}_b\{\hat{\boldsymbol{\beta}}\}$ by $(n-k)^{-1} n$, where k is the dimension of $\mathbf{x}_i$. An alternative adjustment is to replace $\hat{e}_i$ with

$$\tilde{e}_i = (1 - \psi_{ii})^{-0.5} \hat{e}_i,$$

where ψ_{ii} is the i^{th} diagonal element of $\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$. See Horn, Horn and Duncan (1975), Royall and Cumberland (1978) and Cook and Weisberg (1982, section 2.2).

If we observe the value $\mathbf{x}_i$ for an element, but do not observe y_i, then the best predictor of y_i for that element is $\hat{y}_i = \mathbf{x}_i \hat{\boldsymbol{\beta}}$. Likewise, if we know the sum of $\mathbf{x}_i$ for a set of $\mathbf{x}$'s, then the best predictor for the sum of the y_i is the sum of $\mathbf{x}_i \hat{\boldsymbol{\beta}}$. Thus, given a set of N elements that satisfy model (3.1), a set of observations $(y_i, \mathbf{x}_i)$ on a subset denoted by A, and the known values of $\mathbf{x}_i$ for the remaining $N-n$ elements,

$$\hat{Y}_{N-n,\,\text{reg}} = \sum_{i \in \bar{A}} \hat{y}_i = \sum_{i \in \bar{A}} \mathbf{x}_i \hat{\boldsymbol{\beta}},$$

where $\bar{A}$ is the set of elements for which y is not observed, is the best predictor of the sum of the unobserved y's. See Goldberger (1962), Brewer (1963), Royall (1970), Harville (1976) and Graybill (1976, section 12.2). Hence

$$\hat{T}_{y,\,\text{reg}} = \sum_{i \in A} y_i + \hat{Y}_{N-n,\,\text{reg}} \tag{3.6}$$

is the best predictor for the total of N observations.

If the first element in the x–vector is always one, we can partition the x–vector as $\mathbf{x}_i = (1, \mathbf{x}_{1,i})$ and write the regression estimator of the mean as

$$\bar{y}_{\text{reg}} = N^{-1}\hat{T}_{y,\,\text{reg}} = \bar{\mathbf{x}}_N \hat{\boldsymbol{\beta}} = \bar{y}_n + (\bar{\mathbf{x}}_{1,N} - \bar{\mathbf{x}}_{1,n})\hat{\boldsymbol{\beta}}_1, \tag{3.7}$$

where $\hat{\boldsymbol{\beta}}$ of (3.2) is partitioned as $(\hat{\beta}_0, \hat{\boldsymbol{\beta}}_1')'$ and $(\bar{y}_n, \bar{\mathbf{x}}_n)$ is the vector of simple sample means. We call $\bar{\mathbf{x}}_N \hat{\boldsymbol{\beta}}$ the regression estimator of the mean.

Given the model (3.1), the expected value of the mean of y for the finite population of N elements generated by the model is $\bar{\mathbf{x}}_N \boldsymbol{\beta}$ and $\bar{\mathbf{x}}_N \hat{\boldsymbol{\beta}}$ is an unbiased estimator of the finite population mean. This, we believe, is the point at which regression estimation for the finite population mean under more complex designs begins.

4. Design Based Estimation

The development of this section treats the finite population as a sample realization from an infinite population. The use of such models has a long history in survey sampling. Some references through 1970 are Cochran (1939, 1942, 1946), Deming and Stephan (1941), Madow and Madow (1944), Yates (1949), Godambe (1955), Hájek (1959), Rao, Hartley, and Cochran (1962), Konijn (1962), Brewer (1963), Godambe and Joshi (1965), Hanurav (1966), Ericson (1969), Isaki (1970), and Royall (1970).

To discuss the large sample properties of regression estimators we consider sequences of finite populations and associated probability samples. The set of indices of the elements in the N^{th} finite population is $U_N = \{1, \ldots, N\}$, where $N = 1, 2, \ldots$. Associated with the i^{th} element of the N^{th} population is a row vector of characteristics $\mathbf{z}_{iN} = (y_{iN}, \mathbf{x}_{iN})$. Let

$$\mathbf{F}_N = [(y_{1N}, \mathbf{x}_{1N}), (y_{2N}, \mathbf{x}_{2N}), \ldots, (y_{NN}, \mathbf{x}_{NN})]$$

be the set of vectors for the N^{th} finite population. The subscript N on the vectors will often be omitted. The finite population mean is

$$\bar{\mathbf{z}}_N = (\bar{y}_N, \bar{\mathbf{x}}_N) = N^{-1}\sum_{i=1}^{N}(y_i, \mathbf{x}_i). \tag{4.1}$$

We denote the set of indices appearing in the sample selected from the N^{th} finite population by A_N.

When the finite population is a sample from an infinite superpopulation, the probability properties of a sample are determined by the properties of the superpopulation and the properties of the probability mechanism used to select the sample. One can consider the unconditional properties, the properties conditional on the particular finite population, or the properties conditional on some part of the realized sample.

Properties conditional on the finite population depend primarily on the survey design and are often called design properties. Thus an estimator $\hat{\theta}$ is said to be design consistent for the finite population parameter θ_N if, for all $\varepsilon > 0$,

$$\lim_{N,\, n \to \infty} \text{prob}\{|\hat{\theta} - \theta_N| > \varepsilon \,|\, \mathbf{F}_N\} = 0,$$

where the notation means that we condition on the realized finite population $\mathbf{F}_N$ and, hence, the probability is with respect to the design.

Assume the finite population is generated as independent selections from a superpopulation for which $E\{\mathbf{z}_i'\mathbf{z}_i\}$ is positive definite, where $\mathbf{z}_i = (y_i, \mathbf{x}_i)$. We define a superpopulation vector of least squares regression coefficients by

$$\boldsymbol{\beta} = [E\{\mathbf{x}_i'\mathbf{x}_i\}]^{-1} E\{\mathbf{x}_i' y_i\}. \tag{4.2}$$

Given a sample of n observations on $\mathbf{z}_i$ we define the $n \times (k+1)$ matrix $\mathbf{Z} = (\mathbf{y}, \mathbf{X})$ of observations, where the i^{th} row of $\mathbf{Z}$ is $(y_i, \mathbf{x}_i)$. If we assume the model

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}, \tag{4.3}$$

$$E\{\mathbf{u}, \mathbf{u}\mathbf{u}'\} = (\mathbf{0}, \boldsymbol{\Phi}),$$

the generalized least squares estimator of $\boldsymbol{\beta}$ is

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\boldsymbol{\Phi}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Phi}^{-1}\mathbf{y}. \tag{4.4}$$

The model (4.3) serves as motivation for estimators of the form (4.4) but we shall consider estimators where $\boldsymbol{\Phi}$ is a general symmetric positive definite weight matrix, not necessarily the covariance matrix of the errors.

We give the large sample properties of the vector of estimated regression coefficients (4.4) following Fuller (1975). See also Hidiroglou (1974), Scott and Wu (1981), and Robinson and Särndal (1983).

Assume the superpopulation has eighth moments and that the sample design is such that the error in the Horvitz-Thompson estimator of the mean is $O_p(n^{-1/2})$, where the Horvitz-Thompson estimator of the mean is

$$\overline{\mathbf{z}}_{\mathrm{HT}} = (\overline{y}_{\mathrm{HT}}, \overline{\mathbf{x}}_{\mathrm{HT}}) = N^{-1} \sum_{i \in A} \pi_i^{-1} \mathbf{z}_i \tag{4.5}$$

and π_i is the selection probability for element i. Then the error in the vector of regression coefficients is

$$\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_N \mid \mathbf{F}_N = \mathbf{Q}_{xxN}^{-1} \overline{\mathbf{b}}'_{\mathrm{HT}} + O_p(n^{-1}), \tag{4.6}$$

where

$$\boldsymbol{\beta}_N = \mathbf{Q}_{xxN}^{-1} \mathbf{Q}_{xyN}, \tag{4.7}$$

$$(\mathbf{Q}_{xxN}, \mathbf{Q}_{xyN}) = E\{(\hat{\mathbf{Q}}_{xx}, \hat{\mathbf{Q}}_{xy}) \mid \mathbf{F}_N\}, \tag{4.8}$$

$$(\hat{\mathbf{Q}}_{xx}, \hat{\mathbf{Q}}_{xy}) = n^{-1}(\mathbf{X}' \boldsymbol{\Phi}^{-1} \mathbf{X}, \mathbf{X}' \boldsymbol{\Phi}^{-1} \mathbf{y}),$$

$$\overline{\mathbf{b}}_{\mathrm{HT}} = N^{-1} \sum_{i \in A} \pi_i^{-1} \mathbf{b}_i, \tag{4.9}$$

$\mathbf{b}'_i = n^{-1} N \pi_i \zeta_i e_i$, $e_i = y_i - \mathbf{x}_i \boldsymbol{\beta}_N$, and ζ_i is column i of $\mathbf{X}' \boldsymbol{\Phi}^{-1}$. By (4.9) the error in the estimator of $\boldsymbol{\beta}_N$ is approximately the error in a Horvitz-Thompson estimator of the mean. In result (4.6), the $\boldsymbol{\beta}_N$ is defined as a function of the expected values of the sample quantities $(\hat{\mathbf{Q}}_{xx}, \hat{\mathbf{Q}}_{xy})$. Thus $\boldsymbol{\beta}_N$ is not necessarily the ordinary least squares finite population regression coefficient. The vector $\mathbf{b}_i$ of (4.9) is the generalization of the vector $\mathbf{b}_i$ of (3.3). If the limiting distribution of the properly standardized Horvitz-Thompson estimator is normal, and if there is a design consistent estimator of the variance of the Horvitz-Thompson estimator, then it is possible to construct tests and confidence intervals for the coefficients. Assume the design is such that

$$\mathbf{V}_{\bar{z}\bar{z}}^{-1/2}(\overline{\mathbf{z}}_{\mathrm{HT}} - \overline{\mathbf{z}}_N) \mid \mathbf{F}_N \xrightarrow{\mathrm{L}} N(\mathbf{0}, \mathbf{I}), \tag{4.10}$$

as $N, n \to \infty$, where $\mathbf{V}_{\bar{z}\bar{z}}$ is the covariance matrix of $\overline{\mathbf{z}}_{\mathrm{HT}} - \overline{\mathbf{z}}_N$. If $\mathbf{V}_{\bar{z}\bar{z}}$ is $O(n^{-1})$ and the estimator $\hat{\mathbf{V}}_{\bar{z}\bar{z}}$ is consistent for $\mathbf{V}_{\bar{z}\bar{z}}$, then

$$[\hat{V}\{\hat{\boldsymbol{\beta}}\}]^{-1/2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_N) \mid \mathbf{F}_N \xrightarrow{\mathrm{L}} N(\mathbf{0}, \mathbf{I}), \tag{4.11}$$

where

$$\hat{V}\{\hat{\boldsymbol{\beta}}\} = \hat{\mathbf{Q}}_{xx}^{-1} \hat{\mathbf{V}}_{\bar{b}\bar{b}} \hat{\mathbf{Q}}_{xx}^{-1}, = \hat{V}\{\overline{\mathbf{c}}'_{\mathrm{HT}}\}, \tag{4.12}$$

$\hat{\mathbf{V}}_{\bar{b}} = \hat{V}\{\overline{\mathbf{b}}'_{\mathrm{HT}}\}$ is the estimated design variance of $\overline{\mathbf{b}}_{\mathrm{HT}}$ calculated with $\hat{\mathbf{b}}'_i = n^{-1} N \pi_i \zeta'_i \hat{e}_i$ $\hat{e}_i = y_i - \mathbf{x}_i \hat{\boldsymbol{\beta}}$, and $\hat{V}\{\overline{\mathbf{c}}'_{\mathrm{HT}}\}$ is the estimated design variance of $\overline{\mathbf{c}}'_{\mathrm{HT}}$ calculated with $\hat{\mathbf{c}}'_i = \hat{\mathbf{Q}}_{xx}^{-1} \hat{\mathbf{b}}'_i$. The limiting properties hold for stratified samples and for stratified two stage samples under mild restrictions on the sequence of populations.

By analogy to (3.7), a regression estimator of the finite population mean is obtained by evaluating the estimated regression function at the population mean of $\mathbf{x}$ to obtain

$$\overline{y}_{\mathrm{reg}} = \overline{\mathbf{x}}_N \hat{\boldsymbol{\beta}}, \tag{4.13}$$

where $\hat{\boldsymbol{\beta}}$ is of the form (4.4) with a general $\boldsymbol{\Phi}$ matrix. The estimator can be written as $\mathbf{w}'\mathbf{y}$, where the vector of weights can be constructed by minimizing the Lagrangean

$$\mathbf{w}' \boldsymbol{\Phi} \mathbf{w} + (\mathbf{w}' \mathbf{X} - \overline{\mathbf{x}}_N) \boldsymbol{\lambda}$$

and $\boldsymbol{\lambda}$ is the vector of Lagrange multipliers.

If there is a column vectors $\boldsymbol{\gamma}$ such that

$$\mathbf{X} \boldsymbol{\gamma} = \boldsymbol{\Phi} \mathbf{D}_{\pi}^{-1} \mathbf{J} \tag{4.14}$$

for all possible samples, where $\mathbf{D}_{\pi} = \mathrm{diag}(\pi_1, \pi_2, ..., \pi_n)$ and $\mathbf{J}$ is an n–dimensional column vector of ones, then the regression estimator $\overline{\mathbf{x}}_N \hat{\boldsymbol{\beta}}$ of (4.13) with $\hat{\boldsymbol{\beta}}$ defined in (4.4) is a design consistent estimator of $\overline{y}_N$. It follows from (4.11) that

$$[\overline{\mathbf{x}}_N \hat{V}\{\hat{\boldsymbol{\beta}}\} \overline{\mathbf{x}}'_N]^{-1/2}(\overline{\mathbf{x}}_N \hat{\boldsymbol{\beta}} - \overline{y}_N) \xrightarrow{\mathrm{L}} N(0, 1). \tag{4.15}$$

The requirement of (4.14) that $\boldsymbol{\Phi} \mathbf{D}_{\pi}^{-1} \mathbf{J}$ be in the column space of $\mathbf{X}$ is crucial for design consistency. Simple ways to satisfy this requirement are to let one column of $\mathbf{X}$ be the column of ones and to use a multiple of $\mathbf{D}_{\pi}$ as $\boldsymbol{\Phi}$, or to let one column of $\mathbf{X}$ be the elements π_i^{-1} and set $\boldsymbol{\Phi} = \mathbf{I}$, or to let one column of $\mathbf{X}$ be the elements π_i and set $\boldsymbol{\Phi} = \mathbf{D}_{\pi}^2$. If $\mathbf{X}$ is composed of the single column vector with elements π_i and if $\boldsymbol{\Phi} = \mathbf{D}_{\pi}^2$, then the estimator (4.13) reduces to the Horvitz-Thompson estimator of (4.5) for fixed size designs. If $\mathbf{X} = \mathbf{J}$ and $\boldsymbol{\Phi} = \mathbf{D}_{\pi}$, the estimator (4.13) reduces to the ratio estimator,

$$\overline{y}_{\pi} = \left(\sum_{i \in A} \pi_i^{-1} \right)^{-1} \sum_{i \in A} \pi_i^{-1} y_i, \tag{4.16}$$

which is location and scale invariant.

To see the nature of the estimator when (4.14) is satisfied, let, with no loss of generality, $\mathbf{X} = (\mathbf{x}_0, \mathbf{X}_1)$, where $\mathbf{x}_0 = \boldsymbol{\Phi} \mathbf{D}_{\pi}^{-1} \mathbf{J}$ and $\mathbf{x}_i = (x_{0,i}, \mathbf{x}_{1,i})$. Then

$$\overline{y}_{\mathrm{reg}} = \overline{x}_{0,N} \overline{x}_{0,\pi}^{-1} \overline{y}_{\pi} + (\overline{\mathbf{x}}_{1,N} - \overline{x}_{0,N} \overline{x}_{0,\pi}^{-1} \overline{\mathbf{x}}_{1,\pi}) \hat{\boldsymbol{\beta}}_1, \tag{4.17}$$

where

$$\hat{\boldsymbol{\beta}}_1 = [(\mathbf{X}_1 - \mathbf{x}_0 \hat{\boldsymbol{\mu}}_{x1})' \boldsymbol{\Phi}^{-1} (\mathbf{X}_1 - \mathbf{x}_0 \hat{\boldsymbol{\mu}}_{x1})]^{-1}$$
$$\times (\mathbf{X}_1 - \mathbf{x}_0 \hat{\boldsymbol{\mu}}_{x1})' \boldsymbol{\Phi}^{-1} \mathbf{y},$$

$\hat{\boldsymbol{\mu}}_{x1} = \overline{x}_{0,\pi}^{-1} \overline{\mathbf{x}}_{1,\pi}$, and $(\overline{y}_{\pi}, \overline{\mathbf{x}}_{\pi})$ is defined in (4.16). The ratios, such as $\overline{x}_{0,\pi}^{-1} \overline{y}_{\pi}$, can also be written as ratios of Horvitz-Thompson estimators. If $\mathbf{J}$ is in the column space of $\mathbf{X}$, estimator (4.17) is location invariant. If $\boldsymbol{\Phi} = \mathbf{D}_{\pi}$, then $\overline{x}_{0,\pi}^{-1} \overline{x}_{0,N} = 1$, and

$$\overline{y}_{\mathrm{reg}} = \overline{\mathbf{x}}_N \hat{\boldsymbol{\beta}} = \overline{y}_{\pi} + (\overline{\mathbf{x}}_{1,N} - \overline{\mathbf{x}}_{1,\pi}) \hat{\boldsymbol{\beta}}_1, \tag{4.18}$$

where

$$\hat{\boldsymbol{\beta}}_1 = \left[\sum_{i \in A} (\mathbf{x}_{1,i} - \bar{\mathbf{x}}_{1,\pi})' \pi_i^{-1} (\mathbf{x}_{1,i} - \bar{\mathbf{x}}_{1,\pi}) \right]^{-1} \times \sum_{i \in A} (\mathbf{x}_{1,i} - \bar{\mathbf{x}}_{1,\pi})' \pi_i^{-1} (y_i - \bar{y}_\pi). \tag{4.19}$$

Also, when $\boldsymbol{\Phi} = \mathbf{D}_\pi$, the $\boldsymbol{\beta}_N$ of (4.7) is the population regression coefficient

$$\boldsymbol{\beta}_N \left[\sum_{i \in U} \mathbf{x}_i' \mathbf{x}_i \right]^{-1} \sum_{i \in U} \mathbf{x}_i' y_i. \tag{4.20}$$

Because the regression estimator of the mean is a linear combination of regression coefficients, it is a regression coefficient for a linear combination of the original $x-$ variables. To see this, let $\mathbf{x}_i = (x_{0,i}, \mathbf{x}_{1,i}) = (1, \mathbf{x}_{1,i})$, and define a new vector with one in the first position and a second vector with population mean equal to zero obtained by subtracting the original population mean $\bar{\mathbf{x}}_{1,N}$ from the original $\mathbf{x}_{1,i}$ vector. Let $\mathbf{q}_i = (1, \mathbf{x}_{1,i} - \bar{\mathbf{x}}_{1,N})$ be the transformed vector. Then the transformed regression model is

$$y_i = \mathbf{q}_i \boldsymbol{\gamma} + e_i, \tag{4.21}$$

where the finite population coefficient vector is

$$\boldsymbol{\gamma}_N = (\bar{y}_N, \boldsymbol{\beta}_{1,N}')' = \left(\sum_{i \in U} \mathbf{q}_i' \mathbf{q}_i \right)^{-1} \sum_{i \in U} \mathbf{q}_i' y_i. \tag{4.22}$$

The expression for the regression estimator of the mean becomes

$$\bar{y}_{\text{reg}} = \bar{\mathbf{q}}_N \hat{\boldsymbol{\gamma}} = \hat{\gamma}_0, \tag{4.23}$$

where $\hat{\boldsymbol{\gamma}}$ is obtained from (4.4) with $\mathbf{q}_i$ replacing $\mathbf{x}_i$. Because the estimator is a linear estimator of the form $\mathbf{w}' \mathbf{y}$, we can write

$$\bar{y}_{\text{reg}} = \sum_{i \in A} w_i y_i = \sum_{i \in A} \pi_i^{-1} g_i y_i, \tag{4.24}$$

where $w_i = \pi_i^{-1} g_i$. Furthermore, the estimated variance from (4.12) is

$$\hat{V}\{\bar{y}_{\text{reg}}\} = \hat{V}\{\hat{\gamma}_0\} = \hat{V}\left\{ \sum_{i \in A} \pi_i^{-1} (g_i \hat{e}_i) \right\}, \tag{4.25}$$

where it is understood that the estimated design variance of (4.25) is computed for the variable $g_i \hat{e}_i$, $\hat{e}_i = y_i - \mathbf{x}_i \hat{\boldsymbol{\beta}}$, and $\hat{\boldsymbol{\beta}}$ is defined in (4.4) The variance estimator (4.25) is a direct generalization of expression (3.5). By transforming the variables so that the population mean of the auxiliary vector is zero, the first element of the regression vector is the regression estimator of the mean and the first element of (4.12) is an estimator of the variance of the regression estimator that contains a component due to estimating $\boldsymbol{\beta}$. This was pointed out in Hidiroglou, Fuller, and Hickman (1978). Also, see Särndal (1982). Särndal, Swensson and Wretman (1989) suggested the g–factor terminology for the calculation of the estimated variance of a regression estimated total.

From (4.17), we can write

$$\begin{aligned} \bar{y}_{\text{reg}} &= \bar{x}_{0,N} \bar{x}_{0,\pi}^{-1} \left[\bar{y}_\pi - \bar{\mathbf{x}}_{1,\pi} \boldsymbol{\beta}_{1N} - (\bar{y}_N - \bar{\mathbf{x}}_{1,N} \boldsymbol{\beta}_{1N}) \right] \\ &\quad + O_p(n^{-1}), \\ &= \bar{e}_\pi + O_p(n^{-1}), \end{aligned}$$

where $e_i = y_i - \mathbf{x}_i \boldsymbol{\beta}$. Hence, the variance of the regression estimator can be estimated with

$$\hat{V}\{\bar{e}_\pi\} = \hat{V}\left\{ \left(\sum_{i \in A} \pi_i^{-1} \right)^{-1} \sum_{i \in A} \pi_i^{-1} \hat{e}_i \right\}, \tag{4.26}$$

where $\hat{e}_i = y_i - \mathbf{x}_i \hat{\boldsymbol{\beta}}$. Because (4.25) is as easy to compute as (4.26), and is applicable when $\bar{\mathbf{x}}_{1,\pi} - \bar{\mathbf{x}}_{1,N}$ is not $O_p(n^{-1/2})$, the estimator (4.25) is recommended.

The variance of the regression estimator can also be computed using the jackknife or other replication methods, and the use of replication methods is becoming more common. See Frankel (1971), Kish and Frankel (1974), Woodruff and Causey (1976), Royall and Cumberland (1978), and Duchesne (2000). Yung and Rao (1996) showed that (4.25) is identical to a jackknife linearization estimator for stratified multistage designs.

The approach to regression estimation associated with (4.18) and (4.19) falls completely within a design formulation. No models of the population, beyond the existence of moments, are used, through one might argue that one would only consider regression when one feels there is some linear correlation between $\mathbf{x}_{1,i}$ and y_i.

The estimator (4.19) is a very natural estimator because the estimated regression coefficient is a design consistent estimator of the population regression coefficient. It is mildly annoying that (4.18) does not always yield the smallest large sample design variance for the estimated mean. Treating $\hat{\boldsymbol{\beta}}_1$ of (4.18) as a fixed vector, the value that minimizes the variance of the linear combination of means is

$$\boldsymbol{\beta}_{1,\text{dopt}} = \left[V\{\bar{\mathbf{x}}_{1,\pi} \mid \mathbf{F}_N\} \right]^{-1} C\{\bar{\mathbf{x}}_{1,\pi}, \bar{y}_\pi \mid \mathbf{F}_N\}. \tag{4.27}$$

See Cochran (1977, page 201), Fuller and Isaki (1981), Montanari (1987, 1999) and Rao (1994). If there is a design consistent estimator of the variance of $\bar{\mathbf{x}}_{1,\pi}$, then the $\boldsymbol{\beta}_{1,d}$ that minimizes the estimated variance

$$\hat{V}\{\bar{y}_\pi - \bar{\mathbf{x}}_{1,\pi} \boldsymbol{\beta}_{1,d}\}, \tag{4.28}$$

denoted by $\hat{\boldsymbol{\beta}}_{1,\text{dopt}}$, is a consistent estimator of $\boldsymbol{\beta}_{1,\text{dopt}}$. It follows that the estimator

$$\bar{y}_{d,\text{reg}} = \bar{y}_\pi + (\bar{\mathbf{x}}_{1,N} - \bar{\mathbf{x}}_{1,\pi}) \hat{\boldsymbol{\beta}}_{1,\text{dopt}} \tag{4.29}$$

has the minimum limit variance for design consistent estimators of the form $\bar{y}_\pi + (\bar{\mathbf{x}}_{1,N} - \bar{\mathbf{x}}_{1,\pi}) \boldsymbol{\beta}_{1,d}$. Also

$$[\hat{V}\{\bar{e}_\pi\}]^{-1/2} (\bar{y}_{d,\text{reg}} - \bar{y}_N) \xrightarrow{\text{L}} N(0,1), \tag{4.30}$$

where $\hat{V}\{\bar{e}_\pi\}$ is the estimator of (4.26) constructed with $\hat{e}_i = y_i - \bar{y}_\pi - (\mathbf{x}_{1,i} - \bar{\mathbf{x}}_{1,\pi})\hat{\boldsymbol{\beta}}_{1,\text{dopt}}$.

In a large sample sense, (4.29) answers the question of how to construct a regression estimator with optimum design properties. In practice a number of questions remain. The estimator is obtained under the assumption of a large sample and a vector $\mathbf{x}$ of fixed dimension. In practice there may be a number of potential auxiliary variables and if a large number are included in the regression, terms excluded in the large sample approximation become important. This is particularly true for cluster samples where the number of primary sampling units in the sample is small. In such cases, the number of degrees-of-freedom in $\hat{V}\{\bar{\mathbf{x}}_{1,\pi}\}$ is small and the inverse can be unstable. These issues are discussed further in section 9.

The estimator $\boldsymbol{\beta}_{1,\text{dopt}}$ of (4.29) is linear in y for most designs. See Rao (1994). For example, for a stratified design with simple random sampling within strata,

$$\hat{C}\{\bar{\mathbf{x}}_{1,\pi}, \bar{y}_\pi\} = \sum_{h=1}^{H} K_h \sum_{j=1}^{n_h} (\mathbf{x}_{1,hj} - \bar{\mathbf{x}}_{1,h})'(y_{hj} - \bar{y}_h), \quad (4.31)$$

where

$$K_h = W_h^2(1 - f_h)(n_h - 1)^{-1} n_h^{-1} = N^{-2}\pi_h^{-2}(1 - f_h)(n_h - 1)^{-1} n_h,$$

$N^{-1}N_h = W_h$, N_h is the size of stratum h, $f_h = \pi_h = N_h^{-1} n_h$, and n_h is the sample size in stratum h. It follows that the weights associated with estimator (4.29) are

$$w_{hi} = N^{-1}\pi_h^{-1} + (\bar{\mathbf{x}}_{1,N} - \bar{\mathbf{x}}_{1,\pi}) \times \left[\sum_{t=1}^{H} K_t \sum_{j=1}^{n_h} (\mathbf{x}_{1,tj} - \bar{\mathbf{x}}_{1,t})'(\mathbf{x}_{1,tj} - \bar{\mathbf{x}}_{1,t})\right]^{-1} \times K_h(\mathbf{x}_{1,hi} - \bar{\mathbf{x}}_{1,h})'. \quad (4.32)$$

See also Särndal (1996). The weights of (4.32) can be constructed by minimizing $\sum_{hi\in A} w_{hi}^2 K_h^{-1}$ subject to the constraints

$$\sum_{i\in A_h} w_{hi} = N^{-1}N_h, \; h = 1, 2, ..., H,$$

and

$$\sum_{hi\in A} w_{hi}\,\mathbf{x}_{1,hi} = \bar{\mathbf{x}}_{1,N},$$

where A_h is the set of sample elements in stratum h.

The estimator of (4.19) with $\boldsymbol{\Phi} = \mathbf{D}_\pi$ is a function of Horvitz-Thompson estimators of population moments. The estimator (4.17) with $\boldsymbol{\Phi}^{-1} = \text{diag}\{K_t\}$, the diagonal matrix with K_t on the diagonal for elements in stratum t, and dummy variables for stratum effects, gives the estimator of the mean in the class

$$\bar{y}_{\text{reg}} = \bar{y}_\pi + (\bar{\mathbf{x}}_{1,N} - \bar{\mathbf{x}}_{1,\pi})\,\hat{\boldsymbol{\beta}}_1$$

with the smallest estimated design variance. If the true slopes in the strata are the same and if the selection probabilities are proportional to the square roots of the within-stratum variances, then the use of $\boldsymbol{\Phi} = \mathbf{D}_\pi^2$ gives a smaller small sample MSE than the use of $\boldsymbol{\Phi}^{-1} = \text{diag}\{K_t\}$ because the sum of $w_{hi}^2\sigma_h^2$ is smaller. Fuller and Isaki (1981) noted that the design-optimum estimator is often well approximated by the estimator constructed with $\boldsymbol{\Phi} = \mathbf{D}_\pi^2$.

We have introduced regression estimation for the mean, but it is often the totals that are estimated and totals that are used as controls. Consider the regression estimator of the total of y defined by

$$\hat{T}_{y,\text{reg}} = \hat{T}_{y,\pi} + (\mathbf{T}_{x,N} - \hat{\mathbf{T}}_{x,\pi})\,\hat{\boldsymbol{\beta}}_{y\cdot x}, \quad (4.33)$$

where $\mathbf{T}_{x,N}$ is the known total of $\mathbf{x}$ and $(\hat{T}_{y,\pi}, \hat{\mathbf{T}}_{x,\pi})$ is a vector of design consistent estimators of $(T_{y,N}, \mathbf{T}_{x,N})$. By analogy to (4.28), the estimator of the optimum $\boldsymbol{\beta}$ is

$$\hat{\boldsymbol{\beta}}_{y\cdot x} = [\hat{V}\{\hat{\mathbf{T}}_{x,\pi}\}]^{-1}\,\hat{C}\{\hat{\mathbf{T}}'_{x,\pi}, \hat{T}_{y,\pi}\}, \quad (4.34)$$

where $\hat{V}\{\hat{\mathbf{T}}_{x,\pi}\}$ is a design consistent estimator of the variance of $\hat{\mathbf{T}}_{x,\pi}$ and $\hat{C}(\hat{\boldsymbol{T}}'_{x,\pi}, \hat{\boldsymbol{T}}_{y,\pi})$ is a design consistent estimator of the covariance of $\hat{\mathbf{T}}_{x,\pi}$ and $\hat{T}_{y,\pi}$.

The estimator of the total is $N\bar{y}_{\text{reg}}$ for simple random sampling, but the exact equivalence may not hold in more complicated samples, because in such situations the estimated mean may be a ratio estimator. However, if the regression estimator of the two totals is constructed using (4.34), the ratio of the two estimated totals has large sample variance equal to that of the regression estimator of the mean. To see this write the error in the regression estimated totals of y and u as

$$\hat{T}_{y,\text{reg}} - T_{y,N} = \hat{T}_{y,\pi} - T_{y,N} + (\mathbf{T}_{x,N} - \hat{\mathbf{T}}_{x,\pi})\,\boldsymbol{\beta}_{y\cdot x,N} + O_p(Nn^{-1})$$

and

$$\hat{T}_{u,\text{reg}} - T_{u,N} = \hat{T}_{u,\pi} - T_{u,N} + (\mathbf{T}_{x,N} - \hat{\mathbf{T}}_{x,\pi})\,\boldsymbol{\beta}_{u\cdot x,N} + O_p(Nn^{-1}), \quad (4.35)$$

where we are assuming $\hat{T}_{y,\pi} - T_{y,N}$, $\hat{\boldsymbol{\beta}}_{y\cdot x} - \boldsymbol{\beta}_{y\cdot x,N}$ and the corresponding quantities for u, to be $O_p(Nn^{-1/2})$ and $O_p(n^{-1/2})$, respectively. Then the error in $\hat{T}_{u,\text{reg}}^{-1}\hat{T}_{y,\text{reg}}$ is

$$\hat{T}_{u,\text{reg}}^{-1}\hat{T}_{y,\text{reg}} - T_{u,N}^{-1}T_{y,N} = T_{u,N}^{-1}[(\hat{T}_{y,\pi} - T_{y,N}) - R_N(\hat{T}_{u,\pi} - T_{u,N}) + (\mathbf{T}_{x,N} - \hat{\mathbf{T}}_{x,\pi})(\boldsymbol{\beta}_{y\cdot x,N} - R_N\boldsymbol{\beta}_{z\cdot x,N})] + O_p(Nn^{-1}), \quad (4.36)$$

where $R_N = T_{u,N}^{-1} T_{y,N}$. If we construct the regression estimator for R_N starting with $\hat{R} = \hat{T}_{u,\pi}^{-1} \hat{T}_{y,\pi}$, we have

$$\hat{R}_{\text{reg}} = \hat{R} + (\mathbf{T}_{x,N} - \hat{\mathbf{T}}_{x,\pi})\, \hat{\boldsymbol{\beta}}_{R\cdot x}, \tag{4.37}$$

where

$$\hat{\boldsymbol{\beta}}_{R\cdot x} = [\hat{V}\{\hat{\mathbf{T}}_{x,\pi}\}]^{-1} \hat{C}\{\hat{\mathbf{T}}'_{x,\pi}, \hat{R}\}$$

and

$$\hat{C}\{\hat{\mathbf{T}}_{x,\pi}, \hat{R}\} = \hat{C}\{\hat{\mathbf{T}}_{x,\pi},\, T_{u,N}^{-1}(\hat{T}_{y,\pi} - R_N \hat{T}_{u,\pi})\}.$$

It follows that the large-sample-design-optimum coefficient for the ratio is $T_{u,N}^{-1}(\boldsymbol{\beta}_{y\cdot x,N} - R_N\, \boldsymbol{\beta}_{u\cdot x,N})$ and the ratio of design-optimum regression estimators is the large sample design-optimum regression estimator of the ratio.

5. Models and Regression Estimation

In this section we assume that the analyst postulates a detailed superpopulation model. Assume also that the sample is an unequal probability sample or (and) the specified error covariance structure is not a multiple of the identity matrix. Then, only in special cases will the design optimal estimator of (4.29) agree with the best estimator constructed under the model, conditioning on the sample $\mathbf{x}$ – values. To investigate this possible conflict, write the model for the population in matrix notation as

$$\begin{aligned} \mathbf{y}_U &= \mathbf{X}_U \boldsymbol{\beta} + \mathbf{e}_U \\ \mathbf{e}_U &\sim (\mathbf{0}, \boldsymbol{\Sigma}_{eeUU}), \end{aligned} \tag{5.1}$$

where $\mathbf{y}_U = (y_1, y_2, ..., y_N)'$, $\mathbf{e}_U = (e_1, e_2, ..., e_N)'$ and $\mathbf{X}_U = (\mathbf{x}'_1, \mathbf{x}'_2, ..., \mathbf{x}_N)'$. It is assumed that $\boldsymbol{\Sigma}_{eeUU}$ is known or known up to a multiple. The model for a sample of n observation is

$$\begin{aligned} \mathbf{y}_A &= \mathbf{X}_A \boldsymbol{\beta} + \mathbf{e}_A, \\ \mathbf{e}_A &\sim (\mathbf{0}, \boldsymbol{\Sigma}_{eeAA}), \end{aligned}$$

where $\mathbf{y}_A = (y_1, y_2, ..., y_n)'$, $\mathbf{e}_A = (e_1, e_2, ..., e_n)'$, $\mathbf{X}_A = (\mathbf{x}'_1, \mathbf{x}'_2, ..., \mathbf{x}'_n)'$, and we index the sample elements by 1, 2, ..., n, for convenience. We have used the subscript U to identify population quantities, and the subscript A to identify sample quantities, but we will often omit the subscript A to simplify the notation. For example, we may sometimes write the $n \times n$ covariance matrix as $\boldsymbol{\Sigma}_{ee}$. The unknown finite population mean is

$$\bar{y}_N = \bar{\mathbf{x}}_N \boldsymbol{\beta} + \bar{\mathbf{e}}_N. \tag{5.2}$$

Under model (5.1), the best linear, conditionally unbiased predictor of $\theta_N = \bar{y}_N$, conditional on $\mathbf{X}$ is

$$\hat{\theta} = N^{-1} \left[\begin{aligned} &\sum_{i \in A} y_i + (N-n)\, \bar{\mathbf{x}}_{N-n} \hat{\boldsymbol{\beta}} \\ &\quad + \mathbf{J}'_{N-n}\, \boldsymbol{\Gamma}_{\bar{A}A} (\mathbf{y}_A - \mathbf{X}_A\, \hat{\boldsymbol{\beta}}) \end{aligned} \right], \tag{5.3}$$

where $\boldsymbol{\Gamma}_{\bar{A}A} = \boldsymbol{\Sigma}_{ee\bar{A}A} \boldsymbol{\Sigma}_{eeAA}^{-1}$, $\bar{\mathbf{x}}_{N-n} = (N-n)^{-1}(N\bar{\mathbf{x}}_N - n\bar{\mathbf{x}}_n)$, $\boldsymbol{\Sigma}_{ee\bar{A}A} = E\{\mathbf{e}_{\bar{A}}\, \mathbf{e}'_A\}$,

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}' \boldsymbol{\Sigma}_{eeAA}^{-1} \mathbf{X})^{-1} \mathbf{X}' \boldsymbol{\Sigma}_{eeAA}^{-1}\, \mathbf{y},$$

$\mathbf{e}_{\bar{A}} = (e_{n+1}, e_{n+2}, ..., e_N)$, $\mathbf{J}_{N-n}$ is an $N-n$ dimensional column vector of ones, $\bar{\mathbf{x}}_n$ is the simple sample mean, and $\bar{A}$ is the set of elements in U that are not in A. See Royall (1976). Under the model,

$$\hat{\theta} - \bar{y}_N = \mathbf{C}_{x\bar{A}}\, (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) + N^{-1}\, \mathbf{J}'_{N-n} (\boldsymbol{\Gamma}_{\bar{A}A}\, \mathbf{e}_A - \mathbf{e}_{\bar{A}})$$

and

$$\begin{aligned} V\{\hat{\theta} - \bar{y}_N \mid \mathbf{X}_A\} &= \mathbf{C}_{x\bar{A}}\, V\{\hat{\boldsymbol{\beta}}\}\, \mathbf{C}'_{x\bar{A}} \\ &\quad + N^{-2}\, \mathbf{J}'_{N-n} (\boldsymbol{\Sigma}_{ee\bar{A}\bar{A}} - \boldsymbol{\Gamma}_{\bar{A}A} \boldsymbol{\Sigma}_{eeA\bar{A}})\, \mathbf{J}_{N-n}, \end{aligned} \tag{5.4}$$

where

$$\mathrm{C}_{x\bar{A}} = N^{-1}[(N-n)\, \bar{\mathbf{x}}_{N-n} - \mathbf{J}'_{N-n}\, \boldsymbol{\Gamma}_{\bar{A}A}\, \mathbf{X}_A].$$

Design consistency of estimator (5.3) and the situations in which the model estimator reduces to the Horvitz-Thompson estimator have been considered by, among others, Isaki (1970), Royall (1970, 1976), Scott and Smith (1974), Cassel, Särndal, and Wretman (1976, 1979, 1983), Zyskind (1976), Tallis (1978), Isaki and Fuller (1982), Wright (1983), Pfefferman (1984), Tam (1986), Brewer, Hanif and Tam (1988), Montanari (1999), and Gerow and McCulloch (2000).

The estimator (5.3) reduces to $\bar{\mathbf{x}}_N\, \hat{\boldsymbol{\beta}}$ if there is an η such that

$$\mathbf{X}_A \eta = \boldsymbol{\Sigma}_{eeAA}\, \mathbf{J}_n + \boldsymbol{\Sigma}_{eeA\bar{A}}\, \mathbf{J}_{N-n}, \tag{5.5}$$

for all samples with positive probability. If there is also $\boldsymbol{\gamma}$ such that

$$\mathbf{X}_A\, \boldsymbol{\gamma} = \boldsymbol{\Sigma}_{eeAA}\, \mathbf{D}_{\pi}^{-1}\, \mathbf{J}_n \tag{5.6}$$

for all samples with positive probability, then $\hat{\theta}$ of (5.3) is design consistent, where $\mathbf{D}_{\pi}$ was defined for (4.14). Given a $\mathbf{k}$ such that

$$\mathbf{X}_A\, \mathbf{k} = \boldsymbol{\Sigma}_{eeAA} (\mathbf{D}_{\pi}^{-1}\, \mathbf{J}_n - \mathbf{J}_n)\, \boldsymbol{\Sigma}_{eeA\bar{A}}\, \mathbf{J}_{N-n}, \tag{5.7}$$

then $\hat{\theta}$ of (5.3) is expressible as

$$\hat{\theta} = \bar{y}_{\pi} + (\bar{\mathbf{x}}_N - \bar{\mathbf{x}}_{\pi})\, \hat{\boldsymbol{\beta}} \tag{5.8}$$

and if the design is such that $\bar{y}_{\pi}$ is design consistent for $\bar{y}_N$, $\hat{\theta}$ of (5.8) is design consistent for $\bar{y}_N$.

We call a regression model of the form (5.1) for which (5.5) and (5.6), or (5.7), holds a full model. If (5.6) or (5.7) does not hold, we call the model a reduced model or a restricted model. We cannot expect the conditions for a full model to hold for every analysis variable in a general purpose survey because $\boldsymbol{\Sigma}_{ee}$ will be different for different

y's. Therefore, given a reduced model, one might search for a good model estimator in the class of design consistent estimators.

To construct a design consistent estimator of the form $\bar{\mathbf{x}}_N \hat{\boldsymbol{\beta}}$ when model (5.1) is a reduced model, we can add a vector satisfying (5.7) to the $\mathbf{X}$–matrix to create a full model. There are two possible situations associated with this approach. In the first, the population mean (or total) of the added variable is known. With known mean, one can construct the usual regression estimator and the usual design variance estimation formulas are appropriate.

To describe an estimation procedure for the situation in which the population mean of the added variable is not known, let $\mathbf{q} = (q_1, q_2, ..., q_n)'$ denote the added vector, where $\mathbf{q}$ is the vector on the right side of the equality in (5.7). Let $\mathbf{H} = (\mathbf{X}, \mathbf{q})$, where $\mathbf{X}$ is the matrix of auxiliary variables with known population mean vector, $\bar{\mathbf{x}}_N$. We write the full model for the sample as

$$\mathbf{y} = \mathbf{Z}\boldsymbol{\beta}_{y \cdot h} + \mathbf{e}, \qquad (5.9)$$

where $\mathbf{e} \sim (\mathbf{0}, \boldsymbol{\Sigma}_{ee})$. The best linear conditionally unbiased estimator of $\boldsymbol{\beta}_{y \cdot h}$ is

$$\hat{\boldsymbol{\beta}}_{y \cdot h} = (\mathbf{H}' \boldsymbol{\Sigma}_{ee}^{-1} \mathbf{H})^{-1} \mathbf{H}' \boldsymbol{\Sigma}_{ee}^{-1} \mathbf{y}. \qquad (5.10)$$

If the coefficient for $\mathbf{q}$ in (5.9) is not zero, it is not possible to construct a conditionally unbiased estimator of $\bar{\mathbf{h}}_N \boldsymbol{\beta}_{y \cdot h}$ because the $\bar{q}_N$ component of $\bar{\mathbf{h}}_N$ is unknown. However, because $\hat{\boldsymbol{\beta}}_{y \cdot h}$ is unbiased for $\boldsymbol{\beta}_{y \cdot h}$, it is possible to construct a conditionally unbiased estimator of any linear function of $\boldsymbol{\beta}_{y \cdot h}$. Thus, it is natural to replace the unknown $\bar{q}_N$ with the "best available" estimator of $\bar{q}_N$, and a reasonable choice is the regression estimator,

$$\bar{q}_{\text{reg}} = \bar{q}_\pi + (\bar{\mathbf{x}}_N - \bar{\mathbf{x}}_\pi)\, \hat{\boldsymbol{\beta}}_{q \cdot x}, \qquad (5.11)$$

where $\hat{\boldsymbol{\beta}}_{q \cdot x} = (\mathbf{X}' \boldsymbol{\Sigma}_{ee}^{-1} \mathbf{X})^{-1} \mathbf{X}' \boldsymbol{\Sigma}_{ee}^{-1} \mathbf{q}$. Then the estimator (5.3) becomes

$$\hat{\theta} = \bar{y}_\pi + [(\bar{\mathbf{x}}_N, \bar{q}_{\text{reg}}) - (\bar{\mathbf{x}}_\pi, \bar{q}_\pi)]\, \hat{\boldsymbol{\beta}}_{y \cdot h} \qquad (5.12)$$

The estimator (5.12) can be expressed in the familiar regression estimator form,

$$\bar{y}_{\text{reg}} = \bar{y}_\pi + (\bar{\mathbf{x}}_N - \bar{\mathbf{x}}_\pi)\, \hat{\boldsymbol{\beta}}_{y \cdot x}. \qquad (5.13)$$

That is, the regression estimator of the finite population mean of y based on the full model, but with the mean of q_i unknown and estimated with the regression estimator, is the regression estimator with $\boldsymbol{\beta}_{y \cdot x}$ estimated by the generalized least squares regression of y on $\mathbf{x}$ using the covariance matrix $\boldsymbol{\Sigma}_{ee}$. See Park (2002). The estimator is conditionally model unbiased under the reduced model containing only $\mathbf{x}$ if the reduced model is true. If the population coefficient for q_i is not zero, the reduced model is not true. Then the estimator is conditionally model biased, but the estimator is unbiased for the finite population mean under the full model and an unbiased design, because

$$\begin{aligned} E\{\bar{y}_{\text{reg}} - \bar{y}_N\} &= E\{E[\bar{y}_{\text{reg}} - \bar{y}_N \mid \mathbf{H}]\} \\ &= E\{(0, \bar{q}_{\text{reg}} - \bar{q}_N)\, \boldsymbol{\beta}_{y \cdot h} \doteq 0, \end{aligned} \qquad (5.14)$$

where $\bar{y}_{\text{reg}}$ is defined in (5.12) and the approximation is due to the approximate design expectation of the regression estimator $\bar{q}_{\text{reg}}$.

The estimator (5.13) is a linear estimator, where the vector of weights, $\mathbf{w}$, minimizes the Lagrangean

$$\mathbf{w}' \boldsymbol{\Sigma}_{ee}\, \mathbf{w} + [\mathbf{w}' \mathbf{H} - (\bar{\mathbf{x}}_N, \bar{q}_{\text{reg}})]\, \boldsymbol{\lambda}. \qquad (5.15)$$

The estimator is location invariant if the column of ones is in the column space of $\mathbf{X}$.

Because the variable q is the variable whose omission from the full model can produce a bias, it seems prudent to test the coefficient of q before using the reduced model to construct an estimator for the mean of y. This can be done using a model estimator of the variance,

$$\hat{V}\{\hat{\boldsymbol{\beta}}_{y \cdot h} \mid \mathbf{H}\} = (\mathbf{H}' \hat{\boldsymbol{\Sigma}}_{ee}^{-1} \mathbf{H})^{-1}$$

or using the design estimator of variance of (4.12). See Du Mouchel and Duncan (1983) and Fuller (1984).

A working specification for $\boldsymbol{\Sigma}_{ee}$ may be particularly appropriate for two-stage samples, see Royall (1976, 1986) and Montanari (1987). A reasonable model is that in which there is common correlation among items in the same primary sampling unit and zero correlation between units in different primary sampling units. Because the associated $\boldsymbol{\Sigma}_{ee}$ is block diagonal of a particular form, it is relatively easy to invert and hence the estimator based on such a working $\boldsymbol{\Phi}$ is relatively easy to construct. The regression estimator using a $\boldsymbol{\Phi}$ with a non zero correlation for units in the same primary sampling unit is a combination of the estimator based on primary sampling unit totals and that based on elements. See Fuller and Battese (1973). Thus, the use of such a $\boldsymbol{\Phi}$ can avoid variance problems associated with the use of primary sampling unit totals.

6. Maximum Likelihood and Raking Ratio

The theoretical foundation for the regression estimators discussed in section 3 and section 4 is maximum likelihood estimation for the linear model with normal errors. We now consider the likelihood for multinomial variables. Given a simple random sample from a multinomial defined by the entries in a two way table, the logarithm of the likelihood, except for a constant, is

$$\sum_{i=1}^{r} \sum_{j=1}^{c} a_{ij} \log p_{ij}, \qquad (6.1)$$

where a_{ij} is the estimated fraction in cell ij, p_{ij} is the population fraction in cell ij, r is the number of rows, and c is the number of columns. If (6.1) is maximized subject to the restriction $\boldsymbol{\Sigma\Sigma}\, p_{ij} = 1$, one obtains the maximum likelihood estimators $\hat{p}_{ij} = a_{ij}$. If the marginal row fractions $p_{i\cdot, N}$ and the marginal column fractions $p_{\cdot j, N}$ are known, it is natural to maximize the likelihood subject to these constraints by using the Lagrangean

$$\sum_{i=1}^{r} \sum_{j=1}^{c} a_{ij} \log p_{ij} + \sum_{i=1}^{r} \lambda_i \left(\sum_{j=1}^{c} p_{ij} - p_{i\cdot, N} \right) + \sum_{j=r+1}^{r+c} \lambda_j \left(\sum_{i=1}^{r} p_{ij} - p_{\cdot j, N} \right), \quad (6.2)$$

where λ_i, $i = 1, 2, ..., r$, are for the row restrictions and λ_j, $j = 1, 2, ..., c$, are for the column restrictions. There is no explicit expression for the solution to (6.2) and there may be no solution if there are too many empty cells. A procedure that produces estimates close to the maximum likelihood solution is that called *raking ratio or iterative proportional fitting*. The procedure iterates, first making ratio adjustments for the row restrictions, then making ratio adjustments for the column restrictions, then making a ratio adjustments for the row restrictions, *etc.* The method is generally credited to Deming and Stephan (1940). See, for example, Bishop, Fienberg and Holland (1975, Chapter 3).

Deville and Särndal (1992) considered a class of objective functions of the form $\mathbf{\Sigma}_{i \in A} G(w_i, \alpha_i)$, where $G(w, \alpha)$ is a measure of distance between an initial weight α_i and a final weight w_i. The objective function is minimized subject to the constraints

$$\sum_{i \in A} w_i \mathbf{x}_i = \bar{\mathbf{x}}_N. \quad (6.3)$$

Deville and Särndal (1992) used the term *calibrated* to describe weights satisfying (6.3). If the initial weight is $\alpha_i = (\mathbf{\Sigma} \pi_j^{-1})^{-1} \pi_i^{-1}$ and if one is the first element of $\mathbf{x}_i$, the solution to the minimization problem is approximated by a regression estimator of the mean of the form

$$\bar{y}_{\text{reg}} = \bar{y}_\pi + (\bar{\mathbf{x}}_N - \bar{\mathbf{x}}_\pi) \hat{\boldsymbol{\beta}}, \quad (6.4)$$

where

$$\hat{\boldsymbol{\beta}} = \left[\sum_{i \in A} \mathbf{x}_i' \varphi_{ii}^{-1} \mathbf{x}_i \right]^{-1} \sum_{i \in A} \mathbf{x}_i' \varphi_{ii}^{-1} y_i ,$$

and φ_{ii} is the second derivative of $G(w, \alpha)$ with respect to w evaluated at $(w, \alpha) = (\alpha_i, \alpha_i)$. Using this approach, Deville and Särndal (1992) showed that the maximum likelihood and raking ratio estimators have the same limiting distribution as the regression estimator (4.18) with $\mathbf{\Phi} = \mathbf{D}_\pi$. To obtain the raking ratio weights they used the objective function

$$\sum_{i \in A} [w_i \log \alpha_i^{-1} w_i + \alpha_i - w_i], \quad (6.5)$$

and to obtain the maximum likelihood weights they used the objective function

$$\sum_{i \in A} [w_i - \alpha_i - \alpha_i \log \alpha_i^{-1} w_i]. \quad (6.6)$$

Deville, Särndal and Sautory (1993) investigated four estimators in the class. Although weights constructed using different functions could differ considerably, the authors concluded that estimates were quite similar, a result consistent with the theory. Singh and Mohl (1996) and Théberge (1999, 2000) discuss estimators with the calibration property.

7. Population of Auxiliary Vectors Known at Estimation Step

If the **x**-vector is known for all of the population elements, the number of possible regression-type estimators is greatly expanded. Most procedures involve the fitting of an approximating function for the relationship between y and the auxiliary variables. The most used procedure is to assign the population elements to categories on the basis of the auxiliary data and to use these categories as post strata. This procedure is equivalent to approximating the expected value of y given $\mathbf{x}$ by a step function. The estimator is formally equivalent to the regression estimator (4.19) where the **x**–vector is a vector of indicator variables for post-stratum membership.

The application of the procedure often requires the development of criteria to use in forming the post strata. Typically the post strata are formed so that each post stratum contains a minimum number of sample elements and so that the weights for any post stratum are not overly large. Estimation with post strata and the formation of post strata have been studied by Fuller (1966), Holt and Smith (1979), Tremblay (1986) Kalton and Maligalig (1991), Little (1993), Eltinge and Yansaneh (1997), and Lazzeroni and Little (1998), among others. Holt and Smith (1979) argued for the use of a conditional variance estimator for post stratification.

Given the population of **x**–vectors, one can use the sample to estimate a functional relationship between y and $\mathbf{x}$ and then predict the unobserved y. If the procedure is to be design consistent, then a condition similar to (4.14) must hold. One way to ensure design consistency is to require the fitted model to satisfy

$$\sum_{i \in A} \pi_i^{-1} [y_i - f(\mathbf{x}_i, \hat{\boldsymbol{\beta}})] = 0, \quad (7.1)$$

where $f(\mathbf{x}_i, \hat{\boldsymbol{\beta}})$ is the model estimated value for the i^{th} observation.

Firth and Bennett (1998) pointed out that some nonlinear models satisfy (7.1). If the initial model does not satisfy (7.1), an estimated intercept term can be added to create an expanded full model,

$$\tilde{f}_F(\mathbf{x}_i; \hat{\boldsymbol{\beta}}) = f(\mathbf{x}_i; \hat{\boldsymbol{\beta}}) + \left(\sum_{i \in A} \pi_i^{-1} \right)^{-1} \sum_{i \in A} \pi_i^{-1} [y_i - f(\mathbf{x}_i; \hat{\boldsymbol{\beta}})].$$

This is a direct extension of the ideas of difference estimation to the nonlinear case. See Isaki (1970), Cassel, Särndal and Wretman (1976) and Wright (1983). A closely related approach was suggested by Wu and Sitter (2001) in which the fitted function $f(\mathbf{x}_i, \hat{\boldsymbol{\beta}})$ is used as the auxiliary variable in a linear regression estimator.

A number of "local" procedures, other than step functions, can be used to approximate the functional relationship between $\mathbf{x}$ and y. Spline functions and polynomials are linear models that fall within the class of section 4. Estimators that use some kind of local smoothing to estimate population quantities have been considered for finite populations from a model viewpoint by Kuo (1988), Dorfman (1993), Dorfman and Hall (1993), Chambers (1996), and Chambers, Dorfman and Wehrly (1993). Breidt and Opsomer (2000) showed that estimators based on local polynomial regression are design consistent. Firth and Bennett (1998) also considered local fit models.

8. Regression Estimation and Nonresponse

Regression estimation is frequently a part of procedures used to adjust data for unit nonresponse. Regression can be justified on the basis of a model such as (3.1) or on the basis that regression can adjust for unequal response probabilities. See Cassel, Särndal and Wretman (1979, 1983), Little (1982, 1986), Bethlehem (1988), Kott (1994), Fuller, Loughin and Baker (1994) and Fuller and An (1998).

Consider an estimator of the population regression vector of the form (4.4) with $\boldsymbol{\Phi} = \mathbf{D}_\pi$ constructed with the responding units. Denote the estimator by $\tilde{\boldsymbol{\beta}}$ and let p_i be the conditional probability of observing unit i given that the unit is selected for the sample. Then under regularity conditions, the estimator $\tilde{\boldsymbol{\beta}}$ is a consistent estimator of

$$\boldsymbol{\gamma}_N = \left(\sum_{i \in U} \mathbf{x}_i' \, p_i \, \mathbf{x}_i \right)^{-1} \sum_{i \in U} \mathbf{x}_i' \, p_i \, y_i. \tag{8.1}$$

The population mean of y can be expressed as

$$\bar{y}_N = \bar{\mathbf{x}}_N \, \boldsymbol{\gamma}_N + \bar{a}_N \tag{8.2}$$

where $a_i = y_i - \mathbf{x}_i \boldsymbol{\gamma}_N$ and $\bar{a}_N$ is the population mean of the a_i. The regression estimator $\bar{y}_{\text{reg}} = \bar{\mathbf{x}}_N \tilde{\boldsymbol{\beta}}$ will be consistent for $\bar{y}_N$ if the probability limit of $\bar{a}_N$ is zero. The probability limit of $\bar{a}_N$ will be zero if the sequence of finite populations is a sequence of random samples from an infinite population in which

$$y_i = \mathbf{x}_i \, \boldsymbol{\beta} + e_i, \tag{8.3}$$

and the e_i the sample are independent of $\mathbf{x}_i$ with $E\{e_i \mid \mathbf{x}_i\} = 0$.

Alternatively, a sufficient condition for $\bar{a}_N$ to be zero is the existence of a column vector $\boldsymbol{\xi}$ such that

$$\mathbf{x}_i \, \boldsymbol{\xi} = p_i^{-1} \tag{8.4}$$

for $i = 1, 2, \ldots, N$. Thus, if the reciprocal of the response probability is a linear function of the control variables, the regression estimator is a consistent estimator of the mean of y. One way in which (8.4) can be satisfied is for the elements of $\mathbf{x}_i$ to be dummy variables that define subgroups and for the response probabilities to be constant in each subgroup.

If (8.4) holds and if the probability of responding is independent from unit to unit, then the estimated variance based on (4.12) is an appropriate estimator for the variance of the regression estimator of the mean. It is particularly important that a variance estimator of the form (4.12) or (4.25), and not of the form (4.26) be used, because $\bar{\mathbf{x}}_N - \bar{\mathbf{x}}_\pi$ is, in general, not $O_p(n^{-1/2})$ in the presence of nonresponse. Singh and Folsom (2000) make a similar argument for the variance estimator (4.25) when using regression to adjust for coverage error.

Often a preliminary adjustment to the selection probabilities is made for nonresponse and this is followed by regression estimation. The most frequently used response adjustment is to form adjustment cells (post strata) and to ratio adjust the weights of respondents in the cell so that the sum of the weights is equal to the estimated (or known) total for the cell. See, for example, Little and Rubin (1987, page 250). Procedures using an estimated response probability function are discussed by Cassel, Särndal and Wretman (1983), Rosenbaum and Rubin (1983), Folsom and Witt (1994). Fuller and An (1998), and Folsom and Singh (2000). Brick, Waksberg and Keeter (1996) use an estimated contact probability to adjust for frame coverage.

To consider procedures based on estimated response probabilities, assume that the inverse of the response probability for individual i is given by

$$p_i^{-1} = g(\mathbf{z}_i; \boldsymbol{\theta}^0), \tag{8.5}$$

where $\mathbf{z}_i$ is a vector of variables that can be observed for both respondents and nonrespondents, $\boldsymbol{\theta}^{\mathbf{0}}$ is the true value of $\boldsymbol{\theta}$, and $g(\mathbf{z}_i; \boldsymbol{\theta})$ is continuous in $\boldsymbol{\theta}$ with continuous first and second derivatives in an open set containing $\boldsymbol{\theta}^{\mathbf{0}}$ for all $\mathbf{z}_i$. The vector $(y_i, \mathbf{x}_i, \mathbf{z}_i)$ is observed, and we assume that p_i is bounded below by a positive number.

Let δ_i be the indicator variable with $\delta_i = 1$ if a response is obtained and $\delta_i = 0$ if a response is not obtained. Using the vector $(\delta_i, \mathbf{z}_i)$, the parameter $\boldsymbol{\theta}^{\mathbf{0}}$ of the response probability function is estimated. Assume that $\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^{\mathbf{0}} = O_p(n^{-1/2})$, where $\hat{\boldsymbol{\theta}}$ is the estimator of $\boldsymbol{\theta}$. Let $\boldsymbol{\beta}_N$ denote the finite population regression vector for the regression of y on $\mathbf{x}$. Let

$$\tilde{\boldsymbol{\beta}} = \left(\sum_{i \in A} \mathbf{x}_i' \mathbf{x}_i \, \pi_i^{-1} \, \hat{p}_i^{-1} \, \delta_i \right)^{-1} \sum_{i \in A} \mathbf{x}_i' \, y_i \, \pi_i^{-1} \, \hat{p}_i^{-1} \, \delta_i, \tag{8.6}$$

where π_i are the selection probabilities and $\hat{p}_i^{-1} = g(\mathbf{z}_i; \hat{\boldsymbol{\theta}})$. Under conditions of the type used in section 4,

$$\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_N = \mathbf{M}_{xx}^{-1} \sum_{i \in A} \delta_i \, \pi_i^{-1} \, p_i^{-1} \, \mathbf{x}_i' \, a_i \, [1 + p_i \, \mathbf{g}_{1,i} (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^0)] + O_p(n^{-1}),$$

where $\mathbf{g}_{1,i}$ is the row vector of first derivatives of $g(\mathbf{z}_i; \boldsymbol{\theta})$ evaluated at $\boldsymbol{\theta} = \boldsymbol{\theta}^{\mathbf{0}}$ and $\mathbf{M}_{xx} = \boldsymbol{\Sigma}_{i \in A} \mathbf{x}_i' \mathbf{x}_i \pi_i^{-1} p_i^{-1} \delta_i$. If $\mathbf{g}_{1,i}$ is uncorrelated with a_i, then the term involving $\mathbf{g}_{1,i} a_i$ is $O_p(n^{-1})$ and the variance estimator constructed as if $g(\mathbf{z}; \boldsymbol{\theta}^{\mathbf{0}})$ is known is appropriate. The conditions are satisfied if $\mathbf{z}_i$ is a subvector of $\mathbf{x}_i$ and $\mathbf{z}_i$ defines imputation cells (adjustment cells) with equal response rates within a cell.

9. Practical Considerations

If the regression weights are to be used in a general purpose survey, no individual weight used in estimating a total should be less than one. Also, it seems reasonable, on robustness grounds, to avoid very large weights. We discuss some procedures that have been developed to accomplish these objectives.

A number of algorithms produce positive weights with a high probability. Raking ratio procedures produces positive weights for most data configurations. Deville, Särndal and Sautory (1993) discuss the extension of raking ratio to general x–variables and extensions to include bounds on the weights.

Tillé (1998) suggested the use of approximate conditional probabilities, conditional on $\bar{\mathbf{x}}_\pi$, to compute an estimator. His approximation can be extended to produce regression weights that are positive with high probability. Let $\bar{\mathbf{x}}_\pi^{(i)}$ be an estimator obtained by deleting element *i*, or primary sampling unit *i*, and modifying the remaining weights so that $\bar{\mathbf{x}}_\pi^{(i)}$ is unbiased, or consistent to the same order as $\bar{\mathbf{x}}_\pi$, for the population mean of all elements excluding *i*. The estimator $\bar{\mathbf{x}}_\pi^{(i)}$ can be the estimator used to construct jackknife deviates. Let $\hat{\boldsymbol{\Sigma}}_{\bar{x}\bar{x}}$ be an estimator of the covariance matrix of $\bar{\mathbf{x}}_\pi$ and let $\hat{\boldsymbol{\Sigma}}_{\bar{x}\bar{x}(i)}$ be an estimator of the conditional covariance matrix of $\bar{\mathbf{x}}_\pi^{(i)}$ conditional on $i \in A$. Then, in large samples $\bar{\mathbf{x}}_\pi$ and $\bar{\mathbf{x}}_\pi^{(i)}$ are approximately normally distributed and an estimator of the probability that *i* is in the sample given the estimated mean $\bar{\mathbf{x}}_\pi$, is

$$\hat{\pi}_{i|A} = \hat{P}\{ i \in A | \mathbf{F}_N, \bar{\mathbf{x}}_\pi \}$$
$$= \pi_i |\hat{\boldsymbol{\Sigma}}_{\bar{x}\bar{x}}|^{1/2} |\hat{\boldsymbol{\Sigma}}_{\bar{x}\bar{x}(i)}|^{-1/2} \exp\{0.5(\mathbf{G}_{\bar{x}\bar{x}} - \mathbf{G}_{\bar{x}\bar{x}(i)})\} \quad (9.1)$$

where

$$\mathbf{G}_{\bar{x}\bar{x}} = (\bar{\mathbf{x}}_\pi - \bar{\mathbf{x}}_N) \hat{\boldsymbol{\Sigma}}_{\bar{x}\bar{x}}^{-1} (\bar{\mathbf{x}}_\pi - \bar{\mathbf{x}}_N)',$$

$$\mathbf{G}_{\bar{x}\bar{x}(i)} = (\bar{\mathbf{x}}_\pi^{(i)} - \bar{\mathbf{x}}_N^{(i)}) \hat{\boldsymbol{\Sigma}}_{\bar{x}\bar{x}(i)}^{-1} (\bar{\mathbf{x}}_\pi^{(i)} - \bar{\mathbf{x}}_N^{(i)})',$$

and $\bar{\mathbf{x}}_N^{(i)} = (N-1)^{-1}(N\bar{\mathbf{x}}_N - \mathbf{x}_i)$. For simple random sampling, Tillé (1998) showed that the estimator

$$\bar{y}_{p\pi} = N^{-1} \sum_{i \in A} \pi_{i|A}^{-1} y_i, \quad (9.2)$$

where $\pi_{i|A}$ is the conditional probability calculated under the normality assumptions, is approximately equal to the regression estimator. Because the estimator is not calibrated, we suggest a calibrated version obtained by computing the regression estimator with $\hat{\pi}_{i|A}$ as initial weights. The difference between (9.2) and the regression estimator constructed with initial weights $\hat{\pi}_{i|A}$ is $O_p(n^{-1})$. Hence, there is a good chance that the regression weights so constructed will be positive. The variance estimator $\hat{\boldsymbol{\Sigma}}_{\bar{x}\bar{x}(i)}$ is relatively simple to compute for stratified samples but may require considerable computation for other cases. Thus one may choose to approximate $\boldsymbol{\Sigma}_{\bar{x}\bar{x}(i)}$.

Given that the regression weights are being constructed by minimizing an objective function, one can add restrictions to the problem to place bounds on the weights. Huang and Fuller (1978) gave an interative procedure equivalent to constructing a $\boldsymbol{\Phi}$ matrix at each step that reduces the weight on observations whose current weight deviates from the average by a large absolute amount.

To discuss additional procedures associated with quadratic objective functions, assume we have a working covariance matrix, denoted by $\boldsymbol{\Phi}_{ee}$, for the model (5.1) that is to be used to construct a regression estimator. Let $\boldsymbol{\alpha}$ be the column vector of initial weights and assume $\boldsymbol{\Phi}_{ee} \boldsymbol{\alpha}$ is in the column space of $\mathbf{X}$. Then the weights that minimize the conditional model variance are the weights that minimize $\mathbf{w}' \boldsymbol{\Phi}_{ee} \mathbf{w}$ or, equivalently, that minimize

$$(\mathbf{w} - \boldsymbol{\alpha})' \boldsymbol{\Phi}_{ee} (\mathbf{w} - \boldsymbol{\alpha}) \quad (9.3)$$

subject to the constraint

$$\mathbf{w}' \mathbf{X} = \bar{\mathbf{x}}_N. \quad (9.4)$$

Given an objective function, we can add restrictions on the w_i such as

$$L_1 \le w_i \le L_2, \quad i \in A, \quad (9.5)$$

where L_1 and L_2 are nonnegative constants. Minimizing (9.3), subject to the constraints (9.4) and (9.5) is a quadractic programming problem. The use of quadratic programming was suggested by Husain (1969) and was used by Isaki, Tsay and Fuller (2000).

If a large number of control variables are used, it may not be possible to construct weights satisfying the calibration constraints and also falling within reasonable bounds. The practitioner is faced with making compromises. The most common practice is to drop variables from the model. See Bankier, Rathwell and Maijkowski (1992) and Silva and Skinner (1997). To discuss an alternative procedure, consider the situation in which some of the constraints are required but others can be relaxed. Let the matrix of observations on the auxiliary variables be partitioned as $(\mathbf{X}_0, \mathbf{X}_2)$, where $\mathbf{X}_0$ is the set of variables for which exact constraints are required and $\mathbf{X}_2$ is the set for which the constraints can be relaxed. Assume $\boldsymbol{\Phi}_{ee} \boldsymbol{\alpha}$ is in the column space of $\mathbf{X}_0$. Then a generalization of (9.3) and (9.4) is the function

$$(\mathbf{w} - \boldsymbol{\alpha})' \boldsymbol{\Phi}_{ee} (\mathbf{w} - \boldsymbol{\alpha})$$
$$+ (\mathbf{w}' \mathbf{X}_2 - \bar{\mathbf{x}}_{2,N}) \boldsymbol{\Psi} (\mathbf{w}' \mathbf{X}_2 - \bar{\mathbf{x}}_{2,N})' \quad (9.6)$$

and the constraint

$$\mathbf{w}'\,\mathbf{X}_0 - \bar{\mathbf{x}}_{0,N} = \mathbf{0}, \tag{9.7}$$

where $\mathbf{\Phi}_{ee}$ and $\mathbf{\Psi}$ are positive definite symmetric matrices and $\bar{\mathbf{x}}_N = (\bar{\mathbf{x}}_{0,N},\ \bar{\mathbf{x}}_{2,N})$. The $\mathbf{w}$ that minimizes (9.6) subject to (9.7) minimizes the mean squared error of the unbiased linear predictor of $\bar{\mathbf{x}}_N\,\boldsymbol{\beta}$ under the mixed model

$$\mathbf{y} = \mathbf{X}_0\,\boldsymbol{\beta}_0 + \mathbf{X}_2\,\boldsymbol{\beta}_2 + \mathbf{e},$$

where $\boldsymbol{\beta}_2 \sim (\mathbf{0}, \mathbf{\Psi})$, $e \sim (\mathbf{0}, \mathbf{\Phi}_{ee})$, the random vector $\boldsymbol{\beta}_2$ is independent of $\mathbf{e}$, and $\boldsymbol{\beta}_0$ is a fixed vector. See Lazzeroni and Little (1998) for the use of random models for post stratification.

The vector $\mathbf{w}'$ that minimizes (9.6) subject to restriction (9.7) is

$$\mathbf{w}' = \boldsymbol{\alpha}' + (\bar{\mathbf{x}}_N - \bar{\mathbf{x}}_\pi)\,\mathbf{H}_{x\psi x}^{-1}\mathbf{X}'\,\mathbf{\Phi}_{ee}^{-1}, \tag{9.8}$$

where

$$\mathbf{H}_{x\psi x} = \begin{pmatrix} \mathbf{X}_0'\,\mathbf{\Phi}_{ee}^{-1}\,\mathbf{X}_0 & \mathbf{X}_0'\,\mathbf{\Phi}_{ee}^{-1}\,\mathbf{X}_2 \\ \mathbf{X}_2'\,\mathbf{\Phi}_{ee}^{-1}\,\mathbf{X}_0 & \mathbf{\Psi}^{-1} + \mathbf{X}_2'\,\mathbf{\Phi}_{ee}^{-1}\,\mathbf{X}_2 \end{pmatrix}. \tag{9.9}$$

The estimator can be written

$$\bar{y}_{r\,\mathrm{reg}} = \mathbf{w}'\,\mathbf{y} = \bar{y}_\pi + (\,\bar{\mathbf{x}}_N - \bar{\mathbf{x}}_\pi\,)\,\hat{\boldsymbol{\theta}}, \tag{9.10}$$

where $\hat{\boldsymbol{\theta}} = \mathbf{H}_{x\psi y}^{-1}\,\mathbf{X}'\,\mathbf{\Phi}_{ee}^{-1}\,\mathbf{y}$. See Henderson (1963), Robinson (1991), and Rao (2002, Chapter 6).

Husain (1969) considered (9.6) for a simple random sample from a normal distribution with $\mathbf{X}_0 = \mathbf{J}$, $\mathbf{\Phi}_{ee} = \mathbf{I}$, and $\mathbf{\Psi}^{-1} = \gamma^{-1}\,\hat{\mathbf{\Sigma}}_{x,22}$, where $\hat{\mathbf{\Sigma}}_{x,22}$ is the estimated covariance matrix of $\bar{\mathbf{x}}_{2,\pi}$, and γ is a constant to be determined. For this case, Husain showed that the optimal γ is

$$\gamma_{\mathrm{opt}} = [k_2(1-R^2)]^{-1}(n-k_2-2)\,R^2, \tag{9.11}$$

where k_2 is the dimension of $\mathbf{x}_2$ and R^2 is the squared multiple correlation coefficient. Bardsley and Chambers (1984) considered the function (9.6), the division of $\mathbf{x}_i$ into two components, and studied the behavior of the estimator from a model perspective. The procedure associated with (9.5), (9.6) and (9.7) was used by Isaki, Tsay and Fuller (2000). In that application, the vector $\bar{\mathbf{x}}_{0,N}$ contained marginal totals of a multiway table and $\bar{\mathbf{x}}_{2,N}$ contained totals for interior cells. Rao and Singh (1997) studied a closely related estimator in which tolerances are given for the difference between the final estimates for elements of $\bar{\mathbf{x}}_{2,N}$ and the corresponding elements of $\bar{\mathbf{x}}_{2,N}$.

Park (2002) extended Husain's optimality results to a more general $\mathbf{\Psi}$. The $\mathbf{x}_2$ vector can be transformed so that $\hat{V}\{\bar{\mathbf{x}}_{2,\pi}\}$ for the transformed vector is a diagonal matrix and so that $\tilde{\mathbf{X}}_2'\,\mathbf{\Phi}_{ee}^{-1}\,\tilde{\mathbf{X}}_2$ is a diagonal matrix, where $\tilde{\mathbf{X}}_2$ is the part of $\mathbf{X}_2$ that is orthogonal to $\mathbf{X}_0$ in the metric $\mathbf{\Phi}_{ee}$. That is,

$$\tilde{\mathbf{X}}_2 = \mathbf{X}_2 - \mathbf{X}_0\,(\mathbf{X}_0'\,\mathbf{\Phi}_{ee}^{-1}\,\mathbf{X}_0)^{-1}\mathbf{X}_0'\,\mathbf{\Phi}_{ee}^{-1}\,\mathbf{X}_2.$$

Then the diagonal $\mathbf{\Psi}$ that minimizes the approximate variance has elements

$$\psi_{ii} = (m_{ii}V_{\beta\beta ii})^{-1}\beta_i^2, \tag{9.12}$$

where m_{ii} is the i^{th} element of the diagonal matrix $\tilde{\mathbf{X}}_2'\,\mathbf{\Phi}_{ee}^{-1}\,\tilde{\mathbf{X}}_2$ and $V_{\beta\beta ii}$ is the variance of $\hat{\beta}_i$ in the transformed scale. To implement the procedure one must estimate the population parameters or choose realistic values for a general purpose $\mathbf{\Psi}$. If one postulates a superpopulation random model for $\boldsymbol{\beta}$, then the β_i^2 of (9.12) is replaced with $E\{\beta_i^2\}$, where the expectation is the model expectation.

10. Comments

Regression estimation is a flexible and powerful tool for the incorporation of auxiliary information into the estimation process. Closely related procedures, such as raking ratio, have large sample properties equivalent to those of regression estimators. The linearity of such estimators is of paramount importance because it permits the construction of a general purpose data set that provides very good estimators for a wide range of parameters.

Given a concentrated interest in a single y–variable, efficiency gains may be possible by postulating a particular set of auxiliary variables and a particular error covariance matrix. Because of the simple nature of the design consistency requirement, it is easy to test such models for design consistency.

Acknowledgements

This research was partially supported by Cooperative Agreement 43-3AEU-0-80064 between Iowa State University, the U. S. National Agricultural Statistics Service and the U. S. Bureau of the Census. I am deeply indebted to Mingue Park for assistance in literature review, for comments on and repair of theoretical results, and for use of material from his thesis. I thank Michael Hidiroglou, J.N.K. Rao, Harold Mantel, and Jean Opsomer for useful comments on drafts of the manuscript. I thank the Associate Editor for numerous comments that improved the presentation.

Appendix

This appendix contains theorems supporting the limiting properties of the regression estimators discussed in section 4.

Theorem A.1. Let $\{U_N, \mathrm{F}_N, A_N,\ n_N\!: N = k+3, k+4, \ldots\}$ be a sequence of finite populations and samples, where $\mathbf{F}_N$ is a sample from an infinite population with eighth moments, A_N is the sample of size n_N selected from the N^{th} population. Let $\hat{\boldsymbol{\beta}}$ be defined by (4.4) of the text, and let

$$\hat{\mathbf{Q}}_{zz} = n^{-1}\,\mathbf{Z}'\,\mathbf{\Phi}^{-1}\,\mathbf{Z},$$

where $\mathbf{\Phi}$ is a positive definite symmetric $n \times n$ matrix that may be a function of **X** but not of y, **Z** is defined following (4.2), and we omit the subscript N on sample quantities. Assume $\hat{\mathbf{Q}}_{zz}$ is positive definite with probability one. If $\mathbf{\Phi}$ is random, assume the rows of $\mathbf{\Phi}^{-1}\mathbf{Z}$ have bounded fourth moments. Assume the selection probabilities satisfy

$$0 < K_1 < Nn^{-1}\,\pi_i < K_2,$$

where π_i are the selection probabilities. Assume the sample design is such that for any **z** with bounded fourth moments

$$[(\overline{\mathbf{z}}_{\mathrm{HT}} - \overline{\mathbf{z}}_N)',\ (\hat{\mathbf{Q}}_{zz} - \mathbf{Q}_{zzN})] \mid \mathbf{F}_N = O_p(n^{-1/2}), \qquad \text{(A.1)}$$

where

$$\overline{\mathbf{z}}_{\mathrm{HT}} = (\overline{y}_{\mathrm{HT}},\ \overline{\mathbf{x}}_{\mathrm{HT}}) = N^{-1}\sum_{i\in A}\pi_i^{-1}\,\mathbf{z}_i, \qquad \text{(A.2)}$$

$\mathbf{Q}_{zzN} = E\{\hat{\mathbf{Q}}_{zz} \mid \mathbf{F}_N\}$, $\overline{\mathbf{z}}_N$ is the finite population mean of $\mathbf{z}$, $\mathbf{Q}_{zzN}$ is a positive definite matrix for the N^{th} population, and the limit of $\mathbf{Q}_{zzN}$ is positive definite. Then

$$\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_N \mid \mathbf{F}_N = \mathbf{Q}_{xxN}^{-1}\,\overline{\mathbf{b}}'_{\mathrm{HT}} + O_p(n^{-1}), \qquad \text{(A.3)}$$

where $\boldsymbol{\beta}_N = \mathbf{Q}_{xxN}^{-1}\,\mathbf{Q}_{xyN}$, $\overline{\mathbf{b}}_{\mathrm{HT}} = N^{-1}\,\Sigma_{i\in A}\,\pi_i^{-1}\,\mathbf{b}_i$, $\mathbf{b}'_i = n^{-1}\,N\pi_i\,\zeta'_i\,e_i$,

$$\mathbf{Q}_{zzN} = \begin{pmatrix} Q_{yyN} & \mathbf{Q}_{yxN} \\ \mathbf{Q}_{xyN} & \mathbf{Q}_{xxN} \end{pmatrix}, \qquad \text{(A.4)}$$

$e_i = y_i - \mathbf{x}_i\,\boldsymbol{\beta}_N$, and ζ'_i is column i of $\mathbf{X}'\,\mathbf{\Phi}^{-1}$.
Assume the design is such that

$$\mathbf{V}_{\overline{z}\overline{z}}^{-1/2}\,\{\overline{\mathbf{z}}_{\mathrm{HT}} - \overline{\mathbf{z}}_N \mid \mathbf{F}_N\} \xrightarrow{\ \mathrm{L}\ } N(\mathbf{0},\ \mathbf{I}), \qquad \text{(A.5)}$$

as $n_N \to \infty$ for any **z** with finite fourth moments, where $\mathbf{V}_{\overline{z}\overline{z}}$ is the covariance matrix of $\overline{\mathbf{z}}_{\mathrm{HT}} - \overline{\mathbf{z}}_N$. Assume that $\mathbf{V}_{\overline{z}\overline{z}}$ is $O(n^{-1})$ and that the design admits an estimator $\hat{\mathbf{V}}_{\overline{z}\overline{z}}$ such that

$$n(\hat{\mathbf{V}}_{\overline{z}\overline{z}} - \mathbf{V}_{\overline{z}\overline{z}}) \mid \mathbf{F}_N = o_p(1) \qquad \text{(A.6)}$$

for any **z** with bounded fourth moments. Then

$$[\hat{V}\{\hat{\boldsymbol{\beta}}\}]^{-1/2}[\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_N] \mid \mathbf{F}_N \xrightarrow{\ \mathrm{L}\ } N(\mathbf{0},\ \mathbf{I}), \qquad \text{(A.7)}$$

where

$$\hat{V}\{\hat{\boldsymbol{\beta}}\} = \hat{\mathbf{Q}}_{xx}^{-1}\,\hat{\mathbf{V}}_{\overline{b}\overline{b}}\,\hat{\mathbf{Q}}_{xx}^{-1}, \qquad \text{(A.8)}$$

$\hat{\mathbf{V}}_{\overline{b}\overline{b}} = \hat{V}\{\overline{\mathbf{b}}'_{\mathrm{HT}}\}$ is the estimated design variance of $\overline{\mathbf{b}}'_{\mathrm{HT}}$ calculated with $\hat{\mathbf{b}}'_i = n^{-1}N\pi_i\,\zeta'_i\,\hat{e}_i$ and $\hat{e}_i = y_i - \mathbf{x}_i\,\hat{\boldsymbol{\beta}}$.

Proof. The error in $\hat{\boldsymbol{\beta}}$ is

$$\begin{aligned}\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_N &= (\mathbf{X}'\,\mathbf{\Phi}^{-1}\,\mathbf{X})^{-1}\,[\mathbf{X}'\,\mathbf{\Phi}^{-1}\,\mathbf{y} - \mathbf{X}'\,\mathbf{\Phi}^{-1}\,\mathbf{X}\boldsymbol{\beta}_N] \\ &= \hat{\mathbf{Q}}_{xx}^{-1}\,(n^{-1}\,\mathbf{X}'\,\mathbf{\Phi}^{-1}\,\mathbf{e}).\end{aligned}$$

Now $\hat{\boldsymbol{\beta}}$ is a generalized least squares estimator. Therefore

$$\hat{\mathbf{e}}'\,\mathbf{\Phi}^{-1}\,\mathbf{X} = (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})'\,\mathbf{\Phi}^{-1}\,\mathbf{X} = \mathbf{0}$$

and $\mathbf{Q}_{xyN} - \boldsymbol{\beta}'_N\,\mathbf{Q}_{xxN} = \mathbf{Q}_{exN} = \mathbf{0}$. By assumption (A.1)

$$\hat{\mathbf{Q}}'_{ex} = n^{-1}\,\mathbf{X}'\,\mathbf{\Phi}^{-1}\,\mathbf{e} = O_p\,(n^{-1/2}).$$

Thus

$$\begin{aligned}\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_N &= \mathbf{Q}_{xxN}^{-1}\left(n^{-1}\sum_{i\in A}\zeta'_i\,e_i\right) + O_p\,(n^{-1}) \\ &= \mathbf{Q}_{xxN}^{-1}\left(N^{-1}\sum_{i\in A}\pi'_i\,\mathbf{b}'_i\right) + O_p\,(n^{-1}).\end{aligned}$$

The $\mathbf{b}_i$ have bounded fourth moments by the assumptions. Thus, by assumption (A.5)

$$\mathbf{V}_{\beta\beta}^{1/2}\,(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_N) \xrightarrow{\ \mathrm{L}\ } N(\mathbf{0}, \mathbf{I}),$$

where

$$\mathbf{V}_{\beta\beta} = \mathbf{Q}_{xxN}^{-1}\,\mathbf{V}_{\overline{b}\overline{b}}\,\mathbf{Q}_{xxN}^{-1}$$

and $\mathbf{V}_{\overline{b}\overline{b}} = V\{\overline{\mathbf{b}}_{\mathrm{HT}}\}$. Now

$$\begin{aligned} n^{-1}\,\mathbf{X}'\,\mathbf{\Phi}^{-1}\,\hat{\mathbf{e}} &= n^{-1}\,\mathbf{X}'\,\mathbf{\Phi}^{-1}\,\mathbf{e} + n^{-1}\,\mathbf{X}'\,\mathbf{\Phi}^{-1}\,\mathbf{X}\,(\boldsymbol{\beta}_N - \hat{\boldsymbol{\beta}}) \\ &=: N^{-1}\sum_{i\in A}\pi_i^{-1}\,\mathbf{b}'_i + N^{-1}\sum_{i\in A}\pi_i^{-1}\,\mathbf{h}'_i,\end{aligned}$$

where

$$\mathbf{h}'_i = n^{-1}N\pi_i\,\zeta'_i\,\mathbf{x}_i\,\boldsymbol{\delta}_\beta$$

and $\boldsymbol{\delta}_\beta = \boldsymbol{\beta}_N - \hat{\boldsymbol{\beta}}$. For any fixed $\boldsymbol{\delta}$, by (A.6), the estimated variance of $N^{-1}\sum_{i\in A}\pi_i^{-1}(\mathbf{b}'_i + \mathbf{h}'_i)$ is consistent for the variance of the estimator of the mean of $\mathbf{b} + \mathbf{h}$. By assumption, the elements of $\zeta'_i\,\mathbf{x}_i$ have fourth moments. For a fixed $\boldsymbol{\delta}$ the variance of $\overline{\mathbf{h}}_{\mathrm{HT}}$ is $O(n^{-1})$. For $\boldsymbol{\delta} = \boldsymbol{\delta}_\beta$,

$$\hat{V}\{\overline{\mathbf{h}}'_{\mathrm{HT}}\} = o_p\,(n^{-1}),$$

and

$$\hat{V}\{\overline{\mathbf{b}}'_{\mathrm{HT}}\} = V\{\overline{\mathbf{b}}'_{\mathrm{HT}}\} + o_p\,(n^{-1})$$

because $\boldsymbol{\delta}_\beta = O_p(n^{-1/2})$. Result (A.7) then follows from the asymptotic normality of $\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_N$.

Theorem A.2. Let $\mathbf{y}' = (y_1,\ y_2,\ \ldots,\ y_n)$ and $\mathbf{X}' = (\mathbf{x}'_1,\ \mathbf{x}'_1,\ \ldots,\ \mathbf{x}'_n)$. Let $\mathbf{\Phi}$ be a nonsingular symmetric $n \times n$ matrix and let $\mathbf{\Phi}_N$ be a nonsingular symmetric $N \times N$ matrix. Let

$$\overline{y}_\pi,\ \overline{\mathbf{x}}_\pi,\ n^{-1}(\mathbf{X}'\,\mathbf{\Phi}^{-1}\,\mathbf{X}) \quad \text{and} \quad n^{-1}\,\mathbf{X}'\,\mathbf{\Phi}^{-1}\,\mathbf{y}$$

be design consistent estimators for finite population characteristics $\overline{y}_N$, $\overline{\mathbf{x}}_N$, $\mathbf{Q}_{xxN}$ and $\mathbf{Q}_{xyN}$, respectively, where

$$[\mathbf{Q}_{xxN},\ \mathbf{Q}_{xyN}] = [N^{-1}\mathbf{X}'_N \mathbf{\Phi}_N^{-1} \mathbf{X}_N,\ N^{-1}\mathbf{X}_N \mathbf{\Phi}_N^{-1} \mathbf{y}_N]. \quad \text{(A.9)}$$

Let $\boldsymbol{\beta}_N = \mathbf{Q}_{xxN}^{-1}\mathbf{Q}_{xyN}$. Let there be a sequence of column vectors $\{\boldsymbol{\gamma}_N\}$ such that

$$\mathbf{X}\boldsymbol{\gamma}_N = \mathbf{\Phi}\mathbf{D}_\pi^{-1}\mathbf{J}_n \quad \text{(A.10)}$$

for all possible samples, where $\mathbf{D}_\pi = \text{diag}(\pi_1,\ \pi_2,\ ...,\ \pi_n)$ and $\mathbf{J}_n$ is an n–dimensional column vector of ones. Then, the regression estimator $\overline{\mathbf{x}}_N\ \hat{\boldsymbol{\beta}}$ with

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{\Phi}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{\Phi}^{-1}\mathbf{y}, \quad \text{(A.11)}$$

is a design consistent estimator of $\overline{y}_N$.

Proof. If $\hat{\boldsymbol{\beta}}$ is defined by (A.11), then by the properties of generalized least squares estimators,

$$(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})'\mathbf{\Phi}^{-1}\mathbf{X} = \mathbf{0}.$$

If (A.10) holds, then

$$(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})'\mathbf{D}_\pi^{-1}\mathbf{J} = \left(\sum_{i \in A}\pi_i^{-1}\right)(\overline{y}_\pi - \overline{\mathbf{x}}_\pi\hat{\boldsymbol{\beta}}) = 0.$$

It follows that $\overline{y}_{\text{reg}}$ is design consistent because

$$\begin{aligned} 0 &= p\lim_{N\to\infty}\{(\overline{y}_\pi - \overline{\mathbf{x}}_\pi\hat{\boldsymbol{\beta}}_n) \mid \mathbf{F}_N\} \\ &= p\lim_{N\to\infty}\{(\overline{y}_\pi - \overline{\mathbf{x}}_\pi\boldsymbol{\beta}_N) \mid \mathbf{F}_N\} \\ &= p\lim_{N\to\infty}\{(\overline{y}_N - \overline{\mathbf{x}}_N\boldsymbol{\beta}_N) \mid \mathbf{F}_N\}. \end{aligned}$$

Theorem A.3. Let a sequence of populations and samples be as defined in Theorem A.1. Let $\mathbf{z}_i$ be a vector of the form $\mathbf{z}_i = (y_i,\ 1,\ \mathbf{x}_{1,i})$ and let $\mathbf{z}_{1,i} = (y_i,\ \mathbf{x}_{1,i})$. Assume $\overline{\mathbf{z}}_{1,\pi}$ is a design consistent estimator of the population mean $\overline{\mathbf{z}}_{1,N}$ with nonsingular covariance matrix

$$V\{\overline{\mathbf{z}}_{1,\pi} \mid \mathbf{F}_N\} = O(n^{-1}) \quad \text{(A.12)}$$

and

$$n^{1/2}(\overline{\mathbf{z}}_{1,\pi} - \overline{\mathbf{z}}_{1,N}) \mid \mathbf{F}_N \xrightarrow{\text{L}} N(\mathbf{0},\ \mathbf{\Sigma}_{\overline{z}\,\overline{z}}), \quad \text{(A.13)}$$

where $\mathbf{\Sigma}_{\overline{z}\,\overline{z}}$ is the limit of $nV\{\overline{\mathbf{z}}_{1,\pi} \mid \mathbf{F}_N\}$. Assume there is an estimator of the variance of $\overline{\mathbf{z}}_{1,\pi}$, denoted by $\hat{V}\{\overline{\mathbf{z}}_{1,\pi}\}$, such that

$$p\lim_{N\to\infty} n^{1+\delta}(\hat{V}\{\overline{\mathbf{z}}_{1,\pi}\} - V\{\overline{\mathbf{z}}_{1,\pi} \mid \mathbf{F}_N\}) = \mathbf{0} \quad \text{(A.14)}$$

for some $\delta > 0$. Let $\hat{\boldsymbol{\beta}}_{1,\text{dopt}}$ be the vector that minimizes

$$\hat{V}\{\overline{y}_\pi - \overline{\mathbf{x}}_{1,\pi}\boldsymbol{\beta}_{1,d}\} \quad \text{(A.15)}$$

and let $\boldsymbol{\beta}_{1,\text{dopt}}$ be the vector that minimizes $V\{\overline{y}_\pi - \overline{\mathbf{x}}_{1,\pi}\boldsymbol{\beta}_{1,d}\}$. Let $\overline{y}_{d,\text{reg}}$ be defined by (4.29). Then $\overline{y}_{d,\text{reg}}$ has the minimum limit variance for design consistent estimators of the form $\overline{y}_\pi + (\overline{\mathbf{x}}_{1,N} - \overline{\mathbf{x}}_{1,\pi})\boldsymbol{\beta}_{1,d}$. Also

$$[\hat{V}\{\overline{e}_\pi\}]^{-1/2}(\overline{y}_{d,\text{reg}} - \overline{y}_N) \xrightarrow{\text{L}} N(0,\ 1), \quad \text{(A.16)}$$

where $\hat{V}\{\overline{e}_\pi\}$ is the estimator of (A.14) constructed with $\hat{e}_i = y_i - \overline{y}_\pi - (\mathbf{x}_{1,i} - \overline{\mathbf{x}}_{1,\pi})\hat{\boldsymbol{\beta}}_{1,\text{dopt}}$.

Proof. The estimator

$$\hat{\boldsymbol{\beta}}_{1,\text{dopt}} = [\hat{V}\{\overline{\mathbf{x}}_{1,\pi}\}]^{-1}\hat{C}\{\overline{\mathbf{x}}_{1,\pi},\ \overline{y}_\pi\}$$

minimizes the estimated variance of (A.15), and, by assumption (A.14), the estimated variance is consistent for the true variance. Hence, $\hat{\boldsymbol{\beta}}_{1,\text{dopt}}$ is design consistent for $\boldsymbol{\beta}_{1,\text{dopt}}$ and $\boldsymbol{\beta}_{1,\text{dopt}}$ minimizes $V\{\overline{y}_\pi - \overline{\mathbf{x}}_{1,\pi}\boldsymbol{\beta}\}$. Therefore, no estimator of the form (4.29) has a limit distribution with smaller variance.

Now

$$\begin{aligned} \overline{y}_{d,\text{reg}} - \overline{y}_N &= \overline{y}_\pi - \overline{y}_N - (\overline{\mathbf{x}}_{1,N} - \overline{\mathbf{x}}_{1,\pi})\hat{\boldsymbol{\beta}}_{1,\text{dopt}} \\ &= \overline{e}_\pi + o_p(n^{-1/2}), \end{aligned}$$

where $e_i = y_i - \overline{y}_N - (\mathbf{x}_{1,i} - \overline{\mathbf{x}}_{1,N})\boldsymbol{\beta}_{1,\text{dopt}}$. Therefore the variance of the limiting distribution of $n^{1/2}(\overline{y}_{d,\text{reg}} - \overline{y}_N)$ is the variance of $n^{1/2}(\overline{e}_\pi - \overline{e}_N)$. By assumption (A.14), the estimator $\hat{V}\{\overline{\mathbf{z}}_\pi\,\boldsymbol{\gamma}\}$ is a consistent variance estimator of $V\{\overline{\mathbf{z}}_\pi\,\boldsymbol{\gamma}\}$ for any fixed $\boldsymbol{\gamma}$. Because $\hat{\boldsymbol{\beta}}_{1,\text{dopt}} - \boldsymbol{\beta}_{1,\text{dopt}} = o_p(1)$, the estimated variance based on $\hat{e}_i$ converges to the estimated variance based on e_i and (A.16) holds.

References

Anderson, C., and Nordberg, L. (1998). A user's guide to CLAN97. Statistics Sweden, Orebro, Sweden.

Bankier, M.D., Rathwell, S. and Majkowski, M. (1992). Two step generalized least squares estimation in the 1991 Canadian Census. Working Paper-Methodology Branch, Census Operations Section, Social Survey Methods Division. Statistics Canada, Ottawa.

Bankier, M.D., Houle, A.M. and Luc, M. (1997). Calibration estimation in the 1991 and 1996 Canadian census. Statistics Canada (draft), 8 pages.

Bardsley, P., and Chambers, R.L. (1984). Multipurpose estimation from unbalanced samples. *Applied Statistics*, 33, 290-299.

Bethlehem, J.G. (1988). Reduction of nonresponse bias through regression estimation. *Journal of Official Statistics*, 4, 251-260.

Bethlehem, J.G., and Keller, W.J. (1987). Linear weighting of sample survey data. *Journal of Official Statistics*, 3, 141-153.

Bishop, Y.M.M., Fienberg, S.E. and Holland, P.W. (1975). *Discrete Multivariate Analysis*: *Theory and Practice.* MIT Press, Cambridge, MA.

Breidt, F.J., and Opsomer, J.D. (2000). Local polynomial regression estimators in survey sampling. *The Annals of Statistics*, 28, 1026-1053.

Brewer, K.R.W. (1963). Ratio estimation and finite populations: Some results deducible from the assumption of an underlying stochastic process. *Australian Journal of Statistics*, 5, 93-105.

Brewer, K.R.W. (1979). A class of robust sampling designs for large-scale surveys. *Journal of the American Statistical Association*, 74, 911-915.

Brewer, K.R.W., Hanif, M. and Tam, S.M. (1988). How nearly can model-based prediction and design-based estimation be reconciled? *Journal of the American Statistical Association*, 83, 128-132.

Brick, J.M., Waksberg, J. and Keeter, S. (1996). Using data on interruptions in telephone service as coverage adjustments. *Survey Methodology*, 22, 185-197.

Cassel, C.M., Särndal, C.-E. and Wretman, J.H. (1976). Some results on generalized difference estimation and generalized regression estimation for finite populations. *Biometrika*, 63, 615-620.

Cassel, C.M., Särndal, C.-E. and Wretman, J.H. (1979). Prediction theory for finite populations when model-based and design-based principles are combined. *Scandinavian Journal of Statistics*, 6, 97-106.

Cassel, C.M., Särndal, C.-E. and Wretman, J.H. (1983). Some uses of statistical models in connection with the nonresponse problem. In *Incomplete Data in Sample Surveys*, (Eds. W.G. Madow, I. Olkin, and D. Rubin). New York: Academic Press, 3, 143-160.

Chambers, R.L. (1996). Robust case-weighting for multipurpose establishment surveys. *Journal of Official Statistics*, 12, 3-32.

Chambers, R.L., Dorfman, A.H. and Wehrly, T.E. (1993). Bias robust estimation in finite populations using nonparametric calibration. *Journal of the American Statistical Association*, 88, 268-277.

Cochran, W.G. (1939). The use of the analysis of variance in enumeration by sampling. *Journal of the American Statistical Association*, 34, 492-510.

Cochran, W.G. (1942). Sampling theory when the sampling units are of unequal sizes. *Journal of the American Statistical Association*, 37, 199-212.

Cochran, W.G. (1946). Relative accuracy of systematic and stratified random samples for a certain class of populations. *Annals of Mathematical Statistics*, 17, 164-177.

Cochran, W.G. (1977). *Sampling Techniques*, 3rd ed., New York: John Wiley & Sons, Inc.

Cook, R.D., and Weisberg, S. (1982). *Residuals and Influence in Regression*. New York: Chapman and Hall.

Deming, W.E., and Stephan, F.F. (1940). On a least squares adjustment of a sampled frequency table when the expected marginal totals are known. *Annals of Mathematical Statistics*, 11, 427-444.

Deming, W.E., and Stephan, F.F. (1941). On the interpretation of censuses as samples. *Journal of the American Statistical Association*, 36, 45-49.

Deville, J., and Särndal, C.-E. (1992). Calibration estimators in survey sampling. *Journal of the American Statistical Association*, 87, 376-382.

Deville, J., Särndal, C.-E. and Sautory, O. (1993). Generalized raking procedures in survey sampling. *Journal of the American Statistical Association*, 88, 1013-1020.

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 function using nonparametric regression. *Annals of Statistics*, 21, 1452-1475.

Duchesne, P. (2000). A note on jackknife variance estimation for the general regression estimator. *Journal of Official Statistics*, 16, 133-138.

Du Mouchel, W.H., and Duncan, G.J. (1983). Using survey weights in multiple regression analysis of stratified samples. *Journal of the American Statistical Association*, 78, 535-543.

Eltinge, J.L., and Yansaneh, I.S. (1997). Diagnostics for formation of nonresponse adjustment cells, with an application to income nonresponse in the U. S. Consumer Expenditure Survey. *Survey Methodology*, 23, 33-40.

Ericson, W.A. (1969). Subjective Bayesian models in sampling finite populations. *Journal of the Royal Statistical Society*, Series B, 31, 195-224.

Estevao, V., Hidiroglou, M.A. and Särndal, C.-E. (1995). Methodological principles for a generalized estimation system at Statistics Canada. *Journal of Official Statistics*, 11, 181-204.

Firth, D., and Bennett, K.E. (1998). Robust models in probability sampling. *Journal of the Royal Statistical Society*, Series B, 60, 3-21.

Folsom, R.E., and Witt, M.B. (1994). Testing a new attrition nonresponse adjustment method for SIPP. *Proceedings of the Section on Survey Research Methods*, American Statistical Association, 428-433.

Folsom, R.E., and Singh, A.C. (2000). The generalized exponential model for a unified approach to sampling weight calibration for outlier weight treatment, nonresponse adjustment and post-stratification. *Proceedings of the Section on Survey Research Methods*, American Statistical Association, 598-603.

Frankel, M.R. (1971). Inference from survey samples: An empirical investigation. Institute for Social Research, University of Michigan, Ann Arbor.

Fuller, W.A. (1966). Estimation employing post strata. *Journal of the America Statistical Association*, 61, 1172-1183.

Fuller, W.A. (1973). Regression for sample surveys. Paper presented at meeting of International Statistical Institute. August, 1973, Vienna, Austria.

Fuller, W.A. (1975). Regression analysis for sample survey. *Sankhyā*, Series C, 37, 117-132.

Fuller, W.A. (1984). Least squares and related analyses for complex survey designs. *Survey Methodology*, 10, 97-118.

Fuller, W.A., and An, A.B. (1998). Regression adjustments for nonresponse. *Journal of the Indian Society of Agricultural Statistics*, 51, 331-342.

Fuller, W.A., and Battese, G.E. (1973). Transformations for estimation of linear models with nested-error structure. *Journal of the American Statistical Association*, 68, 626-632.

Fuller, W.A., and Isaki, C.T. (1981). Survey design under superpopulation models. In *Current Topics in Survey Sampling*, (Eds., D. Krewski, J.N.K. Rao and R. Platek), New York: Academic Press, 199-226.

Fuller, W.A., Loughin, M.M. and Baker, H.D. (1994). Regression weighting for the 1987-88 National Food Consumption Survey, *Survey Methodology*, 20 75-85.

Fuller, W.A., and Rao, J.N.K. (2001). A regression composite estimator with application to the Canadian labour force survey. *Survey Methodology*, 27, 45-52.

Gambino, J., Kennedy, B. and Singh, M.P. (2001). Regression composite estimation for the Canadian labour force survey: Evaluation and implementation. *Survey Methodology*, 27, 65-74.

Gerow, K., and Mcculloch, C.E. (2000). Simultaneously model unbiased, design-unbiased estimation. *Biometrics* 56, 873-878.

Godambe, V.P. (1955). A unified theory of sampling from finite populations. *Journal of the Royal Statistical Society*, Series B, 17, 269-278.

Godambe, V.P., and Joshi, V.M. (1965). Admissibility and Bayes estimation in sampling finite populations, 1. *Annals of Mathematical Statistics*, 36, 1707-1722.

Goldberger, A.S. (1962). Best linear unbiased prediction in the generalized linear regression model. *Journal of the American Statistical Association*, 57, 369-375.

Graybill, F.A. (1976). *Theory and application of the linear model*. Wadsworth, Belmont, CA.

Hájek, J. (1959). Optimum strategy and other problems in probability sampling. *Casopis Pro Pestovani Matematiky*, 84, 387-423.

Hanurav, T.V. (1966). Some aspects of unified sampling theory. *Sankhyā*, Series A, 28, 175-204.

Henderson, C.R. (1963). Selection index and expected genetic advance. In *Statistical Genetics and Plant Breeding*, 141-163. National Academy Sciences, National Research Council Publication 982, Washington, DC.

Harville, D.A. (1976). Extension of the Gauss-Markov Theroem to include estimation of random effects. *Annals of Statistics*, 4, 384-395.

Hidiroglou, M.A. (1974). Estimation of regression parameters for finite populations. Ph.D. thesis, Iowa State University, Ames, Iowa.

Hidiroglou, M.A., Fuller, W.A. and Hickman, R.D. (1978). *Super Carp*, (sixth edition, 1980) Survey Section, Statistical Laboratory, Iowa State University, Ames, Iowa.

Hidiroglou, M.A., Särndal, C.-E. and Binder, D.A. (1995). Weighting and estimation in business surveys. *Business Survey Methods*, (Eds. Cox, Binder, Chinnappa, Colledge and Kott) New York: John Wiley & Sons, Inc., 477-502.

Holt, D., and Smith, T.M. F. (1979). Post Stratification. *Journal of the Royal Statistical Society*, Serie. A, 142, 33-46.

Horn, S.D., Horn, R.A. and Duncan, D.B. (1975). Estimating heteroscedastic variances in linear models. *Journal of the American Statistical Association*, 70, 380-385.

Huang, E.T., and Fuller, W.A. (1978). Nonnegative regression estimation for sample survey data. *Proceedings of the social statistics section,* American Statistical Association, 300-305.

Husain, M. (1969). Construction of regression weights for estimation in sample surveys. Unpublished M.S. thesis, Iowa State University, Ames, Iowa.

Isaki, C.T. (1970). Survey designs utilizing prior information. Unpublished Ph.D. thesis. Iowa State University.

Isaki, C.T., and Fuller, W.A. (1982). Survey design under the regression superpopulation model. *Journal of the American Statistical Association*, 77, 89-96.

Isaki, C.T., Tsay, J.H. and Fuller, W.A. (2000). Estimation of census adjustment factors. *Survey Methodology*, 26, 31-42.

Jessen, R.J. (1942). Statistical investigation of a sample survey for obtaining farm facts. *Iowa Agriculture Experiment Station Research Bulletin*, 304.

Kalton, G., and Maligalig, D.S. (1991). A Comparison of Methods of Weighting Adjustment for Nonresponse. *Proceedings of the 1991 Annual Research Conference*, U.S. Bureau of the Census, 409-428.

Kish, L., and Frankel, M.R. (1974). Inference from complex samples (with discussion). *Journal of the Royal Statistical Society*, Series B, 36, 1-37.

Konijn, H.S. (1962). Regression analysis for sample surveys. *Journal of the American Statistical Association*, 57, 590-606.

Kott, P.S. (1994). A note on handling nonresponse in sample surveys. *Journal of the American Statistical Association*, 89, 693-696.

Kuo, L. (1988). Classical and prediction approaches to estimating distribution function from survey data. *Proceedings of the Section on Survey Research Methods*, American Statistical Association, 280-285.

Lazzeroni, L.C., and Little, R.J.A. (1998). Random-effects models for smoothing post-stratification weights. *Journal of Official Statistics*, 14, 61-78.

Little, R.J.A. (1982). Models for nonresponse in sample surveys. *Journal of the American Statistical Association*, 77, 237-250.

Little, R.J.A. (1986). Survey nonresponse adjustments for estimates of means. *International Statistical Review*, 54, 139-157.

Little, R.J.A., and Rubin, D.B. (1987). *Statistical Analysis with Missing Data*. New York: John Wiley & Sons, Inc.

Little, R.J.A. (1993). Post-stratification: A modeler's perspective. *Journal of the American Statistical Association*, 88, 1001-1012.

Madow, W.G., and Madow, L.H. (1944). On the theory of systematic sampling. *Annals of Mathematical Statistics*, 15, 1-24.

Mickey, M.R. (1959). Some finite population unbiased ratio and regression estimators. *Journal of the American Statistical Association*, 54, 594-612.

Montanari, G.E. (1987). Post-sampling efficient Q-R prediction in large-sample surveys. *International Statistical Review*, 55, 191-202.

Montanari, G.E. (1999). A study on the conditional properties of finite population mean estimators. *Metron*, 57, 21-35.

Mukhopadhyay, P. (1993). Estimation of a finite population total under regression models: A review. *Sankhyā*, 55, 141-155.

Nieuwenbroek, N., Renssen, R. and Hofman, L. (2000). Towards a generalized weighting system. In *Proceedings of the Second International Conference on Establishment Surveys*, American Statistical Association, Alexandria, Virginia.

Park, M. (2002). Regression estimation of the mean in Survey Sampling. Unpublished Ph.D. dissertation, Iowa State University, Ames, Iowa.

Pfeffermann, D. (1984). Note on large sample properties of balanced samples. *Journal of the Royal Statistical Society*, Series B, 46, 38-41.

Rao, J.N.K. (1994). Estimating totals and distribution functions using auxiliary information at the estimation stage. *Journal of Official Statistics*, 10, 153-165.

Rao, J.N.K. (2002). *Small Area Estimation Theory and Methods*, New York: John Wiley & Sons, Inc.

Rao, J.N.K, Hartley, H.O. and Cochran, W.G. (1962). On a simple procedure of unequal probability sampling without replacement. *Journal of the Royal Statistical Society*, Series B, 24, 482-491.

Rao, J.N.K., and Singh, A.C. (1997). A ridge shrinkage method for range restricted weight calibration in survey sampling. *Proceedings of the section on survey research methods*, American Statistical Association, 57-64.

Robinson, G.K. (1991). The BLUP is a good thing: The estimation of random effects. *Statistical Science*, 6, 15-32.

Robinson, P.M., and Särndal, C.-E. (1983). Asymptotic properties of the generalized regression estimator in probability sampling. *Sankhyā*, Series B, 45, 240-248.

Rosenbaum, P.R., and Rubin, D.B. (1983). The central role of the propensity score in observational studies for casual effects. *Biometrika*, 70, 41-55.

Royall, R.M. (1970). On finite population sampling theory under certain linear regression models. *Biometrika*, 57, 377-387.

Royall, R.M. (1976). The linear least squares prediction approach to two-stage sampling. *Journal of the American Association*, 71, 657-664.

Royall, R.M. (1986). The prediction approach to robust variance estimation in two stage cluster sampling. *Journal of the American Statistical Association*, 81, 119-123.

Royall, R.M., and Cumberland, W.G. (1978). Variance estimation in finite population sampling. *Journal of the American Statistical Association*, 73, 351-358.

Royall, R.M., and Cumberland, W.G. (1981). The finite population linear regression estimator and estimators of its variance, an empirical study. *Journal of the American Statistical Association*, 76, 924-930.

Särndal, C.-E. (1980). On $\pi-$ weighting versus best linear unbiased weighting in probability sampling. *Biometrika*, 67, 639-650.

Särndal, C.-E. (1982). Implications of survey design for generalized regression estimation of linear functions. *Journal of Statistical Planning and Inference*, 7, 155-170.

Särndal, C.-E. (1996). Efficient estimators with simple variance in unequal probability sampling. *Journal of the American Statistics Association*, 91, 1289-1300.

Särndal, C.-E., Swenson, B. and Wretman, J.H. (1989). The weighted residual technique for estimating the variance of the general regression estimator of the finite population total. *Biometrika*, 76, 527-537.

Särndal, C.-E., Swenson, B. and Wretman, J.H. (1992). *Model Assisted Survey Sampling*. New York: Springer-Verlag.

Särndal, C.-E., and Wright, R. (1984). Cosmetic form of estimators in survey sampling. *Scandinavian Journal of Statistics*, 11, 146-156.

Scott, A., and Smith, T.M.F. (1974). Linear superpopulation models in survey and sampling. *Sankhyā*, C, 36, 143-146.

Scott, A., and Wu, C.F. (1981). On the asymptotic distribution of ratio and regression estimators. *Journal of the American Statistical Association*, 76, 98-102.

Silva, P.L.D.N., and Skinner, C.J. (1997). Variable selection for regression estimation in finite populations. *Survey Methodology*, 23, 23-32.

Singh, A.C., and Folsom, R.E. (2000). Bias corrected estimating functions approach for variance estimation adjusted for poststratification. *Proceedings of the Section on Survey Research Methods*, American Statistical Association, 610-615.

Singh, A.C., Kennedy, B. and Wu, S. (2001). Regression composite estimation for the Canadian Labour Force Survey with a rotating design. *Survey Methodology*, 27, 33-44.

Singh, A.C., and Mohl, C.A. (1996). Understanding calibration estimators in survey sampling. *Survey Methodology*, 22, 107-115.

Tallis, G.M. (1978). Note on robust estimation infinite populations. *Sankhyā*, C, 40, 136-138.

Tam, S.M. (1986). Characterization of best model-based predictors in survey sampling. *Biometrika*, 73, 232-235.

Théberge, A. (1999). Extensions of calibration estimators in survey sampling. *Journal of the American Statistical Association*, 94, 635-644.

Théberge, A. (2000). Calibration and restricted weights. *Survey Methodology*, 26, 99-107.

Tillé, Y. (1998). Estimation in surveys using conditional inclusion probabilities: Simple random sampling. *International Statistical Review*, 66, 303-322.

Tremblay, V. (1986). Practical Criteria for Definition of Weighting Classes. *Survey Methodology*, 12, 85-97.

Watson, D.J. (1937). The estimation of leaf area in field crops. *Journal of Agricultural Science*, 27, 474-483.

Woodruff, R.S., and Causey, B.D. (1976). Computerized method for approximating the variance of a complicated estimate. *Journal of the American Statistical Association*, 71, 315-321.

Wright, R.L. (1983). Finite population sampling with multivariate auxiliary information. *Journal of the American Statistical Association*, 78, 879-884.

Wu, C., and Sitter, R.R. (2001). A model-calibration approach to using complete auxiliary information from survey data. *Journal of the American Statistical Association*, 96, 185-193.

Yates, F. (1949). *Sampling Methods for Census and Surveys*. London: Griffin.

Yung, W., and Rao, J.N.K. (1996). Jackknife linearization variance estimators under stratified multi-stage sampling. *Survey Methodology*, 22, 23-31.

Zyskind, G. (1976). On canonical forms, non-negative covariance matrices and best and simple least squares linear estimators in linear models. *Annals of Mathematical Statistics*, 38, 1092-1109.