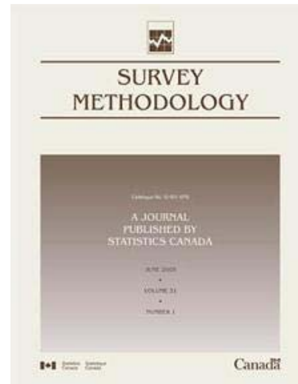
Component of Statistics Canada
Catalogue no. 12-001-X Business Survey Methods Division

Article

Local unconditional best linear unbiased estimators: Applications to survey sampling

by Juliet Popper Shaffer



June 2000

Statistics Canada Statistique Canada

Canada

Survey Methodology, June 2000
Vol. 26, No. 1, pp. 113-116
Statistics Canada, Catalogue No. 12-001

Local unconditional best linear unbiased estimators: Applications to survey sampling

Juliet Popper Shaffer [1]

Abstract

Survey statisticians frequently use superpopulation linear regression models. The Gauss-Markov theorem, assuming fixed regressors or conditioning on observed values of regressors, asserts that the standard estimators of regression coefficients are best linear unbiased. Shaffer (1991) showed that the Gauss-Markov theorem doesn't apply when the regressors are random if some aspects of the population distribution of the regressors are known, and introduced an alternative estimator with better properties than the standard estimator under some conditions. This paper derives some generalizations, and notes an optimality property (locally best linear unbiasedness) of the generalized alternative estimator. Implications for estimation in surveys are noted.

Key Words: Regression analysis; Gauss-Markov theorem; Survey sampling; Unbiased estimation; Optimality; Best linear unbiased estimation.

1. Introduction

In the standard linear regression model for a sample of observations,

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \in, \tag{1}$$

the matrix of regressors, $\boldsymbol{X}$, is assumed to be a known, fixed matrix. Shaffer (1991) showed that when $\boldsymbol{X}$ is assumed to be random, the Gauss-Markov theorem does not hold in general, and described an alternative estimator that is more accurate when $\boldsymbol{\beta}$ is close to zero. Shaffer gave two applications of her results, to estimates of $\boldsymbol{\beta}$ and associated population quantities in multivariate normal superpopulation models and to ratio estimation of population means and totals.

In the present paper, three generalizations of these results are derived.

(a) The results are generalized from a model in which the sample covariance matrix of the errors $\in$ is $\sigma^2\boldsymbol{I}$, where $\boldsymbol{I}$ is the $n \times n$ identity matrix, to the case in which the covariance matrix Σ of $\in$ is $\sigma^2\boldsymbol{B}$, where $\boldsymbol{B}$ is a known, fixed positive-definite matrix, and to some situations in which $\boldsymbol{B}$ is random (since it is the covariance matrix of a randomly-selected sample of regressor values).

(b) A generalized estimator is derived that performs well when the coefficient vector $\boldsymbol{\beta}$ is close to any pre-specified coefficient vector $\boldsymbol{\beta}_0$.

(c) A condition is given for design-unbiasedness of estimators of population means and totals based on the generalized estimator of $\boldsymbol{\beta}$.

Some results under the general model (1) will be given first. Then, modifications that apply to the sample survey situation will be discussed.

Under Model (1) with $\Sigma = \sigma^2\boldsymbol{I}$, the Gauss-Markov theorem asserts that the sample estimator

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y}, \tag{2}$$

is a best linear unbiased estimator (BLUE) if $\boldsymbol{X}$ is regarded as a fixed matrix. If the rows of $\boldsymbol{X}$ are treated as realizations of random vectors x_i, $i = 1, ..., n$, the Gauss-Markov theorem can be interpreted as an assertion that the estimator in (2) has minimum variance in the class of estimators linear in $\boldsymbol{Y}$ and conditionally unbiased, given these realized values of $\boldsymbol{X}$. However, the use of the term "unbiased" without qualification generally means unconditional unbiasedness. If the requirement of unbiasedness is interpreted to mean unbiased unconditionally, *i.e.*, on the average over random vectors with values in $\boldsymbol{X}$, Shaffer (1991) showed that the Gauss-Markov theorem doesn't apply when $\mathrm{E}(\boldsymbol{X}'\boldsymbol{X})$ is known. In that case, the conditionally biased estimator

$$\hat{\boldsymbol{\beta}}^* = [\mathrm{E}(\boldsymbol{X}'\boldsymbol{X})]^{-1}(\boldsymbol{X}'\boldsymbol{Y}) \tag{3}$$

is unconditionally unbiased and has smaller variance than $\hat{\boldsymbol{\beta}}$ when $\boldsymbol{\beta}$ is small. In fact, when $\mathrm{E}(\boldsymbol{X}'\boldsymbol{X})$ is known, no BLUE exists.

Comparison of the variances of (2) and (3) under various modeling assumptions, aside from the implications for estimating the coefficients themselves, gives insight into the conditions under which various estimators of other parameters of the populations have desirable properties, both model-based and design-based.

1. Juliet Popper Shaffer, University of California, Department of Statistics, 367 Evans Hall, #3860, Berkeley, CA 94720-3860, U.S.A. E-mail: shaffer@berkeley.edu.

2. Generalization of the covariance matrix of $\in$

If the covariance matrix of $\in$ is of the form $\sigma^2 \boldsymbol{B}$, where $\boldsymbol{B}$ is a known, fixed positive-definite matrix, the Gauss-Markov theorem applies to the generalized estimator

$$\hat{\beta} = [\boldsymbol{X'}\boldsymbol{B}^{-1}\boldsymbol{X}]^{-1}\boldsymbol{X'}\boldsymbol{B}^{-1}\boldsymbol{Y}. \qquad (4)$$

The proofs in Shaffer (1991) generalize directly to show that, if $\mathrm{E}(\boldsymbol{X'}\boldsymbol{B}^{-1}\boldsymbol{X})$ is known, the estimator

$$\hat{\beta}^* = [\mathrm{E}(\boldsymbol{X'}\boldsymbol{B}^{-1}\boldsymbol{X})]^{-1}\boldsymbol{X'}\boldsymbol{B}^{-1}\boldsymbol{Y} \qquad (5)$$

has smaller variance than (4) when β is sufficiently close to zero. The (unconditional) variances of (4) and (5) are

$$\textstyle\sum_{\hat{\beta}} = \mathrm{E}\,[(\boldsymbol{X'}\boldsymbol{B}^{-1}\boldsymbol{X})^{-1}]\,\sigma^2 \qquad (6)$$

and

$$\textstyle\sum_{\hat{\beta}} = [\mathrm{E}(\boldsymbol{X'}\boldsymbol{B}^{-1}\boldsymbol{X})]^{-1}\,\sigma^2 + \mathrm{Var.}\{[\mathrm{E}(\boldsymbol{X'}\boldsymbol{B}^{-1}\boldsymbol{X})]^{-1}(\boldsymbol{X'}\boldsymbol{B}^{-1}\boldsymbol{X})\beta\}. \qquad (7)$$

When $\beta = 0$, Shaffer shows that (7) is smaller than (6), and therefore, assuming continuity of (7) as a function of β, it is smaller than (6) when β is in a neighborhood of zero.

The results will now be applied in the sample survey context. Let $\boldsymbol{X}_N$ refer to the $N \times p$ matrix, and $\boldsymbol{Y}_N$ to the $N \times 1$ vector, in a finite population. If the $\boldsymbol{N}$ population elements are considered to be a sample from an infinite hypothetical population of potential elements satisfying (1), and if a sample of size n of the finite population is taken, the proofs in Shaffer (1991) generalize directly to show that

$$\hat{\beta}_N^* = [\mathrm{E}(\boldsymbol{X}_N'\,\boldsymbol{B}_N^{-1}\,\boldsymbol{X}_N)]^{-1}\,\boldsymbol{X}_N'\,\boldsymbol{B}_N^{-1}\,\boldsymbol{Y}_N \qquad (8)$$

and

$$\hat{\beta}_n^* = [\mathrm{E}(\boldsymbol{X}_n'\,\boldsymbol{B}_n^{-1}\,\boldsymbol{X}_n)]^{-1}\,\boldsymbol{X}_n'\,\boldsymbol{B}_n^{-1}\,\boldsymbol{Y}_n \qquad (9)$$

have variances smaller than those of their corresponding conditional versions $\hat{\beta}_N$ and $\hat{\beta}_n$ respectively, if β is close to zero, where the expectation in (8) is over the infinite population of hypothetical elements, and the expectation in (9) is over either the same infinite population or over the finite population of N elements satisfying (1). In order to apply these results, the expectations in (8) and (9) have to be known.

If $\boldsymbol{X}_N$ is to be regarded as fixed, the population model can be written as

$$\boldsymbol{Y}_N = \boldsymbol{X}_N\beta + \in_N, \qquad (10)$$

where $\in_N$ is a vector of randomly distributed error terms as in (1). Under Model (10), $\hat{\beta}_N$ and $\hat{\beta}_N^*$ are identical, but $\hat{\beta}_n$ is still distinct from $\hat{\beta}_n^*$. Under Model (10), for a random sample of size n, if

$$\mathrm{E}[(\boldsymbol{X}_n'\,\boldsymbol{B}_n^{-1}\,\boldsymbol{X}_n)/\boldsymbol{n}] = (\boldsymbol{X}_N'\,\boldsymbol{B}_N^{-1}\,\boldsymbol{X}_N)/\boldsymbol{N}, \qquad (11)$$

the alternative estimator can be written in the form

$$\hat{\beta}_n^* = [(\boldsymbol{n}/\boldsymbol{N})(\boldsymbol{X}_N'\,\boldsymbol{B}_N^{-1}\,\boldsymbol{X}_N)]^{-1}\,\boldsymbol{X}_n'\,\boldsymbol{B}_n^{-1}\,\boldsymbol{Y}_n. \qquad (12)$$

In model (10), Equations (11) and (12) will apply if $\boldsymbol{B}_N$ is diagonal and the sampling plan is self-weighting, and under some other conditions and sampling plans, *e.g.*, if $\boldsymbol{B}_N$ is block (cluster) diagonal and complete clusters are sampled. If $\boldsymbol{B}_N$ is diagonal, $\boldsymbol{B}_n$ is not necessarily fixed. For example, suppose a population consists of both men and women, and the variances of the two sexes on the characteristic of interest are known and are different. In that case, if a self-weighting sample is taken, and Model (10) is assumed to hold in both subpopulations, $\boldsymbol{B}_n$ will be diagonal, with entries that are a function of the proportions of the two genders in the sample.

3. Locally best linear unbiased estimation

Under the model (1), the estimator (5) is the locally best linear unbiased estimator (LBLUE) when $\beta = 0$; *i.e.*, the estimator, linear in $\boldsymbol{Y}$ and unbiased for β with smallest variance in a neighborhood of $\beta = 0$. Furthermore, the generalized linear estimator

$$\hat{\beta}^*_{(\beta_0)} = \beta_0 + [\mathrm{E}(\boldsymbol{X'}\boldsymbol{B}^{-1}\boldsymbol{X})]^{-1}[\boldsymbol{X'}\boldsymbol{B}^{-1}(\boldsymbol{Y} - \boldsymbol{X}\beta_0)], \qquad (13)$$

allowing for the addition of a constant, is the LBLUE at $\beta = \beta_0$, for an arbitrary vector β_0. The proof of these results in given in Appendix A. This generalized estimator (13) could be useful in a survey sampling situation in which it was reasonably sure that β would be close to some specified value. The variance of (13) is easily shown to equal (7) with $(\beta - \beta_0)$ substituted for β. (See Appendix A.) Under Model (10) estimators (8), (9), and (12) generalize to

$$\hat{\beta}^*_{(\beta_{0,N})} = \beta_0 + [\mathrm{E}(\boldsymbol{X}_N'\,\boldsymbol{B}_N^{-1}\,\boldsymbol{X}_N)]^{-1}[\boldsymbol{X}_N'\,\boldsymbol{B}_N^{-1}(\boldsymbol{Y}_N - \boldsymbol{X}_N\beta_0)], \qquad (14)$$

$$\hat{\beta}^*_{(\beta_{0,n})} = \beta_0 + [\mathrm{E}(\boldsymbol{X}_n'\,\boldsymbol{B}_n^{-1}\,\boldsymbol{X}_n)]^{-1}[\boldsymbol{X}_n'\,\boldsymbol{B}_n^{-1}(\boldsymbol{Y}_n - \boldsymbol{X}_n\beta_0)], \qquad (15)$$

and

$$\hat{\beta}^*_{(\beta_{0,n})} = \beta_0 + [(n/N)\,\boldsymbol{X}_N'\,\boldsymbol{B}_N^{-1}\,\boldsymbol{X}_N)]^{-1}[\boldsymbol{X}_n'\,\boldsymbol{B}_n^{-1}(\boldsymbol{Y}_n - \boldsymbol{X}_n\beta_0)], \qquad (16)$$

respectively.

4. Conditions for design unbiasedness

Assume the Model (10) holds, and that the unconditionally unbiased estimator can be expressed in the form (16). Suppose there exists a $p \times 1$ –vector $\boldsymbol{g}$ such that $\boldsymbol{B}_N^{-1} \boldsymbol{X}_N \boldsymbol{g} = \mathbf{1}_N$ and, for every sample of size n, $\boldsymbol{B}_n^{-1} \boldsymbol{X}_n \boldsymbol{g} = \mathbf{1}_n$, where $\mathbf{1}_N$ and $\mathbf{1}_n$ are vectors of ones of length N and n, respectively. Then, given a simple random sample,

(a) the estimator

$$\hat{\boldsymbol{Y}}_{(\boldsymbol{\beta}_{0,n})} = \bar{\boldsymbol{X}}'_N \, \hat{\boldsymbol{\beta}}^{*}_{(\boldsymbol{\beta}_{0,n})} \qquad (17)$$

is a design-unbiased estimator of $\bar{\boldsymbol{Y}}_N$, where $\bar{\boldsymbol{X}}'_N = (1/N)\, \mathbf{1}'_N \boldsymbol{X}_N$, and

(b) $\hat{\boldsymbol{Y}}_{(\boldsymbol{\beta}_{0,n})}$ is a generalized difference estimator of $\bar{\boldsymbol{Y}}_N$.

The proof is given in Appendix B.

Note that a vector $\boldsymbol{g}$ satisfying the conditions of this theorem exists if the model includes an intercept (*i.e.*, X_N includes a column of ones) or if $\boldsymbol{B}_N$ is diagonal and the variance is proportional to the values of one of the regressors. Many applications of regression modeling to sample survey estimation are based on models that incorporate these assumptions. Särndal, Swensson and Wretman (1991, pages 231 and 232) discuss these and more general models, and Chapter, 6, section 4 of that reference has examples of commonly applied models incorporating these assumptions. Chapter 6 as a whole discusses both the general difference estimator of $\boldsymbol{N}\bar{\boldsymbol{Y}}_N$ and the analogous general regression estimator based on $\hat{\boldsymbol{\beta}}_n$. The material in that Chapter also suggests generalizations of these results to more complex estimators and sampling plans.

5. Discussion

To apply the results to estimates of properties of a finite population, it will be assumed that the matrix B is diagonal or has the special block-diagonal form and associated sampling plan discussed above. From the results in section 3, it follows that the estimator (17) of $\bar{\boldsymbol{Y}}_N$ has smaller variance than the estimator

$$\hat{\hat{\boldsymbol{Y}}}_{(\hat{\boldsymbol{\beta}}_n)} = \bar{\boldsymbol{X}}_N \, \hat{\boldsymbol{\beta}}_n \qquad (18)$$

when $\boldsymbol{\beta}$ is close to $\boldsymbol{\beta}_0$. Note that (18) can be written

$$\bar{\boldsymbol{Y}}_{\hat{\boldsymbol{\beta}}_n} = \frac{1}{\mathbf{N}} \left[\sum_{i \in s} \boldsymbol{X}'_i \hat{\boldsymbol{\beta}}_n + \sum_{i \notin s} \boldsymbol{X}'_i \hat{\boldsymbol{\beta}}_n \right], \qquad (19)$$

and $\boldsymbol{X}'_i$ is the i^{th} row of $\boldsymbol{X}$, and $\boldsymbol{S}$ is the set of elements in the sample. Royall (1970) showed that the best linear model-unbiased estimator of $\bar{\boldsymbol{Y}}_N$ (unbiased conditionally on the obtained sample) is

$$\frac{1}{\boldsymbol{N}} \left[\sum_{i \in s} \boldsymbol{Y}_i + \sum_{i \notin s} \boldsymbol{X}'_i \hat{\boldsymbol{\beta}}_n \right]. \qquad (20)$$

In some important cases, the first term in (20) is equal to the first term in (19), in which case (20) and (19) are identical. This will be true, for example, if $\boldsymbol{B} = \sigma^2 \boldsymbol{I}$ and the model (10) contains an intercept, or if $p = 1$ and $\boldsymbol{B}$ is diagonal with diagonal entries proportional to the values of the single regressor. In such cases, (20) and (19) are identical, and the design-unbiased and unconditionally-model-unbiased estimator (17) has a smaller expected squared discrepancy from $\bar{\boldsymbol{Y}}_N$ than the best linear conditionally-model-unbiased estimator (20) when $\boldsymbol{\beta}$ is close to $\boldsymbol{\beta}_0$. Furthermore, if the sampling fraction is negligible, (17) has smaller expected squared discrepancy than (20) when $\boldsymbol{\beta}$ is close to $\boldsymbol{\beta}_0$, even without the requirement that the first terms of (20) and (19) be equal.

If $\hat{\boldsymbol{\beta}}$ is replaced by $\hat{\boldsymbol{\beta}}^{*}$ in (20), the resulting estimator is no longer unconditionally unbiased. It can be shown, however, using concepts of dependence (Lehmann 1966) that under the conditions on $\boldsymbol{B}$ noted at the beginning of this section, the resulting estimator will have smaller expected squared discrepancy from $\bar{\boldsymbol{Y}}_N$ than (20) and (19) even without the further restrictions noted in the previous paragraph.

6. Conclusion

Since the conditions under which the estimator (5) of $\boldsymbol{\beta}$ is more efficient than the estimator (4) are very restrictive, and the estimators of population characteristics based on (5) can be derived in other ways, the results given here may be of more theoretical than practical interest. The results do give additional insight into some situations in which simple estimators like the sample mean and the generalized difference estimator are more efficient in estimating the population mean than are ratio estimators, poststratified estimators, regression estimators and other complex estimators. The equations (6) and (7) for comparative variances of (4) and (5) provide an alternative method of comparing respective variances under different regression models and different values of $\boldsymbol{\beta}$. Many of these results hold under very simple sampling plans, but it should be possible to generalize them to more complex, unequal probability sampling plans.

Acknowledgements

The author is grateful to the late Erik N. Torgersen, who suggested the generalized estimator with optimal properties, to Phillip S. Kott, whose suggestion led to the derivation of the design unbiasedness condition, and to anonymous referees for many valuable comments.

Appendix A

Proof that $\hat{\beta}^*_{\beta_0}$ is LBLUE at β_0

Assume model (1), with $\mathrm{Var}(\boldsymbol{Y}|\boldsymbol{X}) = \sigma^2 \boldsymbol{B}$. (The general proof given here applies directly to the model (10) as well.) Consider the sample estimator

$$\hat{\beta}^*_{(\beta_0)} = \beta_0 + [\mathrm{E}(\boldsymbol{X}'\boldsymbol{B}^{-1}\boldsymbol{X}]^{-1}\boldsymbol{X}'\boldsymbol{B}^{-1}(\boldsymbol{Y} - \boldsymbol{X}\beta_0).$$

Let $\tau = \beta - \beta_0$ and $\boldsymbol{Z} = \boldsymbol{Y} - \boldsymbol{X}\beta_0$. Then $\mathrm{E}(\boldsymbol{Z}|\boldsymbol{X}) = \boldsymbol{X}\tau$, $\mathrm{Var}(\boldsymbol{Z}|\boldsymbol{X}) = \sigma^2\boldsymbol{B}$, and $\hat{\tau}^* = [\mathrm{E}(\boldsymbol{X}'\boldsymbol{B}^{-1}\boldsymbol{X})]^{-1}\boldsymbol{X}'\boldsymbol{B}^{-1}\boldsymbol{Z} = \hat{\beta}^*_{(\beta_0)} - \beta_0$.

Thus, the properties of $\hat{\beta}^*_{(\beta_0)}$ at $\beta = \beta_0$ are the same as those of $\hat{\beta}^* = \hat{\beta}^*_{(0)}$ at $\beta = 0$, so without loss of generality it will be shown that $\hat{\beta}^*_{(0)}$ is LBLUE at $\beta = 0$. Also without loss of generality, it will be assumed that $\boldsymbol{B} = \boldsymbol{I}$.

Let $\boldsymbol{C}'(\boldsymbol{X})\boldsymbol{Y}$ be an arbitrary unconditionally-unbiased estimator of β, where $\boldsymbol{C}(\boldsymbol{X})$ is a matrix of functions of $\boldsymbol{X}$, of the same dimensions as $\boldsymbol{X}$. The requirement of unconditional unbiasedness necessitates the restriction $\mathrm{E}[\boldsymbol{C}'(\boldsymbol{X})\boldsymbol{X}] = \boldsymbol{I}$ (Shaffer 1991). Conditioning first on $\boldsymbol{X}$ and then using the expression for unconditional variance, the variance of $\boldsymbol{C}'(\boldsymbol{X})\boldsymbol{Y}$ is $\mathrm{E}[\boldsymbol{C}'(\boldsymbol{X})\,\boldsymbol{C}(\boldsymbol{X})]\,\sigma^2 + \mathrm{Var}(\boldsymbol{C}'(\boldsymbol{X})\,\boldsymbol{X}\beta)$. Since we are considering variance at $\beta = 0$, only the first term is nonzero. Letting $\boldsymbol{C}'(\boldsymbol{X}) = [\mathrm{E}(\boldsymbol{X}'\boldsymbol{X})]^{-1}\boldsymbol{X}'$, the variance of $\hat{\beta}^*$ is $[\mathrm{E}(\boldsymbol{X}'\boldsymbol{X})]^{-1}\sigma^2$.

Let $\tilde{\beta}$ be an arbitrary unconditionally-unbiased estimator of the form $\boldsymbol{C}'(\boldsymbol{X})\boldsymbol{Y}$. Then $\mathrm{Var}(\tilde{\beta}) = \mathrm{Var}(\hat{\beta}^*) + \mathrm{Var}(\tilde{\beta} - \hat{\beta}^*) + 2\mathrm{Cov}(\hat{\beta}^*, \tilde{\beta} - \hat{\beta}^*)$, so $\mathrm{Var}(\hat{\beta}^*) \le \mathrm{Var}(\tilde{\beta})$ if $\mathrm{Cov}(\hat{\beta}^*, \tilde{\beta} - \hat{\beta}^*) \ge 0$, or if $\mathrm{Cov}(\hat{\beta}^*, \tilde{\beta}) \ge \mathrm{Var}(\hat{\beta}^*)$. An easy calculation, using the restriction $\mathrm{E}[\boldsymbol{C}'(\boldsymbol{X})\,\boldsymbol{X}] = \boldsymbol{I}$, shows that $\mathrm{Cov}(\hat{\beta}^*, \tilde{\beta}) = \mathrm{Var}(\hat{\beta}^*)$, which proves that $\hat{\beta}^*_{(\beta_0)}$ is LBLUE at β_0.

Appendix B

Proof of the result in section 4

$$\begin{aligned}\bar{\boldsymbol{X}}'_N\hat{\beta}^*_{(\beta_0)} &= \bar{\boldsymbol{X}}'_N\beta_0 \\ &\quad + (\mathbf{1}/\boldsymbol{N})\mathbf{1}'_N\boldsymbol{X}_N\,[\boldsymbol{n}(\mathbf{1}/\boldsymbol{N})(\boldsymbol{X}'_N\boldsymbol{B}_N^{-1}\boldsymbol{X}_N)^{-1}] \\ &\quad \boldsymbol{X}'_n\boldsymbol{B}_n^{-1}(\boldsymbol{Y}_n - \boldsymbol{X}_n\beta_0) \\ &= \bar{\boldsymbol{X}}'_N\beta_0 + (\mathbf{1}/\boldsymbol{n})\,\boldsymbol{g}'\boldsymbol{X}'_N\boldsymbol{B}_N^{-1}\boldsymbol{X}_N\,(\boldsymbol{X}'_N\boldsymbol{B}_N^{-1}\boldsymbol{X}_N)^{-1} \\ &\quad \boldsymbol{X}'_n\boldsymbol{B}_n^{-1}(\boldsymbol{Y}_n - \boldsymbol{X}_n\beta_0) \\ &= \bar{\boldsymbol{X}}'_N\beta_0 + (\mathbf{1}/\boldsymbol{n})\mathbf{1}'_n(\boldsymbol{Y}_n - \boldsymbol{X}_n\beta_0) \\ &= \bar{\boldsymbol{X}}'_N\beta_0 + \bar{\boldsymbol{Y}}_n\bar{\boldsymbol{X}}'_N\beta_0. \end{aligned} \tag{B.1}$$

where $\boldsymbol{B}_N$ and $\boldsymbol{B}_n$ are the appropriate population and sample matrices, respectively. The final expression in (B.1) is the generalized difference estimator based on a value β_0 chosen independently of the sample. This proves part (b) of the result; since the difference estimator is unbiased for $\bar{\boldsymbol{Y}}$ in a self-weighting sample, the result in (a) follows.

References

Lehmann, E.L. (1966). Some concepts of dependence. *Annals of Mathematical Statistics*, 37, 1137-1153.

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

Särndal, C.-E., Swensson, B. and Wretman, J. (1991). *Model Assisted Survey Sampling*. New York: Springer.

Shaffer, J.P. (1991). The Gauss-Markov theorem and random regressors. *The American Statistician*, 45, 269-273.