MODEL WEIGHTS FOR REGRESSION ESTIMATION

Wayne A. Fuller and Mingue Park
Statistical Laboratory, Snedecor Hall, Iowa State University, Ames, IA 50011-1210

Abstract:

Linear models that form the basis for survey regression estimation and the conditions under which the regression estimators are design consistent are reviewed. Model justification for some commonly used regression estimators is presented. Test for reduced models against design consistent models are discussed.

1. Introduction

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). Cochran (1942) gave the basic theory for regression in survey sampling relying 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.

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).

Various estimators have been proposed for estimating a finite population mean under a regression superpopulation model that postulates a relationship between the study variable and a set of auxiliary variables. Royall (1970, 1976) adapted linear prediction theory to the finite population situation and suggested the best linear model unbiased predictor (BLUP) for a finite population total. Cassel, Särndal and Wretman (1976) and Särndal (1980) proposed a generalized regression predictor that is asymptotically design unbiased and design consistent. Isaki and Fuller (1982) considered predictors of the regression type that are model unbiased and design consistent.Wright (1983) proposed a class of predictors, called QR-predictors, that contains most proposed predictors.

Inference based on prediction theory is sensitive to model misspecification, as illustrated by Hansen, Madow and Tepping (1983). Many techniques for robust inference have been suggested. See Royall (1992), Royall and Cumberland (1981a, 1981b), and Royall and Herson (1973a, 1973b).

Design consistency has been proposed (Isaki and Fuller, 1982; Robinson and Särndal, 1983) as a way of providing protection against model misspecification in the large sample setting. In this paper, we consider the problem of constructing estimators with good model properties, such as model unbiasedness and minimum model variance, that are also design consistent.

Assume the finite population is a realization from the superpopulation model

$$\mathbf{y}_N = \mathbf{X}_N\boldsymbol{\beta} + \mathbf{e}_N \ , \qquad (1)$$

$$\mathbf{e}_N \sim (\mathbf{0}\ ,\ \boldsymbol{\Sigma}_{eeN}) \ ,$$

where

$$\mathbf{y}_N = (y_1, \cdots, y_N)' \ ,$$

$$\mathbf{e}_N = (e_1, \cdots, e_N)' \ ,$$

$$\mathbf{X}_N = (\mathbf{x}_1', \cdots, \mathbf{x}_N')' \ ,$$

and

$$\mathbf{x}_i = (x_{i1}, \cdots, x_{ik}) \ .$$

We assume the covariance matrix $\boldsymbol{\Sigma}_{eeN}$ is a positive definite matrix. Expressions without the subscript N are used to denote the corresponding sample quantities, for example, $\mathbf{y} = (y_1, \cdots, y_n)'$ is the vector of sample observations.

To investigate the large sample properties of certain estimators, we define sequences of populations, samples and sampling designs. The set of indices for the N-th finite population is $U_N = \{1, \cdots, N\}$, where $N = 1, 2, \cdots$. Associated with j-th element of the N-th finite population is a vector of characteristics, denoted by $\mathbf{y}_{jN}$. Let $\mathcal{F}_N = \{\mathbf{y}_{1N}, \cdots, \mathbf{y}_{NN}\}$ be the set of vectors for the N-th finite population. The population mean of y for the N-th finite population is $\bar{y}_N = N^{-1}\sum_{i=1}^{N} y_{iN}$. Let A_N denote the set of indices appearing in the sample selected from the N-th finite population. The sample size is denoted by n_N. The sample size n_N is strictly less than N and $n_N \rightarrow \infty$ as $N \rightarrow \infty$. We assume that samples are selected according to the probability rule $P_N(\cdot)$.

Under the specified sequence of populations, samples, and sampling designs, we define a sequence of estimators

$\hat{\theta}_N$ of the population mean $\bar{y}_N$ to be design consistent, if for all $\epsilon > 0$,

$$\lim_{N\to\infty} P\left\{|\hat{\theta}_N - \bar{y}_N| > \epsilon \,\middle|\, \mathcal{F}_N\right\} = 0 \ ,$$

where the notation indicates that N-th finite population is held fixed and the probability depends only on the sampling design.

2. Design consistent regression estimators

Under some superpopulation models and designs, the BLUP constructed under the model is also design consistent. Assume the superpopulation model in which the covariance matrix of the errors is a multiple of the identity matrix and the column of ones is in the column space of $\mathbf{X}_N$. Without loss of generality, assume the first element of $\mathbf{x}_i$ is identically equal to one, and let $\mathbf{x}_i = (1, \mathbf{x}_{1,i})$. The regression estimator of the population mean of y is

$$\begin{aligned}\bar{y}_{reg} &= \bar{\mathbf{x}}_N \tilde{\boldsymbol{\beta}} \\ &= \bar{y}_n + (\bar{\mathbf{x}}_{1,N} - \bar{\mathbf{x}}_{1,n})\,\tilde{\boldsymbol{\beta}}_1 \ ,\end{aligned} \tag{2}$$

where

$$(\bar{y}_n \ , \ \bar{\mathbf{x}}_{1,n}) = n^{-1} \sum_{i\in A} (y_i \ , \ \mathbf{x}_i) \ ,$$

$$\tilde{\boldsymbol{\beta}} = \left(\tilde{\beta}_0 \ , \tilde{\boldsymbol{\beta}}_1'\right)' = (\mathbf{X}'\mathbf{X})^{-1}\,\mathbf{X}'\mathbf{y} \ ,$$

and $\bar{\mathbf{x}}_{1,N}$ is the population mean of $\mathbf{x}_{1,i}$. See Cochran (1977, p193). The estimator (2) is the BLUP under the model (1) and is also design consistent if the sample is selected by simple random sampling.

Many of the samples encountered in practice are more complicated than simple random nonreplacement samples. Theorem 1 gives conditions under which the regression estimator is design consistent.

Theorem 1. Let $\{\mathcal{F}_N\}$ be a sequence of finite populations, where $\mathcal{F}_N$ is a random sample of size N selected from an infinite superpopulation with finite fourth moments. Let $q_i = (y_i, \mathbf{x}_i)$ be a vector with mean $\bar{q}_N = (\bar{y}_N, \bar{\mathbf{x}}_N)$ for the N-th population. Let a sequence of probability samples be selected from the sequence $\{\mathcal{F}_N\}$. Define the regression estimator of $\bar{y}_N$ by

$$\bar{y}_{reg} = \bar{\mathbf{x}}_N \ \tilde{\boldsymbol{\beta}} \ ,$$

where $\tilde{\boldsymbol{\beta}}$ is a design consistent estimator of a parameter denoted by $\boldsymbol{\beta}_N$. Then

$$p\lim_{N\to\infty} \{(\bar{y}_{reg} - \bar{y}_N)\,|\mathcal{F}_N\} = 0 \ ,$$

if and only if

$$p\lim_{N\to\infty} \{\bar{e}_N|\mathcal{F}_N\} = 0 \ ,$$

where

$$e_i = y_i - \mathbf{x}_i\ \boldsymbol{\beta}_N \ .$$

■

Proof. Omitted.

Consider the construction of an estimator to meet the requirements of Theorem 1. Let a sample design have selection probabilities π_i and define the sample estimator of the mean by

$$\begin{aligned}(\bar{y}_\pi, \bar{\mathbf{x}}_\pi) &= \left(\sum_{i\in A} \pi_i^{-1}\right)^{-1} \sum_{i\in A} \pi_i^{-1}(y_i, \mathbf{x}_i) \\ &=: \sum_{i\in A} a_i(y_i, \mathbf{x}_i) \ ,\end{aligned} \tag{3}$$

where

$$a_i = \left(\sum_{j\in A} \pi_j^{-1}\right)^{-1} \pi_i^{-1} \ .$$

Assume the first element of $\mathbf{x}_i$ is identically equal to one. By analogy to (2), consider an estimator of the form

$$\bar{y}_{reg} = \bar{y}_\pi + (\bar{\mathbf{x}}_{1,N} - \bar{\mathbf{x}}_{1,\pi})\hat{\boldsymbol{\beta}}_1 \ . \tag{4}$$

Assume

$$(\bar{y}_\pi, \bar{\mathbf{x}}_{1,\pi}, \hat{\boldsymbol{\beta}}_1) - (\bar{y}_N, \bar{\mathbf{x}}_{1,N}, \boldsymbol{\beta}_{1,N})|\mathcal{F}_N = O_p(b_N) \ ,$$

where $b_N \to 0$ as $N \to \infty$. Then

$$\begin{aligned}\bar{y}_{reg} - \bar{y}_N &= \bar{y}_\pi - \bar{y}_N + (\bar{\mathbf{x}}_{1,N} - \bar{\mathbf{x}}_{1,\pi})\hat{\boldsymbol{\beta}}_1 \\ &= \bar{y}_\pi - \bar{y}_N + (\bar{\mathbf{x}}_{1,N} - \bar{\mathbf{x}}_{1,\pi})\boldsymbol{\beta}_{1,N} + O_p(b_N^2) \\ &= \bar{e}_\pi + O_p(b_N^2) \ ,\end{aligned} \tag{5}$$

where

$$e_i = y_i - \bar{y}_N - (\mathbf{x}_{1,i} - \bar{\mathbf{x}}_{1,N})\boldsymbol{\beta}_{1,N} \ . \tag{6}$$

The population mean of the e_i of (6) is zero for any $\boldsymbol{\beta}_{1,N}$. Therefore (4) gives a way to construct a design consistent estimator.

The estimator (4) is linear in y and can be written $\bar{y}_{reg} = \sum_{i\in A} w_i y_i$. Regression weights that define a regression estimator of the form (4) can be constructed by minimizing the quadratic objective function

$$(\mathbf{w} - \boldsymbol{\alpha})'\boldsymbol{\Phi}(\mathbf{w} - \boldsymbol{\alpha}) \tag{7}$$

subject to

$$\mathbf{w}'\mathbf{X} - \bar{\mathbf{x}}_N = \mathbf{0} \ , \tag{8}$$

where $\boldsymbol{\alpha}$ is a vector of initial weights and $\boldsymbol{\Phi}$ is a positive definite matrix. One possible $\boldsymbol{\Phi}$-matrix is a diagonal matrix with the diagonal elements being the initial weights. Possible initial weights are $\alpha_i = N^{-1}\pi_i^{-1}$ or a_i of (3).

We observed that the regression estimator (2) is the BLUP under the regression model with homogeneous variances and is also design consistent under simple random sampling. But, for some models and designs the estimator that is conditionally best, given $\mathbf{X}$, need not be a design consistent estimator. Assume the superpopulation model (1). Under the model, the unknown finite population mean is

$$\bar{y}_N = \bar{\mathbf{x}}_N \boldsymbol{\beta} + \bar{e}_N \ . \tag{9}$$

It follows that, under the model, the best linear conditionally unbiased predictor of $\bar{y}_N$, conditional on $\mathbf{X}$, is

$$\begin{aligned} \bar{y}_{BLUP} = N^{-1}\Bigg[\sum_{i\in A} y_i + (N-n)\bar{\mathbf{x}}_{N-n}\hat{\boldsymbol{\beta}} \ + \\ \mathbf{J}'_{N-n}\boldsymbol{\Sigma}_{ee\bar{A}A}\boldsymbol{\Sigma}_{ee}^{-1}\left(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}}\right)\Bigg] \ , \end{aligned} \tag{10}$$

where

$$\hat{\boldsymbol{\beta}} = \left(\mathbf{X}'\boldsymbol{\Sigma}_{ee}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}'\boldsymbol{\Sigma}_{ee}^{-1}\mathbf{y} \ , \tag{11}$$

$$\bar{\mathbf{x}}_{N-n} = (N-n)^{-1}(N\bar{\mathbf{x}}_N - n\bar{\mathbf{x}}_n) \ ,$$

$$\boldsymbol{\Sigma}_{ee\bar{A}A} = \mathrm{E}\left\{\mathbf{e}_{\bar{A}}\mathbf{e}'\right\} \ ,$$

$$\mathbf{e}_{\bar{A}} = (e_{n+1}, \cdots, e_N)' \ ,$$

$\mathbf{J}_{N-n}$ is an $N-n$ dimensional column vector of ones, and $\bar{A}$ is the set of elements in U that are not in A. See Royall (1976). The estimator (10) will be design consistent if the design probabilities, the matrix $\boldsymbol{\Sigma}_{eeN}$ and the matrix $\mathbf{X}_N$ meet certain conditions. These conditions 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). We summarize the results in Theorem 2.

Theorem 2. Let the superpopulation model be given by (1). Assume a sequence of populations, designs and estimators such that

$$\begin{aligned} &[(\bar{y}_{HT}, \bar{\mathbf{x}}_{HT}) - (\bar{y}_N, \bar{\mathbf{x}}_N)] \mid \mathcal{F}_N \\ =&N^{-1}\left\{\sum_{i=1}^{n}\pi_i^{-1}(y_i, \mathbf{x}_i) - (T_{y,N}, \mathbf{T}_{x,N})\right\} \Bigg| \ \mathcal{F}_N \\ =&O_p(n_N^{-\alpha}) \ , \end{aligned} \tag{12}$$

where the π_i are the inclusion probabilities, $(T_{y,N}, \mathbf{T}_{x,N})$ is the total of $(y, \mathbf{x})$ for the N-th population and $\alpha > 0$. Let $\hat{\boldsymbol{\beta}}$ be defined by (11) and let $\{\boldsymbol{\beta}_N\}$ be a sequence such that

$$(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_N) \mid \mathcal{F}_N = O_p(n_N^{-\alpha}) \ . \tag{13}$$

Assume there is a sequence $\{\boldsymbol{\gamma}_N\}$ such that

$$\mathbf{X}\boldsymbol{\gamma}_N = \boldsymbol{\Sigma}_{ee}\mathbf{L}_\pi \ , \tag{14}$$

where $\mathbf{L}_\pi = (\pi_1^{-1}, \cdots, \pi_n^{-1})'$, for every sample form U_N that is possible under the design. Then

$$\left(\bar{\mathbf{x}}_N\hat{\boldsymbol{\beta}} - \bar{y}_N\right) \mid \mathcal{F}_N = O_p(n_N^{-\alpha}) \ . \tag{15}$$

If, in addition

$$\mathbf{X}\boldsymbol{\eta}_N = \boldsymbol{\Sigma}_{ee}\mathbf{J}_n + \boldsymbol{\Sigma}_{eeA\bar{A}}\mathbf{J}_{N-n} \ , \tag{16}$$

where $\mathbf{J}_n$ is a n-dimensional column vector of ones, for a sequence $\{\boldsymbol{\eta}_N\}$ and all possible samples, then $\bar{y}_{BLUP}$ of (10) satisfies

$$(\bar{y}_{BLUP} - \bar{y}_N) \mid \mathcal{F}_N = O_p(n_N^{-\alpha}) \ . \tag{17}$$

Assume there is a sequence $\{\boldsymbol{\zeta}_N\}$ such that

$$\mathbf{X}\boldsymbol{\zeta}_N = \boldsymbol{\Sigma}_{ee}\left(\mathbf{L}_\pi - \mathbf{J}_n\right) - \boldsymbol{\Sigma}_{eeA\bar{A}}\mathbf{J}_{N-n} \ , \tag{18}$$

for every sample from U_N that is possible under the design. Then $\bar{y}_{BLUP}$ of (10) is expressible as

$$\begin{aligned} \bar{y}_{BLUP} &= \bar{y}_{HT} + N^{-1}(N-n)\left(\bar{\mathbf{x}}_{N-n} - \bar{\mathbf{x}}_c\right)\hat{\boldsymbol{\beta}} \\ &= \bar{y}_{HT} + \left(\bar{\mathbf{x}}_N - \bar{\mathbf{x}}_{HT}\right)\hat{\boldsymbol{\beta}} \ , \end{aligned} \tag{19}$$

and

$$(\bar{y}_{BLUP} - \bar{y}_N)\big|\mathcal{F}_N = O_p\left(n_N^{-\alpha}\right) \ , \tag{20}$$

where

$$\bar{\mathbf{x}}_c = (N-n)^{-1}\sum_{i\in A}\left(\pi_i^{-1} - 1\right)\mathbf{x}_i \ .$$

Proof. The sufficient condition for the estimator to be design consistent given in Theorem 1 is

$$p\lim_{N\to\infty}\left(\bar{y}_N - \bar{\mathbf{x}}_N\boldsymbol{\beta}_N \mid \mathcal{F}_N\right) = 0 \ . \tag{21}$$

By assumption (12) and (13), a sufficient condition for (21) is

$$p\lim_{N\to\infty}\left(\bar{y}_{HT} - \bar{\mathbf{x}}_{HT}\hat{\boldsymbol{\beta}} \mid \mathcal{F}_N\right) = 0 \ . \tag{22}$$

A sufficient condition for (22) is

$$\sum_{i=1}^{n}\left(y_i - \mathbf{x}_i\hat{\boldsymbol{\beta}}\right)\ \pi_i^{-1} = 0 \ , \tag{23}$$

for all A with positive probability. By the properties of the generalized least square estimator of (11),

$$\left(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}\right)'\boldsymbol{\Sigma}_{ee}^{-1}\mathbf{X} = \mathbf{0} \ ,$$

for every $\mathbf{X}$ such that $\mathbf{X}'\boldsymbol{\Sigma}_{ee}^{-1}\mathbf{X}$ is nonsingular. Therefore, if there is a $\boldsymbol{\gamma}_N$ such that

$$\boldsymbol{\Sigma}_{ee}^{-1}\mathbf{X}\boldsymbol{\gamma}_N = \mathbf{L}_\pi \ ,$$

condition (23) is satisfied. By assumptions (12), (13) and (14),

$$\begin{aligned}
&\bar{\mathbf{x}}_N\hat{\boldsymbol{\beta}} - \bar{y}_N \\
=&\left(\bar{\mathbf{x}}_N - \bar{\mathbf{x}}_{HT}\right)\hat{\boldsymbol{\beta}} + \left(\bar{y}_{HT} - \bar{y}_N\right) - \left(\bar{y}_{HT} - \bar{\mathbf{x}}_{HT}\hat{\boldsymbol{\beta}}\right) \\
=&\left(\bar{\mathbf{x}}_N - \bar{\mathbf{x}}_{HT}\right)\boldsymbol{\beta}_N + \left(\bar{y}_{HT} - \bar{y}_N\right) \\
&+ \left(\bar{\mathbf{x}}_N - \bar{\mathbf{x}}_{HT}\right)\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_N\right) \\
=&O_p(n_N^{-\alpha}) \ .
\end{aligned} \tag{24}$$

If (16) is satisfied,

$$\begin{aligned}
\bar{y}_{BLUP} =& \Big[\bar{\mathbf{x}}_N\hat{\boldsymbol{\beta}} + N^{-1}\left(\mathbf{J}_n'\boldsymbol{\Sigma}_{ee} + \mathbf{J}_{N-n}'\boldsymbol{\Sigma}_{ee\bar{A}A}\right) \\
&\times \boldsymbol{\Sigma}_{ee}^{-1}\left(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}\right)\Big] \\
=&\bar{\mathbf{x}}_N\hat{\boldsymbol{\beta}} \ .
\end{aligned} \tag{25}$$

Result (17) follows from (24) and (25).

If (18) is satisfied,

$$\begin{aligned}
0 =&N^{-1}\left(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}\right)'\boldsymbol{\Sigma}_{ee}^{-1}\mathbf{X}\boldsymbol{\zeta}_N \\
=&N^{-1}\left(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}\right)'\left[(\mathbf{L}_\pi - \mathbf{J}) - \boldsymbol{\Sigma}_{ee}^{-1}\boldsymbol{\Sigma}_{eeA\bar{A}}\mathbf{J}_{N-n}\right] \\
=&N^{-1}(N-n)\left(\bar{\mathbf{y}}_c - \bar{\mathbf{x}}_c\hat{\boldsymbol{\beta}}\right) \\
&- N^{-1}\left(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}\right)'\boldsymbol{\Sigma}_{ee}^{-1}\boldsymbol{\Sigma}_{eeA\bar{A}}\mathbf{J}_{N-n} \ .
\end{aligned}$$

It follows that $\bar{y}_{BLUP}$ of (10) is

$$\begin{aligned}
\bar{y}_{BLUP} =&N^{-1}\Big[\sum_{i\in A} y_i + (N-n)\bar{y}_c \\
&+ (N-n)\left(\bar{\mathbf{x}}_{N-n} - \bar{\mathbf{x}}_c\right)\hat{\boldsymbol{\beta}}\Big] \\
=&\bar{y}_{HT} + N^{-1}(N-n)\left(\bar{\mathbf{x}}_{N-n} - \bar{\mathbf{x}}_c\right)\hat{\boldsymbol{\beta}} \\
=&\bar{y}_{HT} + \left(\bar{\mathbf{x}}_N - \bar{\mathbf{x}}_{HT}\right)\hat{\boldsymbol{\beta}} \ .
\end{aligned}$$

The error in the predictor is $O_p(n_N^{-\alpha})$ because of assumptions (12) and (13). ■

Theorem 2 gives the conditions under which the best linear unbiased predictor is design consistent. Especially, if (18) is satisfied, the estimator is the regression estimator with the coefficient estimated by the generalized least squares estimator. Theorem 2 also gives a way of constructing the model based design consistent estimator.

Montanari (1987) introduced the *general QR-predictor* as an extension of the *QR-predictor* of Wright (1983) and gave conditions under which the general QR-predictor is design consistent. The general QR-predictor is

$$\bar{y}_{QR} = \bar{\mathbf{x}}_N\hat{\boldsymbol{\beta}} + N^{-1}\sum_{i\in A} r_i\left(y_i - \mathbf{x}_i\hat{\boldsymbol{\beta}}\right) \ , \tag{26}$$

where

$$\hat{\boldsymbol{\beta}} = \left(\mathbf{X}'\mathbf{Q}\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{Q}\mathbf{y} \ ,$$

and $\mathbf{Q}$ is $n\times n$ matrix whose (i,j)-th element denoted by q_{ij}. Conditions under which the estimator (26) is design consistent are

$$p\lim_{N\to\infty}\hat{\boldsymbol{\beta}} = \mathbf{B}_N \tag{27}$$

and

$$\mathbf{c} \in \mathcal{C}(\mathbf{W}_N\mathbf{X}_N) \ , \tag{28}$$

where

$$\mathbf{B}_N = \left(\mathbf{X}_N'\mathbf{W}_N\mathbf{X}_N\right)^{-1}\mathbf{X}_N'\mathbf{W}_N\mathbf{y}_N \ ,$$

$$\mathbf{c} = (c_1, \cdots, c_N)' \ ,$$

$$c_i = 1 - r_i\pi_i \ ,$$

$\mathbf{W}_N$ is $N\times N$ symmetric matrix whose (i,j)-th entry is $q_{ij}\pi_{ij}$ and $\mathcal{C}(\mathbf{X})$ denotes the space spanned by the columns of $\mathbf{X}$.

A specification of $\boldsymbol{\Sigma}_{eeN}$ may be particularly appropriate for two-stage cluster samples, See Royall (1976) and Montanari (1987). A reasonable model is that in which there is common correlation among items in the same primary sampling units and zero correlation between units in different primary sampling units. That is, a potential model for the j-th observation in cluster i is

$$y_{ij} = \mathbf{x}_{ij}\boldsymbol{\beta} + u_{ij} \ , \tag{29}$$

$$u_{ij} = b_i + e_{ij} \ ,$$

$$b_i \sim II(0, \sigma_b^2) \ ,$$

$$e_{ij} \sim II(0, \sigma_e^2) \ ,$$

where e_{ij} is independent of b_k for all i, j and k. Under the model (29), the BLUP defined in (10) is a general QR-predictor with

$$\mathbf{r} = (r_1, \cdots, r_n) = \left(\mathbf{J}_n + \boldsymbol{\Sigma}_{ee}^{-1}\boldsymbol{\Sigma}_{eeA\bar{A}}\mathbf{J}_{N-n}\right) \ ,$$

and

$$\mathbf{Q} = \boldsymbol{\Sigma}_{ee} \ ,$$

where $\boldsymbol{\Sigma}_{ee}$ is a block diagonal matrix in which the i-th block is a $m_i\times m_i$ matrix

$$\sigma_e^2\mathbf{I}_{m_i} + \sigma_b^2\mathbf{J}_{m_i}\mathbf{J}_{m_i}' \ ,$$

and m_i is the number of sampled elements in cluster i. Let a sample of primary sampling units be selected by unequal probability sampling design and a simple random nonreplacement sample of secondary sampling units be selected within a selected primary sampling unit. Under the specified model and design, the condition for the BLUP of the finite population mean to be design consistent given by Montanari (1987) is equivalent to the condition given in (18). Although the two conditions are equivalent under the specified set up, the equivalence of the two conditions does not generally hold. The condition in (18) is easier to check than the condition (28) because we only need the first order inclusion probabilities, the values of auxiliary variables corresponding to sampled elements and the covariance matrix of $\mathbf{e}_N$. Note that condition (28) is a function of $\mathbf{X}_N$, $\mathbf{Q}$, r_i and the first and the second order inclusion probabilities for elements in population.

3. Mixed model regression estimation

The regression model with random components has been heavily used for small area estimation. See Rao (2002). Royall (1976) and Montanari (1987) used the model for cluster samples. The model has also been used for mean estimation in the post stratification setting. See Little (1993) and Lazzeroni and Little (1998).

We consider the mixed model

$$\mathbf{y} = \mathbf{X}_0\boldsymbol{\beta}_0 + \mathbf{X}_2\boldsymbol{\beta}_2 + \mathbf{e} \ , \qquad (30)$$

where $\boldsymbol{\beta}_2 \sim (\mathbf{0}, \boldsymbol{\Psi})$, $\mathbf{e} \sim (\mathbf{0}, \boldsymbol{\Phi})$, $\mathbf{X}_0$ is an $n \times l$ matrix, $\mathbf{X}_2$ is an $n \times (p-l)$ matrix, the random vector $\boldsymbol{\beta}_2$ is independent of $\mathbf{e}$, and $\boldsymbol{\beta}_0$ is a fixed vector.

The best linear model unbiased predictor of $\bar{\mathbf{x}}_N\boldsymbol{\beta}$ is $\mathbf{w}'\mathbf{y}$, where the vector $\mathbf{w}$ is chosen to minimize

$$\mathrm{V}\{\mathbf{w}'\mathbf{y} - \bar{\mathbf{x}}_N\boldsymbol{\beta}\} = \mathrm{V}\{\mathbf{w}'\mathbf{e} + (\mathbf{w}'\mathbf{X}_2 - \bar{\mathbf{x}}_{2,N})\boldsymbol{\beta}_2\} \quad (31)$$

subject to the constraint

$$\mathbf{w}'\mathbf{X}_0 = \bar{\mathbf{x}}_{0,N} \qquad (32)$$

and $\bar{\mathbf{x}}_N = (\bar{\mathbf{x}}_{0,N}, \bar{\mathbf{x}}_{2,N})$ is the population mean of $\mathbf{x}$. If $\boldsymbol{\alpha}$ is a vector of preliminary weights and if $\boldsymbol{\Phi\alpha}$ is in the column space of $\mathbf{X}_0$ then the vector $\mathbf{w}$ can be obtained from the Lagrangean

$$\begin{aligned} Q = {} & (\mathbf{w}-\boldsymbol{\alpha})'\boldsymbol{\Phi}(\mathbf{w}-\boldsymbol{\alpha}) + (\mathbf{w}'\mathbf{X}_2 - \bar{\mathbf{x}}_{2,N})\boldsymbol{\Psi} \\ & \times (\mathbf{w}'\mathbf{X}_2 - \bar{\mathbf{x}}_{2,N})' + 2\boldsymbol{\lambda}'(\mathbf{w}'\mathbf{X}_0 - \bar{\mathbf{x}}_{0,N})' \ , \end{aligned} \qquad (33)$$

where $\boldsymbol{\lambda}$ is a vector of Lagrangian multipliers. The partial derivatives with respect to $\mathbf{w}$ and $\boldsymbol{\lambda}$ are

$$\frac{1}{2}\frac{\partial Q}{\partial \mathbf{w}} = \boldsymbol{\Phi}\mathbf{w} - \boldsymbol{\Phi\alpha} + \mathbf{X}_2\boldsymbol{\Psi}(\mathbf{X}_2'\mathbf{w} - \bar{\mathbf{x}}_{2,N}') + \mathbf{X}_0\boldsymbol{\lambda} \ , \quad (34)$$

and

$$\frac{1}{2}\frac{\partial Q}{\partial \boldsymbol{\lambda}} = \mathbf{X}_0'\mathbf{w} - \bar{\mathbf{x}}_{0,N}' \ .$$

If we multiply (34) by $\mathbf{X}_2'\boldsymbol{\Phi}^{-1}$, multiply (34) by $\mathbf{X}_0'\boldsymbol{\Phi}^{-1}$ and set the results equal to zero, we obtain the linear equation

$$\begin{aligned} & \begin{bmatrix} \mathbf{X}_0'\boldsymbol{\Phi}^{-1}\mathbf{X}_0 & , & \mathbf{X}_0'\boldsymbol{\Phi}^{-1}\mathbf{X}_2\boldsymbol{\Psi} \\ \mathbf{X}_2'\boldsymbol{\Phi}^{-1}\mathbf{X}_0 & , & \mathbf{I} + \mathbf{X}_2'\boldsymbol{\Phi}^{-1}\mathbf{X}_2\boldsymbol{\Psi} \end{bmatrix} \begin{bmatrix} \boldsymbol{\lambda} \\ \mathbf{X}_2'\mathbf{w} - \boldsymbol{\mu}_{x_2}' \end{bmatrix} \\ = & \quad \begin{bmatrix} \bar{\mathbf{x}}_{0,\pi}' - \bar{\mathbf{x}}_{0,N}' \\ \bar{\mathbf{x}}_{2,\pi}' - \bar{\mathbf{x}}_{2,N}' \end{bmatrix} \ , \end{aligned} \qquad (35)$$

where $(\bar{\mathbf{x}}_{0,\pi}, \bar{\mathbf{x}}_{2,\pi}) = \boldsymbol{\alpha}'(\mathbf{X}_0, \mathbf{X}_2)$. Thus, the vector of weights that minimizes the objective function Q is

$$\begin{aligned} \mathbf{w} = & \boldsymbol{\alpha} - \boldsymbol{\Phi}^{-1}\mathbf{X}_2\boldsymbol{\Psi}(\mathbf{X}_2'\mathbf{w} - \bar{\mathbf{x}}_{2,N}') - \boldsymbol{\Phi}^{-1}\mathbf{X}_0\boldsymbol{\lambda} \\ = & \boldsymbol{\alpha} - \left[\boldsymbol{\Phi}^{-1}\mathbf{X}_0 \quad , \quad \boldsymbol{\Phi}^{-1}\mathbf{X}_2\boldsymbol{\Psi}\right] \begin{bmatrix} \boldsymbol{\lambda} \\ \mathbf{X}_2'\mathbf{w} - \bar{\mathbf{x}}_{2,N}' \end{bmatrix} \\ = & \boldsymbol{\alpha} + \boldsymbol{\Phi}^{-1}\mathbf{X} \begin{bmatrix} \mathbf{X}_0'\boldsymbol{\Phi}^{-1}\mathbf{X}_0\mathbf{X}_0'\boldsymbol{\Phi}^{-1}\mathbf{X}_2 \\ \mathbf{X}_2'\boldsymbol{\Phi}^{-1}\mathbf{X}_0\boldsymbol{\Psi}^{-1} + \mathbf{X}_2'\boldsymbol{\Phi}^{-1}\mathbf{X}_2 \end{bmatrix}^{-1} \\ & \times \begin{bmatrix} \bar{\mathbf{x}}_{2,N}' - \bar{\mathbf{x}}_{0,\pi}' \\ \bar{\mathbf{x}}_{2,N}' - \bar{\mathbf{x}}_{2,\pi}' \end{bmatrix} \ , \end{aligned} \qquad (36)$$

where

$$\begin{aligned} \boldsymbol{\lambda} = \mathbf{Q}^{-1} \Big\{ & \left(\bar{\mathbf{x}}_{0,\pi}' - \bar{\mathbf{x}}_{2,N}\right) - \mathbf{X}_0'\boldsymbol{\Phi}^{-1}\mathbf{X}_2 \\ & \times \left(\boldsymbol{\Psi}^{-1} + \mathbf{X}_2'\boldsymbol{\Phi}^{-1}\mathbf{X}_2\right)^{-1} \left(\bar{\mathbf{x}}_{2,\pi}' - \bar{\mathbf{x}}_{2,N}\right) \Big\} \ , \end{aligned}$$

and

$$\begin{aligned} \mathbf{Q} = & \mathbf{X}_0'\boldsymbol{\Phi}^{-1}\mathbf{X}_0 \\ & - \mathbf{X}_0'\boldsymbol{\Phi}^{-1}\mathbf{X}_2\left(\boldsymbol{\Psi}^{-1} + \mathbf{X}_2'\boldsymbol{\Phi}^{-1}\mathbf{X}_2\right)^{-1}\mathbf{X}_2'\boldsymbol{\Phi}^{-1}\mathbf{X}_0 \ . \end{aligned}$$

The estimator defined with the vector of weights (36) is

$$\bar{y}_{rreg} = \bar{y}_\pi + (\bar{\mathbf{x}}_N - \bar{\mathbf{x}}_\pi)\hat{\boldsymbol{\theta}} \ , \qquad (37)$$

where

$$\hat{\boldsymbol{\theta}} = \mathbf{H}_{\Psi xx}^{-1}\mathbf{X}'\boldsymbol{\Phi}^{-1}\mathbf{y} \ ,$$

and

$$\mathbf{H}_{\Psi xx} = \begin{bmatrix} \mathbf{X}_0'\boldsymbol{\Phi}^{-1}\mathbf{X}_0 & , & \mathbf{X}_0'\boldsymbol{\Phi}^{-1}\mathbf{X}_2 \\ \mathbf{X}_2'\boldsymbol{\Phi}^{-1}\mathbf{X}_0 & , & \boldsymbol{\Psi}^{-1} + \mathbf{X}_2'\boldsymbol{\Phi}^{-1}\mathbf{X}_2 \end{bmatrix} \ .$$

The estimator defined in (37) is a design consistent estimator and is the best predictor under the mixed model.

Estimation for the population mean in the present context differs from the situation under the model (29) in that the population mean of $\mathbf{X}_2$ is assumed to be known in estimation under model (30). Thus, in the estimation under model (30) we are estimating a linear combination of fixed and random effects, while the regression estimator under model (29) is an estimator of fixed effects.

To derive the large sample properties of the estimator, consider a sequence of $\mathbf{\Psi}_N$. If $\mathbf{\Psi}_N^{-1}$ is increasing at the same rate as the sample size n, the estimator (37) is a design consistent estimator of the population mean of y.

Theorem 3. Let $\{\mathcal{F}_N, A_N, \mathbf{\Phi}_N, \mathbf{\Psi}_N\}$ be a sequence of populations, samples, and positive definite matrices such that

$$[(\bar{y}_\pi\ ,\ \bar{\mathbf{x}}_\pi) - (\bar{y}_N\ ,\ \bar{\mathbf{x}}_N)]\,|\mathcal{F}_N = O_p\left(n_N^{-\frac{1}{2}}\right) \quad . \tag{38}$$

where n_N is the sample size for the N-th sample and $(\bar{y}_N, \bar{\mathbf{x}}_N)$ is the population mean of $(y, \mathbf{x})$. Assume there exists a sequence $\mathbf{Q}_{z\phi z,N}$ such that

$$\left[n_N^{-1}\mathbf{Z}'\mathbf{\Phi}_N^{-1}\mathbf{Z} - \mathbf{Q}_{z\phi z,N}\right]\,|\mathcal{F}_N = O_p\left(n_N^{-\frac{1}{2}}\right) \tag{39}$$

and

$$\lim_{N\to\infty} N^{-1}\mathbf{Q}_{z\phi z,N} = \mathbf{Q}_{z\phi z} \quad , \tag{40}$$

where $\mathbf{Z} = (\mathbf{z}_1', \cdots \mathbf{z}_n')'$, $\mathbf{z}_i = (y_i, \mathbf{x}_i)$ and $\mathbf{Q}_{z\phi z}$ is a positive definite matrix. Assume

$$\lim_{N\to\infty} n_N^{-1}\left[\mathbf{\Psi}_N^{-1}\right] = \mathbf{\Psi}^{-1} \quad , \tag{41}$$

where $\mathbf{\Psi}$ is a positive definite matrix. Then, estimator (37) satisfies

$$\begin{aligned}\bar{y}_{rreg} - \bar{y}_N &= \bar{y}_\pi - \bar{y}_N + (\bar{\mathbf{x}}_N - \bar{\mathbf{x}}_\pi)\boldsymbol{\theta}_N + O_p\left(n_N^{-1}\right)\\ &= O_p\left(n_N^{-\frac{1}{2}}\right) \quad ,\end{aligned} \tag{42}$$

where

$$\begin{aligned}\boldsymbol{\theta}_N &= \left[\mathbf{Q}_{x\phi x,N} + \mathbf{\Lambda}_{\phi,N}\right]^{-1}\mathbf{Q}_{x\phi y,N}\\ &= \mathbf{H}_{x\psi x,N}^{-1}\mathbf{Q}_{x\phi y,N} \quad ,\end{aligned}$$

$$\mathbf{H}_{x\psi x,N} = \mathbf{Q}_{x\phi x,N} + \mathbf{\Lambda}_{\phi,N} \quad ,$$

and

$$\mathbf{\Lambda}_{\phi,N} = \frac{1}{n_N}\begin{bmatrix}\mathbf{0} & , & \mathbf{0}\\ \mathbf{0} & , & \mathbf{\Psi}_N^{-1}\end{bmatrix} \quad .$$

Proof. The estimator (37) is

$$\bar{y}_{rreg} = \bar{y}_\pi + (\bar{\mathbf{x}}_N - \bar{\mathbf{x}}_\pi)\boldsymbol{\theta}_N + (\bar{\mathbf{x}}_N - \bar{\mathbf{x}}_\pi)\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_N\right) \quad . \tag{43}$$

The population characteristic $\boldsymbol{\theta}_N$ is

$$\begin{aligned}\boldsymbol{\theta}_N =& \mathbf{H}_{x\psi x,N}^{-1}\mathbf{Q}_{x\phi y,N}\\ =& \mathbf{H}_{x\psi x,N}^{-1}\left[\mathbf{Q}_{x\phi x,N}\boldsymbol{\theta}_N + \mathbf{Q}_{x\phi a,N} + \mathbf{\Lambda}_N\boldsymbol{\theta}_N - \mathbf{\Lambda}_N\boldsymbol{\theta}_N\right]\\ =& \boldsymbol{\theta}_N + \mathbf{H}_{x\psi x,N}^{-1}\left[\mathbf{Q}_{x\phi a,N} - \mathbf{\Lambda}_N\boldsymbol{\theta}_N\right] \quad ,\end{aligned} \tag{44}$$

where $a_i = y_i - \mathbf{x}\boldsymbol{\theta}_N$ and $\mathbf{Q}_{x\phi a,N} = \mathbf{Q}_{x\phi y,N} - \mathbf{Q}_{x\phi x,N}\boldsymbol{\theta}_N$. By (44),

$$\mathbf{Q}_{x\phi a,N} - \mathbf{\Lambda}_N\boldsymbol{\theta}_N = \mathbf{0} \quad .$$

The error of $\hat{\boldsymbol{\theta}}$ in estimating $\boldsymbol{\theta}_N$ is

$$\begin{aligned}\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_N &= \left(\frac{1}{n_N}\mathbf{H}_{x\psi x}\right)^{-1}\frac{1}{n_N}\mathbf{Q}_{x\phi y} - \boldsymbol{\theta}_N\\ &= \left(\frac{1}{n_N}\mathbf{H}_{x\psi x}\right)^{-1}\left\{\frac{1}{n_N}\mathbf{Q}_{x\phi x}\boldsymbol{\theta}_N + \frac{1}{n_N}\mathbf{Q}_{x\phi a}\right.\\ &\qquad\left. - \frac{1}{n_N}\mathbf{Q}_{x\phi x}\boldsymbol{\theta}_N - \mathbf{\Lambda}_N\boldsymbol{\theta}_N\right\}\\ &= \left(\frac{1}{n_N}\mathbf{H}_{x\psi x}\right)^{-1}\left(\frac{1}{n_N}\mathbf{Q}_{x\phi a} - \mathbf{\Lambda}_N\boldsymbol{\theta}_N\right)\\ &= \left(\frac{1}{n_N}\mathbf{H}_{x\psi x}\right)^{-1}\left(\frac{1}{n_N}\mathbf{Q}_{x\phi a} - \mathbf{Q}_{x\phi a,N}\right) \quad ,\end{aligned} \tag{45}$$

where $\mathbf{Q}_{x\phi x} = \mathbf{X}'\mathbf{\Phi}^{-1}\mathbf{X}$, $\mathbf{Q}_{x\phi y} = \mathbf{X}'\mathbf{\Phi}^{-1}\mathbf{y}$ and $\mathbf{Q}_{x\phi a} = \mathbf{Q}_{x\phi y} - \mathbf{Q}_{x\phi x}\boldsymbol{\theta}_N$. By the assumption (39),

$$\left(\frac{1}{n_N}\mathbf{Q}_{x\phi a} - \mathbf{Q}_{x\phi a,N}\right) = O_p\left(n_N^{-\frac{1}{2}}\right) \quad ,$$

and $\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_N = O_p\left(n_N^{-\frac{1}{2}}\right)$ because $\left({n_N}^{-1}\mathbf{H}_{x\psi x}\right)$ is bounded. The result follows from (39) and (43). ∎

If $\mathbf{\Psi}_N$ is fixed or $\mathbf{\Psi}_N \to \infty$ as $n \to \infty$, then the estimator $\hat{\boldsymbol{\theta}}$ approaches the weighted least squares estimator and we obtain design consistency for $\bar{y}_{rreg}$ because $\hat{\boldsymbol{\theta}}$ converges to the population analog of the weighted least square estimator.

The proof of Theorem 4 is a proof of the assertion that the estimator $\bar{y}_{rreg}$ defined in (37) is the best linear conditionally unbiased predictor for the population mean of y under the mixed model.

Theorem 4. Consider the mixed model (30). Assume that there exists column vector $\mathbf{c}_1$ such that

$$\mathbf{\Phi}\boldsymbol{\alpha} = \mathbf{X}_0\mathbf{c}_1 \quad .$$

Let $\bar{y}_{rreg} = \mathbf{w}'\mathbf{y}$, where $\mathbf{w}$ is the vector of weights that minimizes

$$(\mathbf{w}-\boldsymbol{\alpha})'\mathbf{\Phi}(\mathbf{w}-\boldsymbol{\alpha}) + (\mathbf{w}'\mathbf{X}_2 - \bar{\mathbf{x}}_{2,N})\mathbf{\Psi}(\mathbf{w}'\mathbf{X}_2 - \bar{\mathbf{x}}_{2,N})' \quad , \tag{46}$$

subject to

$$\mathbf{w}'\mathbf{X}_0 - \bar{\mathbf{x}}_{0,N} = \mathbf{0} \quad , \tag{47}$$

where $\bar{\mathbf{x}}_N = (\bar{\mathbf{x}}_{0,N}\ ,\ \bar{\mathbf{x}}_{2,N})$ is the population mean of $\mathbf{x}$. Then $\bar{y}_{rreg}$ is the best linear conditionally unbiased predictor of $\bar{\mathbf{x}}_{0,N}\boldsymbol{\beta}_0 + \bar{\mathbf{x}}_{2,N}\boldsymbol{\beta}_2$.

Proof. Under the restriction (47), the objective function (46) is

$$
\begin{aligned}
&\mathbf{w}'\mathbf{\Phi}\mathbf{w} - 2\mathbf{w}'\mathbf{\Phi}\boldsymbol{\alpha} + \boldsymbol{\alpha}'\mathbf{\Phi}\boldsymbol{\alpha} \\
&\quad + (\mathbf{w}'\mathbf{X}_2 - \bar{\mathbf{x}}_{2,N})\mathbf{\Psi}(\mathbf{w}'\mathbf{X}_2 - \bar{\mathbf{x}}_{2,N})' \\
&= \mathbf{w}'\mathbf{\Phi}\mathbf{w} - 2\mathbf{w}'\mathbf{X}_0\mathbf{c}_1 + \boldsymbol{\alpha}'\mathbf{\Phi}\boldsymbol{\alpha} \\
&\quad + (\mathbf{w}'\mathbf{X}_2 - \bar{\mathbf{x}}_{2,N})\mathbf{\Psi}(\mathbf{w}'\mathbf{X}_2 - \bar{\mathbf{x}}_{2,N})' \\
&= \mathbf{w}'\mathbf{\Phi}\mathbf{w} - 2\boldsymbol{\mu}_{x_0}\mathbf{c}_1 + \boldsymbol{\alpha}'\mathbf{\Phi}\boldsymbol{\alpha} \\
&\quad + (\mathbf{w}'\mathbf{X}_2 - \bar{\mathbf{x}}_{2,N})\mathbf{\Psi}(\mathbf{w}'\mathbf{X}_2 - \bar{\mathbf{x}}_{2,N})' \\
&= \mathbf{w}'\mathbf{\Phi}\mathbf{w} + (\mathbf{w}'\mathbf{X}_2 - \bar{\mathbf{x}}_{2,N})\mathbf{\Psi}(\mathbf{w}'\mathbf{X}_2 - \bar{\mathbf{x}}_{2,N})' + C_1 \ , \qquad (48)
\end{aligned}
$$

where , $C_1 = \boldsymbol{\alpha}'\mathbf{\Phi}\boldsymbol{\alpha} - 2\bar{\mathbf{x}}_{0,N}\mathbf{c}_1$, is a constant that is independent of $\mathbf{w}$. The conditional expectation of the error of a linear predictor $\mathbf{w}'\mathbf{y}$ under the model is

$$
\mathrm{E}\{(\mathbf{w}'\mathbf{y} - \bar{\mathbf{x}}_{0,N}\boldsymbol{\beta}_0 - \bar{\mathbf{x}}_{2,N}\boldsymbol{\beta}_2)|\mathbf{X}\} = \mathbf{w}'\mathbf{X}_0\boldsymbol{\beta}_0 - \bar{\mathbf{x}}_{0,N}\boldsymbol{\beta}_0 \ .
$$

Thus, the sufficient and necessary condition for $\mathbf{w}'\mathbf{y}$ to be unbiased for the population mean of y is

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

which is equivalent to (47). Under the constraint (47), the conditional variance of a linear estimator is

$$
\begin{aligned}
&\mathrm{V}\{(\mathbf{w}'\mathbf{y} - \bar{\mathbf{x}}_{0,N}\boldsymbol{\beta}_0 - \bar{\mathbf{x}}_{2,N}\boldsymbol{\beta}_2)|\mathbf{X}\} \\
&\quad = \mathbf{w}'\mathbf{\Phi}\mathbf{w} + (\mathbf{w}'\mathbf{X}_2 - \bar{\mathbf{x}}_{2,N})\mathbf{\Psi}(\mathbf{w}'\mathbf{X}_2 - \bar{\mathbf{x}}_{2,N})' \ .
\end{aligned}
$$

Thus, minimizing the objective function (46) subject to (47) is equivalent to minimizing the conditional variance of a linear predictor under the restriction for a linear estimator to be conditionally unbiased. ■

4. Construction of a model based design consistent regression estimator

In the previous section, we obtained the condition for the BLUP to be design consistent. In this section, we consider the problem of constructing a model based design consistent regression estimator when condition (14) or (18) is not satisfied. We call a regression model for which (14) holds a *full model*. If (14) does not hold, we call the model a *reduced model*.

We can not expect condition (14) for a full model to hold for every y in a general purpose survey because $\mathbf{\Sigma}_{ee}$ will be different for different y's. Therefore, given a reduced model, we search for a good model estimator under the model (1) in the class of design consistent estimators of the form (4). As we have seen in the previous section, the estimator of the form (4) is design consistent if the estimator of the regression coefficient is a design consistent estimator of a constant.

By (4), the requirement of design consistency is essentially a requirement that the estimated regression function pass through the design consistent estimator of the population mean vector. To force the regression through $(\bar{\mathbf{x}}_{1,\pi} \ , \ \bar{y}_\pi)$, we compute the regression of $y - \bar{y}_\pi$ on $\mathbf{x}_1 - \bar{\mathbf{x}}_{1,\pi}$. The transformed regression model for the sample can be written

$$
(\mathbf{I} - \mathbf{J}\mathbf{a}')\mathbf{y} = (\mathbf{I} - \mathbf{J}\mathbf{a}')\mathbf{X}_1\boldsymbol{\beta}_1 + (\mathbf{I} - \mathbf{J}\mathbf{a}')\mathbf{e} \ , \qquad (49)
$$

where $\mathbf{X}_1 = (\mathbf{x}'_{1,1}, \cdots, \mathbf{x}'_{1,n})'$, $(\bar{y}_\pi \ , \ \bar{\mathbf{x}}_{1,\pi}) = \mathbf{a}'(\mathbf{y} \ , \ \mathbf{X}_1)$, $\mathbf{a} = (a_1, \cdots, a_n)'$ and a_i were defined in (3). The regression estimator of the mean is expressed as

$$
\bar{y}_{reg} = \sum_{i\in A} w_i \, y_i =: \sum_{i\in A}(a_i + b_i)\, y_i \ , \qquad (50)
$$

where b_i is to be determined. If the estimator is to be location invariant we require $\sum_{i\in A} w_i = 1$ or equivalently,

$$
\sum_{i\in A} b_i = 0 \ . \qquad (51)
$$

The restriction that $\sum_{i\in A} w_i\mathbf{x}_{1,i} = \bar{\mathbf{x}}_{1,N}$ becomes

$$
\sum_{i\in A} b_i(\mathbf{x}_{1,i} - \bar{\mathbf{x}}_{1,\pi}) = \bar{\mathbf{x}}_{1,N} - \bar{\mathbf{x}}_{1,\pi} \ . \qquad (52)
$$

To simplify the expressions, let

$$
\mathbf{b} = (b_1, \cdots, b_n)' \ , \ \mathbf{z}_i = (1, \mathbf{x}_{1,i} - \bar{\mathbf{x}}_{1,\pi}) \ ,
$$

$$
\bar{\mathbf{z}}_c = (0, \bar{\mathbf{x}}_{1,N} - \bar{\mathbf{x}}_{1,\pi}) \ ,
$$

and $\mathbf{Z} = (\mathbf{z}'_1 \cdots, \mathbf{z}'_n)'$. Then the b_i that give the minimum variance of $\hat{\boldsymbol{\beta}}_1$ are obtained by minimizing the Lagrangian

$$
\mathbf{b}'\mathbf{\Sigma}_{ee}\mathbf{b} + \sum_{j=1}^{p}\lambda_j\left(\sum_{i=1}^{n} b_i z_{ji} - \bar{z}_{cj}\right) \qquad (53)
$$

with respect to $\mathbf{b}$. The solution vector is

$$
\mathbf{b}' = \bar{\mathbf{z}}_c\left(\mathbf{Z}'\mathbf{\Sigma}_{ee}^{-1}\mathbf{Z}\right)^{-1}\mathbf{Z}'\mathbf{\Sigma}_{ee}^{-1} \ . \qquad (54)
$$

Thus, the regression estimator (50) with $\mathbf{b}$ of (54) is

$$
\bar{y}_{reg,1} = (\mathbf{a} + \mathbf{b})'\mathbf{y} = \bar{y}_\pi + (\bar{\mathbf{x}}_{1,N} - \bar{\mathbf{x}}_{1,\pi})\hat{\boldsymbol{\beta}}_1 \ , \qquad (55)
$$

where $\hat{\boldsymbol{\beta}}_1$ is the second component of

$$
\left(\hat{\beta}_0, \hat{\boldsymbol{\beta}}'_1\right)' = \left(\mathbf{Z}'\mathbf{\Sigma}_{ee}^{-1}\mathbf{Z}\right)^{-1}\mathbf{Z}'\mathbf{\Sigma}_{ee}^{-1}\mathbf{y} \ . \qquad (56)
$$

In constructing the regression estimator (50), we obtained design consistency by forcing the regression line through the design consistent estimator of the population mean. We can also construct a design consistent estimator by adding a vector satisfying (14) to the $\mathbf{X}$ matrix if the original matrix $\mathbf{X}$ does not satisfy (14). This creates a full model from the original reduced model.

To describe this approach let z denote the added variable, where $\mathbf{z}' = (z_1, \cdots, z_n)'$ satisfies

$$\mathbf{z} = \mathbf{\Sigma}_{ee}\mathbf{L}_\pi \quad . \tag{57}$$

where $\mathbf{L}_\pi$ is defined in (14). We will find it convenient to work with the vector of deviations

$$\mathbf{z}_d = \mathbf{z} - \mathbf{X}(\mathbf{X}'\mathbf{\Sigma}_{ee}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{\Sigma}_{ee}^{-1}\mathbf{z} \quad , \tag{58}$$

where $\mathbf{X}$ is the original matrix of auxiliary variables with known population mean vector, $\bar{\mathbf{x}}_N$. The vector $\mathbf{z}_d$ is orthogonal in the metric $\mathbf{\Sigma}_{ee}$ to $\mathbf{X}$. Under this approach our full model for the sample is

$$\begin{aligned} \mathbf{y} =& \mathbf{Z}_1\boldsymbol{\beta}_{y,Z_1} + \mathbf{e} \quad , \\ & \mathbf{e} \sim (\mathbf{0}, \mathbf{\Sigma}_{ee}) \quad , \end{aligned} \tag{59}$$

where

$$\mathbf{Z}_1 = (\mathbf{z}_d \ , \ \mathbf{X}) \quad .$$

There are two possible situations associated with this approach. In the first, the population mean of the added variable, $\bar{z}_{d,N}$, is known. In this case, the resulting estimator

$$\bar{y}_{reg} = \bar{\mathbf{z}}_{1,N}\hat{\boldsymbol{\beta}}_{y,Z_1} \quad , \tag{60}$$

where

$$\bar{\mathbf{z}}_{1,N} = (\bar{z}_{d,N} \ , \ \bar{\mathbf{x}}_N) \quad ,$$

and

$$\hat{\boldsymbol{\beta}}_{y,Z_1} = (\mathbf{Z}_1'\mathbf{\Sigma}_{ee}^{-1}\mathbf{Z}_1)^{-1}\mathbf{Z}_1'\mathbf{\Sigma}_{ee}^{-1}\mathbf{y} \quad ,$$

is the best linear, conditionally unbiased predictor under the full model (59). If $\mathbf{\Sigma}_{ee}$ is known and if an equal probability sample is selected, then the regression estimator (60) is calculable.

If the population mean of the added variable is not known, the mean of the added variable z_d can be estimated with a design consistent estimator. A design consistent estimator of $\bar{z}_{d,N}$ is

$$\bar{z}_{d,\pi} = \left(\sum_{i\in A}\pi_i^{-1}\right)^{-1}\sum_{i\in A}\pi_i^{-1}z_{d,i} \quad . \tag{61}$$

Then a regression estimator of the population mean of y can be constructed by replacing the unknown mean of z_d with the estimated mean to obtain

$$\bar{y}_{reg,2} = (\bar{z}_{d,\pi}, \bar{\mathbf{x}}_N)\hat{\boldsymbol{\beta}}_{y,Z_1} \quad , \tag{62}$$

where $\hat{\boldsymbol{\beta}}_{y,Z_1}$ is of (60). The regression estimator (62) has the form of (55). For the estimator to be location invariant, we assume the first element of $\mathbf{x}_i$ is identically equal to one and let the matrix $\mathbf{X} = (\mathbf{J}_n, \mathbf{X}_1)$.

Theorem 5. The regression estimator of (62) can be written as

$$\bar{y}_{reg,2} = \bar{y}_\pi + (\bar{\mathbf{x}}_N - \bar{\mathbf{x}}_\pi)\hat{\boldsymbol{\beta}}_{y,X} \quad , \tag{63}$$

where

$$(\bar{y}_\pi, \bar{\mathbf{x}}_\pi) = \left(\sum_{i\in A}\pi_i^{-1}\right)^{-1}\sum_{i\in A}\pi_i^{-1}(y_i, \mathbf{x}_i) =: \sum_{i\in A}a_i(y_i, \mathbf{x}_i) \quad ,$$

$$a_i = \left(\sum_{j\in A}\pi_j^{-1}\right)^{-1}\pi_i^{-1} \quad ,$$

and

$$\hat{\boldsymbol{\beta}}_{y,X} = \left(\mathbf{X}'\mathbf{\Sigma}_{ee}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{\Sigma}_{ee}^{-1}\mathbf{y} \quad .$$

Also, the vector of weights used to define the regression estimator (63) is

$$\mathbf{w} = \mathbf{a}' + (\bar{\mathbf{x}}_N - \bar{\mathbf{x}}_\pi)\left(\mathbf{X}'\mathbf{\Sigma}_{ee}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{\Sigma}_{ee}^{-1} \quad , \tag{64}$$

where $\mathbf{a} = (a_1, \cdots, a_n)'$, is identically equal to the vector of weights defined in (55).

Proof. By construction $\mathbf{Z}_1'\mathbf{\Sigma}_{ee}^{-1}\mathbf{Z}_1$ is block diagonal with $\mathbf{z}_d'\mathbf{\Sigma}_{ee}^{-1}\mathbf{z}_d$ as one block and $\mathbf{X}'\mathbf{\Sigma}_{ee}^{-1}\mathbf{X}$ as the other block. Thus $\hat{\boldsymbol{\beta}}_{y,X_1}$ is expressed as

$$\begin{aligned} \hat{\boldsymbol{\beta}}_{y,Z_1} &= \begin{bmatrix} \hat{\beta}_{y,z_d} \\ \hat{\boldsymbol{\beta}}_{y,X} \end{bmatrix} \\ &= \begin{bmatrix} \left(\mathbf{z}_d'\mathbf{\Sigma}_{ee}^{-1}\mathbf{z}_d\right)^{-1}\mathbf{z}_d'\mathbf{\Sigma}_{ee}^{-1}\mathbf{y} \\ \left(\mathbf{X}'\mathbf{\Sigma}_{ee}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{\Sigma}_{ee}^{-1}\mathbf{y} \end{bmatrix} \\ &= \begin{bmatrix} \{\mathbf{L}_\pi'\mathbf{z}_d\}^{-1}\{\mathbf{L}_\pi'\mathbf{y} - \mathbf{L}_\pi'\mathbf{X}\left(\mathbf{X}'\mathbf{\Sigma}_{ee}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{\Sigma}_{ee}^{-1}\mathbf{y}\} \\ \left(\mathbf{X}'\mathbf{\Sigma}_{ee}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{\Sigma}_{ee}^{-1}\mathbf{y} \end{bmatrix} \quad , \end{aligned}$$

because

$$\begin{aligned} \mathbf{z}_d'\mathbf{\Sigma}_{ee}^{-1}\mathbf{z}_d &= \mathbf{L}_\pi'\mathbf{\Sigma}_{ee}\mathbf{L}_\pi - \mathbf{L}_\pi'\mathbf{X}\left(\mathbf{X}'\mathbf{\Sigma}_{ee}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{L}_\pi \\ &= \mathbf{L}_\pi'\left\{\mathbf{\Sigma}_{ee}\mathbf{L}_\pi - \mathbf{X}\left(\mathbf{X}'\mathbf{\Sigma}_{ee}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{L}_\pi\right\} \\ &= \mathbf{L}_\pi'\left\{\mathbf{z} - \mathbf{X}\left(\mathbf{X}'\mathbf{\Sigma}_{ee}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{\Sigma}_{ee}^{-1}\mathbf{z}\right\} \\ &= \mathbf{L}_\pi'\mathbf{z}_d \quad . \end{aligned}$$

The regression estimator (62) is

$$\begin{aligned} \bar{y}_{reg} &= (\bar{z}_{d,\pi}, \bar{\mathbf{x}}_N)\hat{\boldsymbol{\beta}}_{y,Z_1} \\ &= \bar{z}_{d,\pi}\hat{\boldsymbol{\beta}}_{y,z_d} + \bar{\mathbf{x}}_N\hat{\boldsymbol{\beta}}_{y,X} \\ &= \left(\sum_{i\in A}\pi_i^{-1}\right)^{-1}(\mathbf{L}_\pi'\mathbf{z}_d)(\mathbf{L}_\pi'\mathbf{z}_d)^{-1} \\ &\quad \times \left(\sum_{i\in A}\pi_i^{-1}\right)\left\{\bar{y}_\pi - \bar{\mathbf{x}}_\pi\hat{\boldsymbol{\beta}}_{y,X}\right\} + \bar{\mathbf{x}}_N\hat{\boldsymbol{\beta}}_{y,X} \\ &= \bar{y}_\pi + (\bar{\mathbf{x}}_N - \bar{\mathbf{x}}_\pi)\hat{\boldsymbol{\beta}}_{y,X} \quad . \end{aligned}$$

The matrix $\mathbf{Z}$ that was used to define the vector $\mathbf{b}$ on (55) is expressed

$$\begin{aligned}\mathbf{Z} &= \begin{pmatrix}\mathbf{J}_n & \mathbf{X}_1 - \mathbf{J}_n\bar{\mathbf{x}}_{1,\pi}\end{pmatrix} \\ &= \begin{pmatrix}\mathbf{J}_n & \mathbf{X}_1\end{pmatrix}\begin{pmatrix}1 & -\bar{\mathbf{x}}_{1,\pi} \\ \mathbf{0} & \mathbf{I}\end{pmatrix} \\ &=: \begin{pmatrix}\mathbf{J}_n & \mathbf{X}_1\end{pmatrix}\mathbf{T}\ ,\end{aligned}$$

where

$$\mathbf{T} = \begin{pmatrix}1 & -\bar{\mathbf{x}}_{1,\pi} \\ \mathbf{0} & \mathbf{I}\end{pmatrix}\ .$$

By using the inverse of partitioned matrix, the vector $\mathbf{b}$ in (55) is

$$\begin{aligned}\mathbf{b} &= \begin{pmatrix}0 & \bar{\mathbf{x}}_{1,N} - \bar{\mathbf{x}}_{1,\pi}\end{pmatrix} \\ &\times \left[\mathbf{T}'\begin{pmatrix}\mathbf{J}'_n \\ \mathbf{X}'_1\end{pmatrix}\boldsymbol{\Sigma}_{ee}^{-1}\begin{pmatrix}\mathbf{J}_n & \mathbf{X}_1\end{pmatrix}\mathbf{T}\right]^{-1}\mathbf{T}'\begin{pmatrix}\mathbf{J}'_n \\ \mathbf{X}'_1\end{pmatrix}\boldsymbol{\Sigma}_{ee}^{-1} \\ &= \begin{pmatrix}0 & \bar{\mathbf{x}}_{1,N} - \bar{\mathbf{x}}_{1,\pi}\end{pmatrix}\mathbf{T}^{-1}\left(\mathbf{X}'\boldsymbol{\Sigma}_{ee}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}'\boldsymbol{\Sigma}_{ee}^{-1} \\ &= \begin{pmatrix}0 & \bar{\mathbf{x}}_{1,N} - \bar{\mathbf{x}}_{1,\pi}\end{pmatrix}\begin{pmatrix}1 & \bar{\mathbf{x}}_{1,\pi} \\ \mathbf{0} & \mathbf{I}\end{pmatrix} \\ &\quad \times \left(\mathbf{X}'\boldsymbol{\Sigma}_{ee}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}'\boldsymbol{\Sigma}_{ee}^{-1} \\ &= (\bar{\mathbf{x}}_N - \bar{\mathbf{x}}_\pi)\left(\mathbf{X}'\boldsymbol{\Sigma}_{ee}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}'\boldsymbol{\Sigma}_{ee}^{-1}\ .\end{aligned}$$

The result follows from (55) and (64). ■

Thus the regression estimator of the finite population mean based on the full model, but with the mean of $\mathbf{z}$ unknown and estimated, is the regression estimator with $\boldsymbol{\beta}_{y,x}$ estimated by the generalized least squares regression of y on $\mathbf{x}$ using the covariance matrix $\boldsymbol{\Sigma}_{ee}$. The estimator is conditionally model unbiased under the reduced model containing only $\mathbf{x}$ if the reduced model is true. If the coefficient for $\mathbf{z}_d$ is not zero, the reduced model is not true. Then the estimator is conditionally model biased, but the estimator is unconditionally unbiased for the finite population mean because

$$\begin{aligned}&E\left\{E\left[\bar{y}_\pi + (\bar{\mathbf{x}}_N - \bar{\mathbf{x}}_\pi)\hat{\boldsymbol{\beta}}_{y,X}\right]\right\} \\ =&E\left\{\bar{\mathbf{x}}_\pi\boldsymbol{\beta}_{y,X} + \bar{z}_{d,\pi}\boldsymbol{\beta}_{y,z_d} + (\bar{\mathbf{x}}_N - \bar{\mathbf{x}}_\pi)\boldsymbol{\beta}_{y,X} \mid \mathcal{F}\right\} \\ \doteq&\bar{z}_{d,N}\boldsymbol{\beta}_{y,z_d} + \bar{\mathbf{x}}_N\boldsymbol{\beta}_{y,X}\end{aligned} \tag{65}$$

where the approximation is due to the use of the ratio estimator $\bar{z}_{d,\pi}$ defined on (61).

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

$$\hat{V}\left\{\hat{\boldsymbol{\beta}}_{y,Z_1}\middle|\mathbf{Z}_1\right\} = \left(\mathbf{Z}'_1\hat{\boldsymbol{\Sigma}}_{ee}^{-1}\mathbf{Z}_1\right)^{-1}$$

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

5. Bibliography

Brewer, K.R.W. (1963). Ratio estimation and finite population: some results deductible 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.

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: W.G. Madow and I. Olkin (eds.), *Incomplete Data in Sample Surveys*, Vol. 3. Academic Press, New York, pp. 143–160.

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. (1977). *Sampling Techniques*. 3rd ed. Wiley, New York.

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.

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

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

Hansen, M.H., Madow, W.G. and Tepping, B.J. (1983). An evaluation of model-dependent and probability-sampling inferences in sample surveys. *Journal of the American Statistical Association*. **78**, 776–793.

Isaki, C.T. (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.

Jessen, R.J. (1969). Statistical investigation of a sample survey for obtaining farm facts. *Iowa Agriculture Experiment Station Research Bulletin*. 304.

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

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

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.

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

Rao, J. N. K. (2002). *Small area estimation : Theory and Methods*. Wiley, New York.

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

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. (1992). Robust and optimal design under prediction models for finite populations. *Survey Methodology*. **18**, 179–185.

Royall, R.M. and Cumberland, W.G. (1981a). An empirical study of the ratio estimator and estimators of its variance. *Journal of the American Statistical Association*. **76**, 66–77.

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

Royall, R.M. and Herson, J. (1973a). Robust estimation in finite populations I. *Journal of the American Statistical Association*. **68**, 880–889.

Royall, R.M. and Herson, J. (1973b). Robust estimation in finite populations II: Stratification on a size variable. *Journal of the American Statistical Association*. **68**, 890–893.

Särndal, C.E. (1980). On π inverse weighting versus best linear unbiased weighting in probability sampling. *Biometrika*. **67**, 639–650.

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

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

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

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

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

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

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.