

**DESIGN-BASED RANDOM PERMUTATION MODELS
WITH AUXILIARY INFORMATION [¶]**

Wenjun Li [§]

Division of Preventative and Behavioral Medicine
University of Massachusetts Medical School, Worcester, MA 01655

Edward J. Stanek III

Department of Public Health
University of Massachusetts, Amherst, MA 01003

Corresponding address:

Wenjun Li, Ph.D.
Biostatistics Research Center
Division of Preventive and Behavioral Medicine
University of Massachusetts Medical School
Shaw Building S2-230
55 Lake Avenue North,
Worcester, MA 01655

[¶] The content of this article is a part of the first author's dissertation conducted at the Department of Biostatistics and Epidemiology, University of Massachusetts, Amherst, Massachusetts. This research was partially supported by a NIH grant (NIH-PHS-R01-HD36848). The authors wish to thank Drs. Julio Singer, John Buonaccorsi, and Carol Bigelow for their constructive comments.

[§] Corresponding author, email address: Wenjun.Li@UMASSMED.EDU

ABSTRACT

We develop design-based estimators of the population mean using auxiliary information under simple random without replacement sampling by extending the random permutation model of Stanek, Singer and Lencina (2004) and Stanek and Singer (2004). A key step in the development is representing the population mean, defined as a non-stochastic average of unit values, as the sum of random variables constructed from a random permutation of the population. Using methods similar to those in model-based approaches, the random variable representing the sum of non-sampled units is predicted to form a design-based estimator of the population mean.

The random permutation model is extended by the joint permutation of the response and auxiliary variables, with auxiliary information incorporated through centering the auxiliary variables on their respective means. Auxiliary values are assumed to be known for all population units. The estimators are required to be linear functions of the sample, unbiased and have minimum mean squared error. The estimators are identical to model-assisted and calibration estimators. However, the results provide the building blocks for extending the design-based random permutation model theory to include covariates in more complicated sample designs. This theory has already been extended to predict realized random effects in a mixed model using similar design-based methods applied to a random permutation model for a clustered population. The results developed here broaden the scope of the theory by directly extending it to include covariates. The concordance of results with other less systematic approaches is an added appeal of the methodology.

Key words: auxiliary variable, design-based inference, prediction, finite sampling random permutation model, simultaneous permutation

Running title: Random Permutation models with auxiliary information

1. INTRODUCTION

Estimators can be made more precise by accounting for auxiliary information such as gender, age, income and chronic disease-bearing history that are partially or completely known in a population. Methods of improving estimation with auxiliary information have been discussed in a model-based approach (Bolfarine and Zacks 1992; Valliant et al. 2000) and using model-assisted or calibration methods (Cassel et al. 1977; Cochran 1977; Särndal and Wright 1984; Deville and Särndal 1992; Särndal et al. 1992). In a model-based approach, the result is the best linear unbiased predictor (BLUP) (Ghosh and Rao 1994; Rao 1997). The model-assisted approach combines ideas from model-based and design-based approaches, resulting in generalized regression (GREG) estimators. The calibration approach produces a weighed estimator with weights that minimize the distance to benchmark weights, while “calibrated” to known population quantities on some set of auxiliary variables.

The actual sample design plays no role in inference based on the model-based approach. Additional superpopulation model assumptions beyond the sample design are required for development of GREG estimators. The calibration approach lacks an integrated theoretical framework. All of these approaches require additional assumptions, or lack an integrated theory that builds simply on the sample design.

We develop a design-based estimator of the population mean that accounts for auxiliary information. The development extends use of the random permutation model for simple random sampling (SRS), and two stage cluster sampling (Stanek, Singer and Lencina (2004), Stanek and Singer (2004)) to account for auxiliary variables. The method avoids the limitations of the superpopulation model approach, but takes advantage of some of the optimization tools used in developing predictors with superpopulation model assumptions. Beginning with a vector of responses for each subject in a finite population, a SRS of subjects is represented by a random permutation probability model that permutes subject’s response vectors. Auxiliary information is

incorporated through a simple linear transformation. *The results provide a direct design-based method to account for auxiliary information when estimating the population mean.*

This paper is organized as follows. We first present definitions and notation, and introduce the random permutation model. We next present a simultaneous random permutation model in a simple setting with a response and one auxiliary variable, and use it to derive the best linear unbiased estimator (BLUE) of the population mean. Subsequently, we extend the results to scenarios with multiple auxiliary variables. We conclude by illustrating the method with an example, and include results from a small scale simulation evaluating empirical estimators.

2. DEFINITIONS AND NOTATION

Let the population consist of N subjects labeled $s = 1, 2, \dots, N$, where we assume the labels are non-informative, serving only to identify the subjects. Associated with subject s is a non-stochastic potentially observable response vector, $(y_s \ \mathbf{x}_s')'$, where y_s denotes the response of interest, and $\mathbf{x}_s = ((x_{ks}))$ is a $p \times 1$ vector of auxiliary variables. We represent the population mean by the vector of means, $(\mu_y \ \boldsymbol{\mu}_x')'$, where $\boldsymbol{\mu}_x = ((\mu_{x_k}))$ is a $p \times 1$ vector. The population variance is defined by the $(p+1) \times (p+1)$ matrix $\frac{N-1}{N} \boldsymbol{\Sigma}$, where $\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_y^2 & \boldsymbol{\sigma}'_{yx} \\ \boldsymbol{\sigma}_{yx} & \boldsymbol{\Sigma}_x \end{pmatrix}$,

$\boldsymbol{\sigma}_{yx} = (\sigma_{yx_1} \ \sigma_{yx_2} \ \dots \ \sigma_{yx_p})'$ and $\boldsymbol{\Sigma}_x = ((\sigma_{x_k x_{k'}}))$ is a $p \times p$ matrix. Individual terms are

defined in the usual manner, such that $\mu_y = \frac{1}{N} \sum_{s=1}^N y_s$, $\mu_{x_k} = \frac{1}{N} \sum_{s=1}^N x_{ks}$, $\sigma_y^2 = \frac{1}{N-1} \sum_{s=1}^N (y_s - \mu_y)^2$,

$\sigma_{x_k}^2 = \frac{1}{N-1} \sum_{s=1}^N (x_{ks} - \mu_{x_k})^2$ and $\sigma_{yx_k} = \frac{1}{N-1} \sum_{s=1}^N (y_s - \mu_y)(x_{ks} - \mu_{x_k})$.

3. THE RANDOM PERMUTATION MODEL

We define a random permutation model as the set of all possible permutations of subjects in the population, where each permutation is equally likely. We represent a random permutation by the sequence of $i = 1, \dots, N$ random variables, referring to a random variable in the permutation by its position, Y_i . A simple random sample of size n is defined as the first $i = 1, \dots, n$ variates in a random permutation. Following Stanek, Singer and Lencina (2004), we explicitly represent these random variables in terms of a set of indicator random variables U_{is} , $i = 1, 2, \dots, N$, that have a value of one if subject s is in position i in a permutation, and zero otherwise. Using this notation, response for the subject in position i in a permutation is

represented by the random variable $Y_i = \sum_{s=1}^N U_{is} y_s$; using vectors, $Y_i = \mathbf{U}_i' \mathbf{y}$ where

$\mathbf{U}_i = (U_{i1} \ U_{i2} \ \dots \ U_{iN})'$ represents a vector of random variables that selects a subject in position i and $\mathbf{y} = ((y_s))$ represents the population response vector. Defining

$\mathbf{U}_{N \times N} = (\mathbf{U}_1 \ \mathbf{U}_2 \ \dots \ \mathbf{U}_N)'$, the random variables representing a permutation of response are given by $\mathbf{Y} = \mathbf{U} \mathbf{y}$. Similar vectors can be defined to represent a permutation of auxiliary variables, $\mathbf{X}_k = \mathbf{U} \mathbf{x}_k$, $k = 1, \dots, p$. The vector and matrix notation simplifies representation of the problem and development of the expected value and variance of the vectors.

We use properties of the indicator random variables to evaluate the expected value and variance of the random variables. Taking expectation over all possible permutations,

$$E(U_{is}) = \frac{1}{N} \text{ for all } i = 1, \dots, N; s = 1, \dots, N. \text{ As a result, } E(Y_i) = E(\mathbf{U}_i' \mathbf{y}) = \frac{1}{N} \mathbf{1}_N' \mathbf{y} = \mu_y \text{ for all}$$

$$i = 1, \dots, N, \text{ where } \mathbf{1}_N \text{ is an } N \times 1 \text{ column vector. In a similar manner, } E(U_{is} U_{i^*s^*}) = \frac{1}{N(N-1)}$$

when $i \neq i^*$ and $s \neq s^*$, $E(U_{is}^2) = \frac{1}{N}$ when $i = i^*$ and $s = s^*$, and $E(U_{is}U_{i^*s^*}) = 0$ otherwise.

Using these expressions, we can show that $\text{var}(Y_i) = \frac{N-1}{N}\sigma_y^2$ for all $i = 1, \dots, N$, and

$$\text{cov}(Y_i, Y_{i^*}) = -\frac{1}{N}\sigma_y^2 \text{ when } i \neq i^*.$$

We combine these results to express the expected value and variance of a random permutation using vector notation. As a result, $E(\mathbf{Y}) = \mu_y \mathbf{1}_N$; and $\text{var}(\mathbf{Y}) = \sigma_y^2 \mathbf{P}_{N,N}$, where $\mathbf{P}_{a,b} = \mathbf{I}_a - b^{-1} \mathbf{J}_a$, \mathbf{I}_N is an $N \times N$ identity matrix, and $\mathbf{J}_N = \mathbf{1}_N \mathbf{1}_N'$. Similar expressions apply to auxiliary variates. In addition, $\text{cov}(\mathbf{Y}, \mathbf{X}_k) = \sigma_{y,x_k} \mathbf{P}_{N,N}$ and for $k \neq k^*$,

$$\text{cov}(\mathbf{X}_k, \mathbf{X}_{k^*}) = \sigma_{x_k, x_{k^*}} \mathbf{P}_{N,N}.$$

4. SIMULTANEOUS RANDOM PERMUTATION MODEL

We next represent random variables arising from the random permutation of response and auxiliary variates simultaneously. For simplicity, we first assume that there is a single auxiliary variable (i.e. $p = 1$) corresponding to the subject's age (which we represent by x_s). Age is assumed to be known for all subjects in the population.

The simultaneous random permutation model is defined by concatenating the permutation vectors for response and the auxiliary variates. The resulting model is an example of a seemingly unrelated regression model (Zellner 1963) with both regressions corresponding to simple mean models such that

$$\begin{pmatrix} \mathbf{Y} \\ \mathbf{X} \end{pmatrix} = \begin{pmatrix} \mathbf{1}_N & \mathbf{0} \\ \mathbf{0} & \mathbf{1}_N \end{pmatrix} \begin{pmatrix} \mu_y \\ \mu_x \end{pmatrix} + \mathbf{E} \quad (1)$$

$$\mathbf{Z} = \mathbf{G}\boldsymbol{\mu} + \mathbf{E}$$

where $\mathbf{Z} = (\mathbf{Y}' \ \mathbf{X}')'$, $\mathbf{G} = \mathbf{I}_2 \otimes \mathbf{1}_N$, $\boldsymbol{\mu} = (\mu_y \ \mu_x)'$, and where \otimes denotes the Kronecker product (Graybill 1976). In this model, $E(\mathbf{Z}) = \mathbf{G}\boldsymbol{\mu}$ and $\text{var}(\mathbf{Z}) = \boldsymbol{\Sigma} \otimes \mathbf{P}_{N,N}$. The mean age, μ_x , is known since age is assumed to be known for all subjects. This implies that the N random variables arising from permuting the auxiliary variable are constrained by

$\frac{1}{N} \sum_{i=1}^N X_i = \mu_x$. We eliminate μ_x from model (1) by subtracting $\begin{pmatrix} \mathbf{0} \\ \mathbf{1}_N \mu_x \end{pmatrix}$ from both sides (which

is equivalent to multiplying each term in the model by $\mathbf{R} = \begin{pmatrix} \mathbf{I}_N & \mathbf{0} \\ \mathbf{0} & \mathbf{P}_{N,N} \end{pmatrix}$). Representing the

elements of \mathbf{X}^* by $X_i^* = X_i - \mu_x$, the transformed model is given by

$$\begin{pmatrix} \mathbf{Y} \\ \mathbf{X}^* \end{pmatrix} = \begin{pmatrix} \mathbf{1}_N \\ \mathbf{0}_N \end{pmatrix} \mu_y + \mathbf{E}^* \quad (2)$$

$$\mathbf{Z}^* = \mathbf{G}^* \mu_y + \mathbf{E}^*$$

where $\mathbf{Z}^* = \mathbf{RZ}$ and $\mathbf{G}^* = \mathbf{RG}$. Since $\mathbf{P}_{N,N}$ is idempotent, $E(\mathbf{Z}^*) = \mathbf{G}^* \mu_y$, $\text{cov}(\mathbf{Z}^*) = \boldsymbol{\Sigma} \otimes \mathbf{P}_{N,N}$.

Model (2) is equivalent to model (1) with the auxiliary variate centered at zero.

5. SAMPLING AND PARTITIONING THE RANDOM VARIABLES

We represent random variables for a SRS of size n by the first n elements of the random permutation vectors. The sampled (\mathbf{Z}_I^*) and remaining (\mathbf{Z}_{II}^*) portions of \mathbf{Z}^* can be

obtained by pre-multiplication by $\mathbf{K} = \begin{pmatrix} \mathbf{I}_2 \otimes (\mathbf{I}_n \ \mathbf{0}_{n \times (N-n)}) \\ \mathbf{I}_2 \otimes (\mathbf{0}_{(N-n) \times n} \ \mathbf{I}_{N-n}) \end{pmatrix}$, i.e., $\mathbf{KZ}^* = \begin{pmatrix} \mathbf{Z}_I^* \\ \mathbf{Z}_{II}^* \end{pmatrix}$. Since \mathbf{K} is

non-stochastic, it follows that $E(\mathbf{Z}_I^*) = (\mu_y \mathbf{1}_n' \ \mathbf{0}_n')'$, $E(\mathbf{Z}_{II}^*) = (\mu_y \mathbf{1}_{N-n}' \ \mathbf{0}_{N-n}')'$, and

$$\text{cov} \begin{pmatrix} \mathbf{Z}_I^* \\ \mathbf{Z}_{II}^* \end{pmatrix} = \begin{pmatrix} \mathbf{V}_I & \mathbf{V}_{I,II} \\ \mathbf{V}_{I,II}' & \mathbf{V}_{II} \end{pmatrix}, \text{ where } \mathbf{V}_I = \boldsymbol{\Sigma} \otimes \mathbf{P}_{n,N}, \mathbf{V}_{II} = \boldsymbol{\Sigma} \otimes \mathbf{P}_{(N-n),N},$$

$$\mathbf{V}_{I,II} = \mathbf{V}_{II,I}' = -\frac{1}{N} \boldsymbol{\Sigma} \otimes \mathbf{J}_{n \times (N-n)}, \text{ where } \mathbf{J}_{n \times (N-n)} = \mathbf{1}_n \mathbf{1}_{N-n}'.$$

Consequently, the partitioned model that reflects SRS sampling can be represented as

$$\begin{pmatrix} \mathbf{Z}_I^* \\ \mathbf{Z}_{II}^* \end{pmatrix} = \begin{pmatrix} \mathbf{G}_I^* \\ \mathbf{G}_{II}^* \end{pmatrix} \mu_y + \mathbf{E}^*, \quad (3)$$

$$\text{where } \mathbf{G}_I^* = (\mathbf{1}_n' \quad \mathbf{0}_n')' \text{ and } \mathbf{G}_{II}^* = (\mathbf{1}_{N-n}' \quad \mathbf{0}_{N-n}')'.$$

6. PARAMETER OF INTEREST

We assume that the parameter of interest is the population mean of the response variate, which is defined as a linear function of the permuted random variables, namely,

$$\mu_y = c \sum_{i=1}^n Y_i + c \sum_{i=n+1}^N Y_i, \text{ where } c = \frac{1}{N}, \text{ or equivalently as,}$$

$$\mu_y = \mathbf{C}' \mathbf{Z}^* = \mathbf{C}_I' \mathbf{Z}_I^* + \mathbf{C}_{II}' \mathbf{Z}_{II}^*, \quad (4)$$

$$\text{where } \mathbf{C} = c(\mathbf{1}_N' \quad \mathbf{0}_N')', \mathbf{C}_I = c(\mathbf{1}_n' \quad \mathbf{0}_n')' \text{ and } \mathbf{C}_{II} = c(\mathbf{1}_{N-n}' \quad \mathbf{0}_{N-n}')'.$$

After sampling, only the second term in the right hand side of (4) will be unknown; thus, estimating μ_y , is equivalent to

find a predictor of $\mathbf{C}_{II}' \mathbf{Z}_{II}^* = c \sum_{i=n+1}^N Y_i$. We develop the best linear unbiased predictor (BLUP) of

$\mathbf{C}_{II}' \mathbf{Z}_{II}^*$, and refer to the estimator of μ_y as the BLUE of μ_y .

7. BLUE OF A LINEAR FUNCTION

We require the predictor of $\mathbf{C}_{II}' \mathbf{Z}_{II}^*$ to be a linear function of the sample,

$$\mathbf{w}' \mathbf{Z}_I^* = \sum_{i=1}^n w_{yi} Y_i + \sum_{i=1}^n w_{xi} X_i^*, \text{ to be unbiased, i.e., } E(\mathbf{w}' \mathbf{Z}_I^*) = E(\mathbf{C}_{II}' \mathbf{Z}_{II}^*), \text{ and have minimum}$$

MSE. The estimator of μ_y is given by

$$P = (\mathbf{C}'_I + \mathbf{w}') \mathbf{Z}_I^* = c \sum_{i=1}^n Y_i + \left(\sum_{i=1}^n w_{yi} Y_i + \sum_{i=1}^n w_{xi} X_i^* \right). \quad (5)$$

The unbiased constraint implies that $\mathbf{w}' \mathbf{G}_I^* - \mathbf{C}'_I \mathbf{G}_I^* = 0$. The variance of P is

$$\text{var}(P) = \mathbf{w}' \mathbf{V}_I \mathbf{w} - 2 \mathbf{w}' \mathbf{V}_{I,II} \mathbf{C}_{II} + \mathbf{C}'_{II} \mathbf{V}_{II} \mathbf{C}_{II}. \quad \text{With such setup, the prediction theorem of Royall}$$

(1976) can be applied. The minimum variance unbiased estimator is obtained by minimizing the function

$$\Phi(\mathbf{w}) = \mathbf{w}' \mathbf{V}_I \mathbf{w} - 2 \mathbf{w}' \mathbf{V}_{I,II} \mathbf{C}_{II} + 2 \left\{ \mathbf{w}' \mathbf{G}_I^* - \mathbf{C}'_I \mathbf{G}_I^* \right\} \lambda, \quad (6)$$

where λ is a Lagrangian multiplier, resulting in

$$\hat{\mathbf{w}} = \mathbf{V}_I^{-1} \left\{ \mathbf{V}_{I,II} + \mathbf{G}_I^* \left(\mathbf{G}_I^{*'} \mathbf{V}_I^{-1} \mathbf{G}_I^* \right)^{-1} \left(\mathbf{G}_{II}^{*'} - \mathbf{G}_I^{*'} \mathbf{V}_I^{-1} \mathbf{V}_{I,II} \right) \right\} \mathbf{C}_{II}. \quad (7)$$

After simplification, (7) reduces to,

$$\hat{\mathbf{w}} = \frac{1}{n} \begin{pmatrix} 1-f \\ -\beta \end{pmatrix} \otimes \mathbf{1}_n, \quad (8)$$

where $f = n/N$, $\beta = \sigma_{xy} / \sigma_x^2$. Consequently, the estimator and its variance are

$$\begin{aligned} \hat{P} &= \frac{1}{N} \left\{ \sum_{i=1}^n Y_i + \sum_{i=n+1}^N \left(\bar{Y}_I - \frac{\beta}{1-f} (\bar{X}_I - \mu_x) \right) \right\} \\ &= f \bar{Y}_I + (1-f) \left[\bar{Y}_I - \frac{\beta}{1-f} (\bar{X}_I - \mu_x) \right] \\ &= \bar{Y}_I - \beta (\bar{X}_I - \mu_x) \end{aligned} \quad (9)$$

and

$$\text{var}(\hat{P}) = (1 - \rho^2) (1-f) \frac{\sigma_y^2}{n}, \quad (10)$$

where $\rho = \sigma_{xy} / \sigma_x \sigma_y$ is the correlation coefficient of Y on X , and \bar{Y}_I and \bar{X}_I are the sample mean of the response variable and the auxiliary variable, respectively. This expression is similar

to the expressions commonly seen in linear regression estimators, but includes a finite population correction factor.

The expressions in equation (9) have interesting interpretations. The first expression is divided into the sum of random variables in the sample and a predictor of the random variables in the remainder. Since the weighted total of the sample and remainder random variables equals the population mean, it is clear that for a realized sample, we simply substitute the values observed for the sample random variables in the estimator. Rather than predicting the value of each remaining random variable by the sample mean, \bar{Y}_I , we include an adjustment that accounts for the auxiliary variable. The adjustment ‘shrinks’ the observed difference between the auxiliary sample mean and the population auxiliary mean. For example, if the value observed for the auxiliary sample mean, \bar{X}_I , exceeds the population mean, μ_x , and the response and auxiliary random variables are positively correlated, we may anticipate that the sample mean, \bar{Y}_I , will also exceed the population mean μ_y . The predictor of the remaining random variables compensates for this over estimation of the sample random variables. The correction is proportional to the auxiliary difference, $\bar{X}_I - \mu_x$, and weighted by the regression coefficient, β inflated by one over the finite population correction factor $1 - f$.

The second expression in equation (9) is very similar to the BLUP of a realized cluster mean based on a mixed model fit to a two stage cluster sample. The similarity provides an intuition connecting BLUP in the two stage sampling context to adjustments for auxiliary variables in SRS. In the context of two stage cluster sampling with equal size clusters, the BLUP of a realized cluster mean under a two stage random permutation model is given by

$$\hat{T}_i = \tilde{f}\bar{Y}_i + (1 - \tilde{f}) \left[\bar{Y}_i - \frac{\tilde{\beta}}{1 - \tilde{f}} (\bar{Y}_i - \bar{Y}) \right],$$

where $\tilde{f} = m/M$ is the common sampling fraction of units in clusters, \bar{Y}_i is the sample mean of the realized cluster, \bar{Y} is the average of these means over sampled clusters, and

$\tilde{\beta} = \frac{(1-\tilde{f})\sigma_e^2}{m\sigma^2 + (1-\tilde{f})\sigma_e^2}$ is a function of the variance between clusters, σ^2 , and the average

variance within clusters, σ_e^2 (see Table 1, Stanek and Singer (2004, p1125)). The predictor is the sum of two terms. The first term substitutes the values observed for the sample units in the realized cluster in the predictor. The second term predicts the contribution of the remaining units in the realized cluster by adjusting the realized cluster sample mean. The adjustment is proportional to the difference between the realized cluster sample mean, and the overall mean.

The similarity of the BLUP of a realized cluster mean in two stage cluster sampling to the estimator of the population mean with auxiliary information in SRS is self-evident. The difference between the sample mean for the auxiliary variable and the population mean in SRS is replaced by the difference between the realized cluster sample mean and the overall mean in two-stage cluster sampling. The regression coefficient relating the response to the auxiliary variable in SRS is replaced by a coefficient that is a function of the within and between cluster variance components.

8. EXTENSION TO CASES WITH MULTIPLE COVARIATES

Extension of the above results to scenarios with multiple auxiliary variables is straightforward. Suppose p auxiliary variables are available for use in estimation. The matrix of random variables representing a joint permutation of a response variable and the p covariates is given by $\mathbf{U}(\mathbf{y} \ \mathbf{x}_1 \ \cdots \ \mathbf{x}_p) = (\mathbf{Y} \ \mathbf{X}_1 \ \cdots \ \mathbf{X}_p)$. A simultaneous random permutation model can be defined similar to (1), with $\mathbf{Z} = \text{vec}(\mathbf{Y} \ \mathbf{X}_1 \ \cdots \ \mathbf{X}_p)$ and $\mathbf{G} = (\mathbf{I}_{p+1} \otimes \mathbf{1}_N)$. It follows that $E(\mathbf{Z}) = \mathbf{G}\boldsymbol{\mu}$, and that $\text{cov}(\mathbf{Z}) = \boldsymbol{\Sigma} \otimes \mathbf{P}_{N,N}$. When auxiliary means μ_k , $k = 1, \dots, p$, are known, a

similar centering transformation can be applied to \mathbf{Z} , such that $\mathbf{Z}^* = \begin{pmatrix} \mathbf{I}_N & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_p \otimes \mathbf{P}_{N,N} \end{pmatrix} \mathbf{Z}$. A reparameterized model incorporating known auxiliary information has a form identical to Model (2), but where $\mathbf{G}^* = \begin{pmatrix} \mathbf{1}'_N & \mathbf{0}'_{Np} \end{pmatrix}'$. With these differences, a linear unbiased minimum variance estimator can be derived following the same steps in Section 7 (Appendix 1). The unique solution for the vector of coefficients \mathbf{w} is

$$\hat{\mathbf{w}} = \frac{1}{n} \begin{pmatrix} 1-f \\ -\boldsymbol{\beta} \end{pmatrix} \otimes \mathbf{1}_n,$$

where $\boldsymbol{\beta} = \boldsymbol{\Sigma}_X^{-1} \boldsymbol{\sigma}_{Xy} = (\beta_1 \ \beta_2 \ \cdots \ \beta_p)'$, $k = 1, 2, \dots, p$. Accordingly, the estimator and its variance are

$$\begin{aligned} \hat{P} &= f\bar{Y}_I + (1-f) \left[\bar{Y}_I - \frac{\boldsymbol{\beta}'}{1-f} (\bar{\mathbf{X}}_I - \boldsymbol{\mu}_x) \right] \\ &= \bar{Y}_I - \boldsymbol{\beta}' (\bar{\mathbf{X}}_I - \boldsymbol{\mu}_x) \end{aligned} \quad (11)$$

and

$$\text{var}(\hat{P}) = (1 - \rho_{yX}^2)(1-f) \frac{\sigma_y^2}{n}, \quad (12)$$

where \bar{Y}_I is the sample mean of the response variable and $\bar{\mathbf{X}}_I$ is the vector of sample means of the p auxiliary variable, and $\rho_{yX}^2 = \boldsymbol{\sigma}'_{Xy} \boldsymbol{\Sigma}_X^{-1} \boldsymbol{\sigma}_{Xy} / \sigma_y^2$ is the squared multiple correlation coefficient of Y on \mathbf{X}^* . Notice that the only difference between (12) and (10) is the multiple correlation coefficient ρ_{yX}^2 instead of ρ^2 . This expression is similar to the expressions commonly seen in multiple linear regression models (Graybill 1976), but includes a finite population correction factor. Results (11) and (12) are also the same as difference estimators with optimal coefficients (Montanari 1987) and the GREG estimator (Särndal et al. 1992).

9. EMPIRICAL ESTIMATES

The estimator in equation (9) and (11) are a function of the regression coefficients, which, in turn, are a function of variance components in the population. In practice, although Σ_x may be known, σ_{xy} will need to be estimated. Sarndal et al. (1992, p229) recommends estimating both terms. When $p=1$, the estimate of β is given by $b_1 = \hat{\sigma}_{xy} / \hat{\sigma}_{xx}$, where

$$\hat{\sigma}_{xy} = \frac{1}{n-1} \sum_{i=1}^n (Y_i - \bar{Y}_I)(X_i - \bar{X}_I) \text{ and } \hat{\sigma}_x^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_I)^2.$$

An alternative estimator uses the known population variance of X to estimate β by $b_2 = \hat{\sigma}_{xy} / \sigma_{xx}$.

We conducted a small scale simulation study to compare different empirical estimators in the context of a simple problem where there is interest in estimating the smoking rate $\mu_y = \pi_y$ in a population of size N based on a SRS with both smoking status and gender are recorded on sample subjects. We assume the response variable is an indicator of smoking status ($y_s = 1$ if subject s is a smoker, and zero otherwise), and the auxiliary variable is an indicator of male gender ($x_s = 1$ if subject s is male, and zero otherwise). We also assume that the proportion of males in the population, $\mu_x = \pi_x$, is known. By including the gender status of the sample subjects, theoretically, we can reduce the variance of the estimate of smoking prevalence in the population over the simple sample proportion, with the percent reduction given by $100(1 - \rho^2)$. In practice, since the regression parameter is not known, some of the reduction in the MSE is lost due to uncertainty in the regression coefficient.

The simulation study generated a series of hypothetical populations of sizes 50, 100, 200, 400, 800 and 1,600; each with an auxiliary variable representing 50% men ($\pi_x = 0.5$). In all populations, the overall prevalence of smokers is assumed to be 40% ($\pi_y = 0.4$). The prevalence of smoking in men ranged from 40% to 68% translating into a relative risk of smoking

for men ranging from 1 to 5.67. We evaluated the estimators by comparing the average MSE over 25,000 independent simple random samples. To compare the MSE of the estimators, we expressed the MSE of estimators that account for gender as a percentage of the MSE of the estimator corresponding to the simple sample smoking prevalence. Three estimators are compared, differing by whether b_1 , b_2 , or β is used for the regression coefficient.

The simulation results are illustrated in Figures 1 and 2 and summarize the reduction in MSE that may result by accounting for auxiliary information. Figure 1 presents results when $N=100$ and $n=25$. Results were very similar to those in Figure 1 with larger N and larger size samples. Figure 2 presents results when $N=50$ and $n=25$. In both figures, notice that although a reduction in the MSE will almost always occur when the regression coefficient is known after accounting for auxiliary information. In the simulation, the MSE of the estimator based on b_2 was slightly lower but basically equivalent to the estimator based on b_1 in all settings. Using the empirical estimators, a reduction in the MSE will not occur unless there is a sufficiently large difference (measured here by the relative risk of smoking for males) in response by the auxiliary variable. From Figure 1, a reduction in the MSE will occur when $RR>1.75$, while from Figure 2, a reduction will occur when $RR>1.8$.

10. EXAMPLE

As a simple example, suppose there is interest in estimating the smoking rate $\mu_y = \pi_y$ in a population of size N based on a SRS with both smoking status and gender recorded on sample subjects. We assume that the proportion of males in the population, $\mu_x = \pi_x$, is known, and represent the sample estimate of the proportion smoking as $\bar{Y}_I = \hat{\pi}_y$, the proportion male in the sample as $\bar{X}_I = \hat{\pi}_x$, and the proportion of male smokers in the sample as $\hat{\pi}_{xy}$. With this

notation, $\frac{N-1}{N} \hat{\sigma}_y^2 = \hat{\pi}_y (1 - \hat{\pi}_y)$, $\frac{N-1}{N} \hat{\sigma}_x^2 = \hat{\pi}_x (1 - \hat{\pi}_x)$, $\frac{N-1}{N} \sigma_x^2 = \pi_x (1 - \pi_x)$, and

$\frac{N-1}{N} \hat{\sigma}_{xy} = \hat{\pi}_{xy} - \hat{\pi}_x \hat{\pi}_y$. Using these estimators, we estimate β by $b_1 = \frac{\hat{\pi}_{xy} - \hat{\pi}_x \hat{\pi}_y}{\hat{\pi}_x (1 - \hat{\pi}_x)}$ or

$b_2 = \frac{\hat{\pi}_{xy} - \hat{\pi}_x \hat{\pi}_y}{\pi_x (1 - \pi_x)}$ and estimate ρ by $\hat{\rho} = \frac{\hat{\pi}_{xy} - \hat{\pi}_x \hat{\pi}_y}{\sqrt{\pi_x (1 - \pi_x)} \sqrt{\hat{\pi}_y (1 - \hat{\pi}_y)}}$. Substituting these

expressions into the estimator in (9) using b_2 results in

$$\begin{aligned} \hat{P}_2 &= \frac{1}{N} \left\{ n \hat{\pi}_y + \left[(N - n) \hat{\pi}_y - N b_2 (\hat{\pi}_x - \pi_x) \right] \right\} \\ &= \hat{\pi}_y - \left(\frac{\hat{\pi}_{xy} - \hat{\pi}_x \hat{\pi}_y}{\pi_x (1 - \pi_x)} \right) (\hat{\pi}_x - \pi_x) \end{aligned}$$

with variance estimated by $\text{var}(\hat{P}) = (1 - \hat{\rho}^2)(1 - f) \frac{\hat{\sigma}_y^2}{n}$.

As a particular application, suppose that $N = 100$ and $n = 20$ so that $f = 0.20$, and assume that half the population is male, i.e. $\pi_x = 0.5$. Among the sample, suppose further that thirty percent smoke ($\hat{\pi}_y = 0.3$), sixty percent are male ($\hat{\pi}_x = 0.6$), and that twenty-five percent of the sample subjects are male smokers ($\hat{\pi}_{xy} = 0.25$). Using this information,

$$b_2 = \frac{0.25 - 0.6(0.3)}{0.5(1 - 0.5)} = 0.28, \quad \hat{\rho} = \frac{0.25 - 0.6(0.3)}{\sqrt{0.5(1 - 0.5)} \sqrt{0.3(1 - 0.3)}} = 0.3055, \text{ and the proportion of}$$

smokers in the population is estimated by

$$\begin{aligned} \hat{P}_2 &= \frac{1}{100} \left\{ 6 + \left[(80) 0.3 - (100) 0.28(0.6 - 0.5) \right] \right\} \\ &= \frac{1}{100} \left\{ 6 + \left[24 - 28(0.1) \right] \right\} \\ &= \frac{1}{100} \{ 6 + 21.2 \} \\ &= 0.272 \end{aligned}$$

Thus, rather than using the sample estimate of the smoking prevalence (i.e., $\hat{\pi}_y = 0.3$), the estimate of smoking prevalence accounting for gender is given by $\hat{P} = 0.272$. To form this estimate, the predicted number of smokers among the non-sampled subjects is reduced from 24 to 21.2 to account for the larger percentage of male subjects in the sample (relative to the percentage of males in the population). The same estimate can be obtained by directly applying the regression results, $\hat{P}_2 = \hat{\pi}_y - b_2(\hat{\pi}_x - \pi_x)$ or $\hat{P}_2 = 0.3 - 0.28(0.6 - 0.5) = 0.272$, but this expression does not reveal the essential prediction of non-sampled random variables underlying the method development. Using b_1 to estimate β , $\hat{P}_1 = 0.271$, a result that corresponds to the simple weighted average prevalence of smoking, using population gender weights. These estimators are identical to “post-stratified” estimators of prevalence rate, as discussed in calibration (Deville and Särndal 1992) and model-assisted approaches (Särndal et al. 1992). The estimated standard error is given by $\sqrt{\text{var}(\hat{P}_2)} = 0.3922$. The standard error is $\sqrt{1 - \hat{\rho}^2} = 0.9522$ of the estimated standard error based solely on the sample (given by

$$\sqrt{\text{var}(\hat{\pi}_y)} = \sqrt{(1-f) \frac{N}{N-1} \frac{\hat{\pi}_y(1-\hat{\pi}_y)}{n}}.$$

11. DISCUSSION

The survey sampling literature has struggled to reconcile design-based and model-based theories of estimation/prediction. Model based methods recently popularized by Valiant, Dorfman and Royall (2000), have a theoretical structure based on the prediction-based methods developed by Royall (1973,1976). This structure is important, since it allows methods to be extended relatively easily to different applications with increasing complexity. The limitation of the theory is that it does not account for the sample design.

A similar theory has not previously emerged using design-based methods. Instead, a mixture of approaches, such as GREG or calibration approaches (Sarndal, et al. 1992) have

been developed. These approaches combine model-based and design-based ideas, or begin with ad-hoc functional forms of estimators, and optimize them in special settings. These approaches have been successful in addressing many practical problems in a design-based framework. However, they have not led to an approach that provides the consistent conceptual and theoretical base, or that can be readily extended to applications with increasing complexity.

We have illustrated how the design-based random permutation model theory can be extended to include auxiliary variables in a straight forward manner. These results extend the scope of the random permutation model theory to a broader class of problems. Previous developments of the theory have identified subtleties in interpreting random effects in simple random sampling (Stanek, Singer, and Lencina (2004)) and developed predictors of realized random effects in balanced two stage sampling problems with response error (Stanek and Singer, 2004). Current research is extending these results to clustered population settings where clusters are of different size, and there is unequal probability sampling, and to settings where there is missing data. In each case, the same basic approach is considered, with estimators (or predictors) developed based on a clear optimization theory.

The present results illustrate how the theory can be used to account for covariates. The fact that the results coincide with results developed by GREG or calibration approaches strengthens the appeal of the random permutation model approach as a design-based competitor to the model-based superpopulation theory. Still, much more work is needed to extend the methods. Extensions are being developed to more complex settings, including two stage designs with cluster and unit covariates, longitudinal studies, and settings where randomization of units to treatments. We consider the basic results developed here to provide a foundation for additional work in these directions.

Reference

Bolfarine, H. and S. Zacks (1992). Prediction Theory for Finite Populations. New York, Springer-Verlag.

- Brewer, K. R. W. (1995). Combining design-based and model-based inference (Chapter 30). Business Survey Methods. B. B. Cox, DA; Chinnappa, BN; Christianson, A; Colledge, MJ; Kott, PS. New York, John Wiley & Sons, Inc.: 589-606.
- Brewer, K. R. W. (1999). "Cosmetic calibration with unequal probability sampling." Survey Methodology **25**: 205-212.
- Brewer, K. R. W. (1999). "Design-based or prediction-based inference? Stratified random vs stratified balanced sampling." International Statistical Review **67**(1): 35-47.
- Brewer, K. R. W., M. Hanif, et al. (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., C.-E. Särndal, et al. (1977). Foundations of Inference in Survey Sampling. New York, NY, Wiley.
- Cassel, C. M., C. E. Särndal, et al. (1976). "Some Results on Generalized Difference Estimation and Generalized Regression Estimation for Finite Populations." Biometrika **63**: 615-620.
- Cochran, W. G. (1977). Sampling Techniques. New York, John Wiley and Sons.
- Deville, J. C. and C. E. Särndal (1992). "Calibration Estimators in Survey Sampling." Journal of the American Statistical Association **87**: 376-382.
- Ghosh, M. and J. N. K. Rao (1994). "Small area estimation: an appraisal." Statistical Science **9**(1): 55-76.
- Graybill, F. A. (1976). Theory and application of the linear model. Belmont, CA, Wadsworth Publishing Company, Inc.
- Horvitz, D. G. and D. J. Thompson (1952). "A generalization of sampling without replacement from a finite universe." Journal of the American Statistical Association **47**(260): 663-685.
- Montanari, G. E. (1987). "Post-sampling Efficient QR-prediction in Large-sample Surveys." International Statistical Review **55**: 191-202.
- Rao, C. R. (1967). Least square theory using an estimated dispersion matrix and its applications to measurement of signals. Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability, Berkeley, CA.
- Rao, J. N. K. (1997). "Developments in sample survey theory: an appraisal." Canadian Journal of Statistics **25**(1): 1-21.
- Rao, J. N. K. (1999). "Some Current Trends in Sample Survey Theory and Methods." Sankhya: The Indian Journal of Statistics: Series B **61**(1): 1-57.
- Royall, R.M. (1973). "The prediction approach to finite population sampling theory: application to the hospital discharge survey." Rockville, Md. DHEW No. (HSM) 73-1329:1-31.
- Royall, R. M. (1976). "The linear least-squares prediction approach to two-stage sampling." Journal of the American Statistical Association **71**: 657-664.
- Särndal, C. E., B. Swensson, et al. (1992). Model assisted survey sampling. New York, Springer-Verlag.
- Särndal, C. E. and R. L. Wright (1984). "Cosmetic Form of Estimators in Survey Sampling." Scandinavian Journal of Statistics **11**: 146-156.
- Stanek, E. J., J. d. M. Singer, et al. (2004). "A unified approach to estimation and prediction under simple random sampling." Journal of Statistical Planning and Inference **121**: 325-338.
- Stanek III, E. J. and J. M. Singer (2004). "Predicting Random Effects from Finite population clustered samples with response error." Journal of the American Statistical Association **99**(468): 1119-1130.
- Valliant, R., A. H. Dorfman, et al. (2000). Finite Population Sampling and Inference, a Prediction Approach. New York, John Wiley & Sons.
- Wu, C. B. and R. R. Sitter (2001). "A model-calibration approach to using complete auxiliary information from survey data." Journal of the American Statistical Association **96**(453): 185-193.

Zellner, A. (1963). "Estimators of Seemingly Unrelated Regression Equations: Some Exact Finite Sample Results." Journal of the American Statistical Association **58**: 977-992.

Figure 1. Percent Change in MSE of Smoking Prevalence Estimates Accounting for Gender by Relative Risk of Smoking (N=100, n=25)

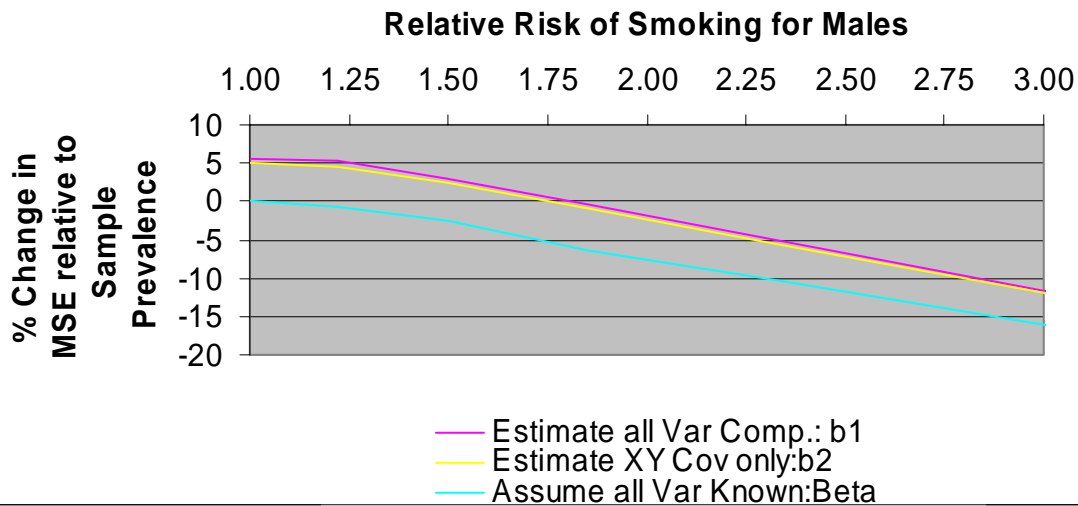
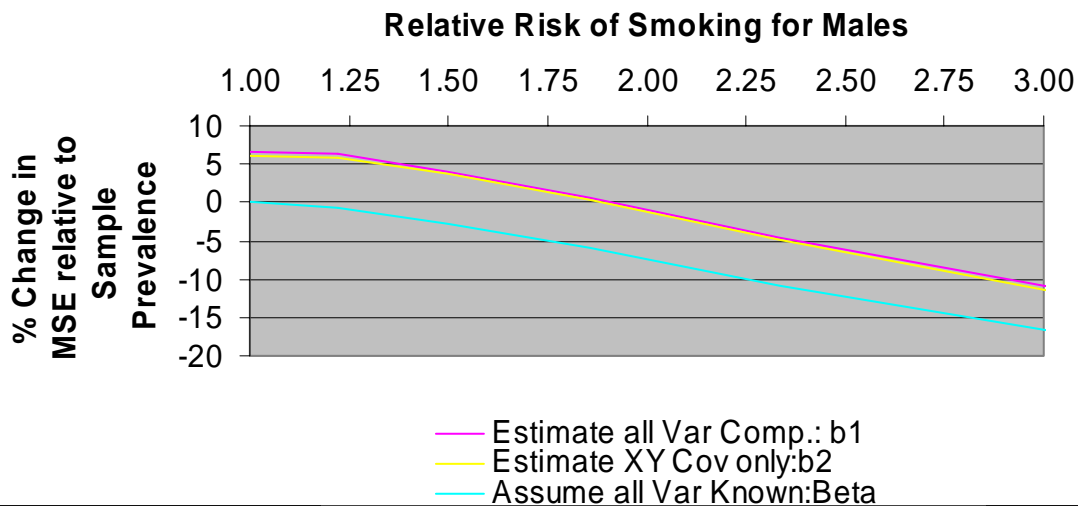


Figure 2. Percent Change in MSE of Smoking Prevalence Estimates Accounting for Gender by Relative Risk of Smoking (N=50, n=25)



Appendix 1 Proof of result (11) and (12)

When there are p auxiliary variables, a simultaneous random permutation model can be defined similar to (1) with $\mathbf{Z} = \text{vec}(\mathbf{Y} \ \mathbf{X}_1 \ \cdots \ \mathbf{X}_p)$ and $\mathbf{G} = (\mathbf{I}_{p+1} \otimes \mathbf{1}_N)$. When $\boldsymbol{\mu}_X$ is known,

let $\mathbf{X}^* = \mathbf{P}_N \mathbf{X}$, where $\mathbf{X} = (\mathbf{X}_1 \ \mathbf{X}_2 \ \cdots \ \mathbf{X}_p)$. The transformation can be represented as

$$\mathbf{Z}^* = \mathbf{RZ}, \text{ where } \mathbf{R} = \begin{pmatrix} \mathbf{I}_N & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_p \otimes \mathbf{P}_N \end{pmatrix}, \text{ and } \mathbf{RG} = \begin{pmatrix} \mathbf{G}^* & \mathbf{0}_{N(p+1)} \end{pmatrix} \text{ where } \mathbf{G}^* = \begin{pmatrix} \mathbf{1}'_N & \mathbf{0}'_{Np} \end{pmatrix}', \text{ and}$$

$\mathbf{E}^* = \mathbf{RE}$. The reparameterized model that incorporates the constraints thus has a form identical

to Model (2). To represent the SRS sampling, let $\mathbf{K} = \begin{pmatrix} \mathbf{I}_{p+1} \otimes (\mathbf{I}_n \ \mathbf{0}_{n \times (N-n)}) \\ \mathbf{I}_{p+1} \otimes (\mathbf{0}_{(N-n) \times n} \ \mathbf{I}_{N-n}) \end{pmatrix}$, and thus the

portioned simultaneous permutation model takes the same form of (3),

$$\begin{pmatrix} \mathbf{Z}_I^* \\ \mathbf{Z}_{II}^* \end{pmatrix} = \begin{pmatrix} \mathbf{G}_I^* \\ \mathbf{G}_{II}^* \end{pmatrix} \mu_y + \mathbf{E}^*,$$

where $\mathbf{G}_I^* = \begin{pmatrix} \mathbf{1}'_n & \mathbf{0}'_{np} \end{pmatrix}'$ and $\mathbf{G}_{II}^* = \begin{pmatrix} \mathbf{1}'_{N-n} & \mathbf{0}'_{(N-n)p} \end{pmatrix}'$.

A linear function of the random variables can be defined similar to (4), with

$\mathbf{C} = (\mathbf{c}' \ \mathbf{0}'_{Np \times 1})'$, $\mathbf{C}_I = (\mathbf{c}'_I \ \mathbf{0}'_{np \times 1})'$ and $\mathbf{C}_{II} = (\mathbf{c}'_{II} \ \mathbf{0}'_{(N-n)p \times 1})'$. The estimator of $\mathbf{C}'_{II} \mathbf{Z}_{II}^*$ can then

be defined as a linear function of the sample, $\mathbf{w}' \mathbf{Z}_I^*$, where $\mathbf{w} = (\mathbf{w}'_y \ \mathbf{w}'_x)'$,

$\mathbf{w}_x = (\mathbf{w}'_1 \ \mathbf{w}'_2 \ \cdots \ \mathbf{w}'_p)'$, is a $(p+1)n \times 1$ vector of coefficients. With these, a linear unbiased

minimum variance estimation can be derived by minimizing the function

$$\Phi(\mathbf{w}) = \mathbf{w}' \mathbf{V}_I \mathbf{w} - 2 \mathbf{w}' \mathbf{V}_{I,II} \mathbf{C}_{II} + 2 \{ \mathbf{w}' \mathbf{G}_I^* - \mathbf{C}'_{II} \mathbf{G}_{II}^* \} \lambda.$$

Specifically, differentiating the above function with respect to \mathbf{w} and λ , and setting the derivatives to zeros results in the following estimating equations

$$\begin{pmatrix} \mathbf{V}_I & \mathbf{G}_I^* \\ \mathbf{G}_I^{*'} & \mathbf{0} \end{pmatrix} \begin{pmatrix} \hat{\mathbf{w}} \\ \hat{\lambda} \end{pmatrix} = \begin{pmatrix} \mathbf{V}_{I,II} \mathbf{C}_{II} \\ \mathbf{G}_{II}^{*'} \mathbf{C}_{II} \end{pmatrix},$$

where $\mathbf{G}_I^* = (\mathbf{1}'_n \quad \mathbf{0}'_{np})'$ and $\mathbf{G}_{II}^* = (\mathbf{1}'_{N-n} \quad \mathbf{0}'_{(N-n)p})'$. The unique solution is

$$\hat{\mathbf{w}} = \mathbf{V}_I^{-1} \left\{ \mathbf{V}_{I,II} + \mathbf{G}_I^* (\mathbf{G}_I^{*'} \mathbf{V}_I^{-1} \mathbf{G}_I^*)^{-1} (\mathbf{G}_{II}^{*'} - \mathbf{G}_I^{*'} \mathbf{V}_I^{-1} \mathbf{V}_{I,II}) \right\} \mathbf{C}_{II}. \quad (\text{A.1})$$

Since $\mathbf{V}_I^{-1} = \boldsymbol{\Sigma}^{-1} \otimes \mathbf{P}_{n,N}^{-1}$, $\mathbf{V}_{I,II} = \boldsymbol{\Sigma} \otimes \left(-\frac{1}{N} \mathbf{J}_{n \times (N-n)} \right)$, $\mathbf{P}_{n,N}^{-1} \mathbf{1}_n = \frac{N}{N-n} \mathbf{1}_n$ and

$\mathbf{1}'_n \mathbf{P}_{n,N}^{-1} \mathbf{1}_n = \frac{Nn}{N-n}$, we have the following useful identities, $(\mathbf{G}_I^{*'} \mathbf{V}_I^{-1} \mathbf{G}_I^*)^{-1} = \frac{N-n}{Nn} (\mathbf{u}'_1 \boldsymbol{\Sigma}^{-1} \mathbf{u}_1)^{-1}$ with

$\mathbf{u}_1 = (1 \quad \mathbf{0}_{1 \times p})$ and $\mathbf{V}_I^{-1} \mathbf{V}_{I,II} = -\frac{1}{N-n} \mathbf{I}_{p+1} \otimes \mathbf{J}_{n \times (N-n)}$. Subsequently, (A.1) can be simplified as

follows,

$$\begin{aligned} \hat{\mathbf{w}} &= \left\{ -\frac{1}{N-n} \mathbf{I}_{p+1} \otimes \mathbf{J}_{n \times (N-n)} + \frac{1}{n} (\mathbf{u}'_1 \boldsymbol{\Sigma}^{-1} \mathbf{u}_1)^{-1} \left[(\boldsymbol{\Sigma}^{-1} \mathbf{u}_1 \mathbf{u}'_1) \otimes (\mathbf{P}_{n,N}^{-1} \mathbf{1}_n \mathbf{1}'_{N-n}) \right] \right\} \mathbf{C}_{II} \\ &= \left(\frac{\mathbf{1}'_{N-n} \mathbf{c}_{II}}{N-n} \right) \left\{ -\mathbf{u}_1 \otimes \mathbf{1}_n + \frac{N}{n} \left((\mathbf{u}'_1 \boldsymbol{\Sigma}^{-1} \mathbf{u}_1)^{-1} \boldsymbol{\Sigma}^{-1} \mathbf{u}_1 \right) \otimes \mathbf{1}_n \right\}. \end{aligned}$$

Since $\boldsymbol{\Sigma}^{-1} = \begin{pmatrix} \sigma_y^2 & \sigma'_{xy} \\ \sigma_{xy} & \boldsymbol{\Sigma}_x \end{pmatrix}^{-1} = \begin{pmatrix} (\sigma_y^2 - \sigma'_{xy} \boldsymbol{\Sigma}_x^{-1} \sigma_{xy})^{-1} & -\sigma_y^{-2} \sigma'_{xy} (\boldsymbol{\Sigma}_x - \sigma_{xy} \sigma_y^{-2} \sigma'_{xy})^{-1} \\ -\boldsymbol{\Sigma}_x^{-1} \sigma_{xy} (\sigma_y^2 - \sigma'_{xy} \boldsymbol{\Sigma}_x^{-1} \sigma_{xy})^{-1} & (\boldsymbol{\Sigma}_x - \sigma_{xy} \sigma_y^{-2} \sigma'_{xy})^{-1} \end{pmatrix}$,

$$(\mathbf{u}'_1 \boldsymbol{\Sigma}^{-1} \mathbf{u}_1)^{-1} (\boldsymbol{\Sigma}^{-1} \mathbf{u}_1) = (\sigma_y^2 - \sigma'_{xy} \boldsymbol{\Sigma}_x^{-1} \sigma_{xy}) \left(\frac{(\sigma_y^2 - \sigma'_{xy} \boldsymbol{\Sigma}_x^{-1} \sigma_{xy})^{-1}}{-\boldsymbol{\Sigma}_x^{-1} \sigma_{xy} (\sigma_y^2 - \sigma'_{xy} \boldsymbol{\Sigma}_x^{-1} \sigma_{xy})^{-1}} \right) = \left(\frac{1}{-\boldsymbol{\Sigma}_x^{-1} \sigma_{xy}} \right) = \left(\frac{1}{-\boldsymbol{\beta}} \right)$$

where $\boldsymbol{\beta} = \boldsymbol{\Sigma}_x^{-1} \sigma_{xy}$. Therefore, we have

$$\hat{\mathbf{w}} = \left(\frac{\mathbf{1}'_{N-n} \mathbf{c}_{II}}{N-n} \right) \left\{ -\mathbf{u}_1 \otimes \mathbf{1}_n + \frac{N}{n} \left[\left(\frac{1}{-\boldsymbol{\beta}} \right) \otimes \mathbf{1}_n \right] \right\}.$$

Consequently,

$$\begin{aligned}
 \hat{P} &= (\mathbf{C}_I + \mathbf{w})' \mathbf{Z}_I^* \\
 &= \mathbf{c}'_I \mathbf{Y}_I + \mathbf{c}'_{II} \mathbf{1}_{N-n} \left\{ \bar{Y}_I - \frac{N}{N-n} \boldsymbol{\beta}' (\bar{\mathbf{X}}_I - \boldsymbol{\mu}_x) \right\} \\
 &= f \bar{Y}_I + (1-f) \left\{ \bar{Y}_I - \frac{1}{1-f} \boldsymbol{\beta}' (\bar{\mathbf{X}}_I - \boldsymbol{\mu}_x) \right\} \\
 &= \bar{Y}_I - \boldsymbol{\beta}' (\bar{\mathbf{X}}_I - \boldsymbol{\mu}_x)
 \end{aligned}$$

and

$$\begin{aligned}
 \text{var}(\hat{P}) &= (\mathbf{C}_I + \mathbf{w})' \mathbf{V}_I (\mathbf{C}_I + \mathbf{w}) \\
 &= (1 - \rho_{yX}^2) \left(\frac{1-f}{n} \sigma_y^2 \right),
 \end{aligned}$$

where $\rho_{yX}^2 = \boldsymbol{\sigma}'_{Yy} \boldsymbol{\Sigma}_X^{-1} \boldsymbol{\sigma}_{Xy} / \sigma_0^2$ is the multiple correlation coefficient of \mathbf{Y} on \mathbf{X} .