Research Article **Open Access**

Special Issue: Proceedings of the 24th International Workshop on Matrices and Statistics

Stephen Haslett

Positive semidefiniteness of estimated covariance matrices in linear models for sample survey data

DOI 10.1515/spma-2016-0020
Received July 16, 2015; accepted April 9, 2016

Abstract: Descriptive analysis of sample survey data estimates means, totals and their variances in a design framework. When analysis is extended to linear models, the standard design-based method for regression parameters includes inverse selection probabilities as weights, ignoring the joint selection probabilities. When joint selection probabilities are included to improve estimation, and the error covariance is not a diagonal matrix, the unbiased sample based estimator of the covariance is the Hadamard product of the population covariance, the elementwise inverse of selection probabilities and joint selection probabilities, and a sample selection matrix of rank equal to the sample size. This Hadamard product is however not always positive definite, which has implications for best linear unbiased estimation. Conditions under which a change in covariance structure leaves BLUEs and/or BLUPs are known. Interestingly, this class of "equivalent" matrices for linear models includes non-positive semi-definite matrices. The paper uses these results to explore how the estimated covariance from the sample can be modified so that it meets necessary conditions to be positive semidefinite, while retaining the property that fitting a linear model to the sampled data yields the same BLUEs and/or BLUPs as when the original Hadamard product is used.

1 Introduction

Traditional or design-based sample survey design theory, as developed in [8] or [4], uses the probabilities of selection to form estimators, for example of means and totals, with estimation of the variance of these estimators being based on selection and joint selection probabilities. Many such estimators for a wide range of survey designs, including expansion estimators for stratified, clustered, and probability proportional designs, can be subsumed under the heading of Horvitz-Thompson estimators [11].

The alternative conceptualisation of sample surveys is model-based. See, for example [15], and consequent papers by the same authors.

During the 1980's and early 1990's there was much debate about which method was best. The model based approach was of considerable interest to academics doing research on sample design and analysis theory, and the design based approach was entrenched among practitioners and statisticians at official statistics agencies. Official Statisticians argued, quite correctly, that the model-based estimators were not robust to model failure. This led in time to a rapprochement, with development of model assisted sampling with the publication of [16]. The model assisted approach had the major advantage that it made explicit the underlying models that practitioners use when deciding on the "best" design for a particular survey, because design unbiasedness alone (which is simple to ensure at least in theory) is insufficient to ensure a good estimator.

Stephen Haslett: Massey University, New Zealand
and The Australian National University, Australia
E-mai: s.j.haslett@massey.ac.nz, stephen.haslett@anu.edu.au

Interestingly, and this is why designing a good sampling scheme remains somewhat an art, no uniformly best (minimum variance) sample design exists [1].

The model-assisted approach had been foreshadowed, for example in [5], and the joint design and model based approach was considered in [9] and used in [7].

To be explicit, suppose for a given population that the population mean of the variable y is:

$$\bar{Y} = \frac{1}{N}\sum_{i=1}^{N} Y_i$$

Let $\chi_i = 1$ if $i \in s$ and $\chi_i = 0$ if $i \notin s$ where s is the sample, so that $\chi_i = 1$ for n of the Npopulation units. In the design-based context, $Y_i = y_i$ for i =1,2,...,N and it is only the $\{\chi_i : i = 1, 2, \ldots, N\}$ which are random variables.

Then the only design-unbiased estimators of the mean are of the form:

$$\bar{y} = \frac{1}{N}\sum_{i=1}^{N} \chi_i y_i/\pi_i$$

with $E(\chi_i) = \pi_i$ for i =1,2,...,N where E denotes design expectation and π_i is the selection probability for the ith unit; this is the Horvitz-Thompson estimator.

In the model-based approach, the population is now itself considered to be a sample from a superpopulation with model-expectation $\mathbf{E}$ and model-variance $\mathbf{V}$, so that each of the $\{Y_i : i = 1, 2, \ldots, N\}$ is random with respect to the superpopulation, and so is $\bar{Y}$.

Design-based, model-based and model-assisted estimators, and their variances and estimated variances can be derived in this framework. For further details, see [7] or [9].

The importance of this broad framework is that it allows time series, as well as linear and generalized linear models to be fitted to sample survey data in a way that provides better and more accurate estimates both of the parameters and of the variance of those estimates.

2 Linear Models

2.1 Design based estimation for linear models

One of the standard ways of fitting linear models, particularly regression models, to survey data is to use inverse selection probabilities as weights.

The linear model for the survey data can be specified as:

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

where $\mathbf{Y}$ is an n×1 vector of responses, $\mathbf{X}$ is an n×p matrix of auxiliary variables, β a p×1 vector of parameters, and $\mathbf{e}$ an n×1 vector of errors with variance of $\mathbf{V}(\mathbf{e}) = \mathbf{E}(\mathbf{e}\mathbf{e}^T) = \mathbf{V_e}$. In general, neither $\mathbf{X}$ nor $\mathbf{V_e}$ need be of full rank.

For sample survey data, the data $\mathbf{Y} = (y_{1,} y_2, \ldots, y_i, \ldots y_n)^T$have selection probabilities $\boldsymbol{\Pi}_0 = diag(\pi_i)$ where i =1,2,... n with a parallel definition for the $N - n$ non-sampled elements.

Then the standard solution for the case where $\mathbf{X}$ is full rank is

$$\tilde{\beta} = (\mathbf{X}^T\boldsymbol{\Pi}_0^{-1}\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol{\Pi}_0^{-1}\mathbf{Y}, \tag{2}$$

with estimated variance given by

$$\tilde{\mathbf{V}}(\tilde{\boldsymbol{\beta}}) = [(\mathbf{X}^T\boldsymbol{\Pi}_0^{-1}\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol{\Pi}_0^{-1}]\,\mathbf{V_e}\,[\boldsymbol{\Pi}_0^{-1}\mathbf{X}(\mathbf{X}^T\boldsymbol{\Pi}_0^{-1}\mathbf{X})^{-1}]. \tag{3}$$

See, for example, [3, 12, 17]. This design-based solution (2) is essentially weighted least squares (with weights equal to the inverse selection probabilities) but nevertheless has parallels with ordinary least squares in that the covariance $\mathbf{V_e}$ is not accounted for. However, $\mathbf{V_e}$ is accounted for in the estimated variance, which generally differs from the appropriately scaled version of $(\mathbf{X}^T\boldsymbol{\Pi}_0^{-1}\mathbf{X})^{-1}$ that would be the estimated variance of $\tilde{\boldsymbol{\beta}}$ under simple random sampling. There is an additional minor difference from ordinary (or even weighted) least squares: $\mathbf{X}^T\boldsymbol{\Pi}_0^{-1}\mathbf{X}$ is an design unbiased estimator of $\mathbf{X}_P^T\mathbf{X}_P$, and $\mathbf{X}^T\boldsymbol{\Pi}_0^{-1}\mathbf{Y}$ is a design unbiased estimator of $\mathbf{X}_P^T\mathbf{Y}_P$, where $\mathbf{X}_P$ and $\mathbf{Y}_P$ are the finite population analogues of $\mathbf{X}$ and $\mathbf{Y}$, but since the ratio of expectations is not the expectation of the ratio, $\tilde{\boldsymbol{\beta}}$ is not quite a design unbiased estimator of $\boldsymbol{\beta}$.

3 Simultaneous adjustment of estimates for selection probabilities, joint selection probabilities and covariance structure for design-based estimates from linear models

Consider a population as a sample from a superpopulation and the linear model based on that population

$$\mathbf{Y}_P = \mathbf{X}_P\boldsymbol{\beta}+\mathbf{e}_P, \tag{4}$$

with $\mathbf{E}(\mathbf{e}_P) = 0$, where $\mathbf{E}$ denotes expectation with respect to the superpopulation, and $\mathbf{V}(\mathbf{e}_P) = \mathbf{E}(\mathbf{e}_P\mathbf{e}_P^T) = \mathbf{V}_{\mathbf{e}_P}$ is the $N{\times}N$ covariance matrix with ijth element v_{ij}. Then, because the sampling scheme is non-informative [2] so that the selection probabilities do not depend on (i.e., are independent of) the errors $\mathbf{e}_P$ in (4), standard results apply and the best linear unbiased estimate (BLUE) of the superpopulation parameter $\boldsymbol{\beta}$ in the full rank case is

$$\hat{\boldsymbol{\beta}}_P = (\boldsymbol{X}_P^T\mathbf{V}_{e_P}^{-1}\boldsymbol{X}_P)^{-1}\boldsymbol{X}_P^T\mathbf{V}_{e_P}^{-1}\boldsymbol{Y}_P. \tag{5}$$

3.1 Design unbiased estimation of the population covariance matrix

Suppose now we have a sample s, selected from the population P. Using a probability based sampling scheme with selection probabilities $\{\pi_i : i = 1, 2, \dots, N\}$ and joint selection probabilities $\{\pi_{ij} : i = 1, 2, \dots, N; j = 1, 2, \dots, N\}$ (and noting that $\pi_i = \pi_{ii}$ for $i = 1, 2, \dots, N$), define χ_P to be the $N{\times}N$ matrix with ijth element χ_{ij} equal to one if $i \in s$ and $j \in s$ and zero otherwise. Note that $\boldsymbol{\chi}_P$ depends on the sample s that is drawn, has n non-zero diagonal elements all equal to one, and $n(n-1)$ off-diagonal elements equal to one, with all other elements equal to zero. Also provided the design is noninformative, $\mathrm{E}(\boldsymbol{\chi}_P) = \boldsymbol{\Pi}_P$ where E is expectation with respect to the design, and $\boldsymbol{\Pi}_P$ has ijth element π_{ij} for $i = 1, 2, \dots, N; j = 1, 2, \dots, N$.

Define the matrix $\boldsymbol{\Pi}_P^{\odot-}$ to have ijth element $1/\pi_{ij}$ for $i = 1, 2, \dots, N; j = 1, 2, \dots, N$.

Note that for $i = 1, 2, \dots, N; j = 1, 2, \dots, N$, we have

$$\mathrm{E}(\chi_{ij}v_{ij}\pi_{ij}^{-1}) = v_{ij}\pi_{ij}^{-1}\mathrm{E}(\chi_{ij}) = v_{ij}\pi_{ij}^{-1}\pi_{ij} = v_{ij},$$

which in matrix form is

$$\mathrm{E}(\boldsymbol{\chi}_P \odot \mathbf{V}_{\mathbf{e}_P} \odot \boldsymbol{\Pi}_P^{\odot-}) = \mathbf{V}_{\mathbf{e}_P}, \tag{6}$$

where $\odot$ denotes the Hadamard (or elementwise) product, where $\boldsymbol{\chi}_P$ and $\boldsymbol{\Pi}_P^{\odot-}$ are non-negative matrices (i.e., all entries positive or zero), and $\mathbf{V}_{\mathbf{e}_P}$ is positive semidefinite.

It is the matrix $\boldsymbol{\chi}_P \odot V_{\mathbf{e}_P} \odot \boldsymbol{\Pi}_P^{\odot-}$ which is the principal focus of the remainder of this paper.

Note that, after suitable permutation, there is only an $n{\times}n$ submatrix of $\boldsymbol{\chi}_P$ that is non-zero, so that $\boldsymbol{\chi}_P \odot V_{\mathbf{e}_P} \odot \boldsymbol{\Pi}_P^{\odot-}$ has ijth element v_{ij}/π_{ij} if $i \in s$ and $j \in s$ and is zero otherwise, with the convention that the diagonal elements are v_{ii}/π_i if $i \in s$ and zero otherwise, and so (after the same permutation) $\boldsymbol{\chi}_P \odot V_{\mathbf{e}_P} \odot \boldsymbol{\Pi}_P^{\odot-}$ also has only an $n{\times}n$ submatrix that is non-zero.

Note too that if $\mathbf{V}_{\mathbf{e}_P} = \sigma_{\mathbf{e}_P}^2\mathbf{I}$, where $\mathbf{I}$ is the identity matrix and $\sigma_{\mathbf{e}_P}^2$ is a scale factor, $\boldsymbol{\chi}_P \odot V_{\mathbf{e}_P} \odot \boldsymbol{\Pi}_P^{\odot-}$ reduces to a diagonal matrix with ith element v_{ii}/π_i if $i \in s$ and 0 if $i \notin s$.

3.2 Improved approximate to BLUE for design based estimation from sample survey data

The permutation or re-ordering of $\boldsymbol{\chi}_P \odot V_{\mathbf{e}_P} \odot \boldsymbol{\Pi}_P^{\odot -}$ so that the sampled elements occur in the first n rows and n columns is straightforward. Denote this re-ordering by $\mathbf{V}_{\mathbf{e}_P,\Pi_P,s}$ or, since there is no ambiguity, by $\mathbf{V}_{\mathbf{e}_P,s}$.

Now, $\mathbf{V}_{\mathbf{e}_P,s}$ has a generalized inverse (which is also the Moore-Penrose inverse)

$$\mathbf{V}^{+}_{\mathbf{e}_P,s} = \begin{pmatrix} \mathbf{V}^{-1}_{\mathbf{e}_P,s,n} & 0 \\ 0 & 0 \end{pmatrix}, \tag{7}$$

where $\mathbf{V}_{\mathbf{e}_P,s,n}$ is $n{\times}n$ with ijth element v_{ij}/π_{ij}. Because $0 < \pi_i \le 1$ for $i = 1,2,\ldots,N$, the inverse $\mathbf{V}^{-1}_{\mathbf{e}_P,s,n}$ exists provided $\mathbf{V}_{\mathbf{e}_P}$ is full rank.

After the rows of $\mathbf{X}_P$ have also been appropriately permuted to match the permutation for the rows and columns of $\mathbf{V}^{+}_{\mathbf{e}_P,s}$ we have

$$\mathbf{X}_P^T V^{+}_{\mathbf{e}_P,s}\mathbf{X}_P = \begin{pmatrix} \mathbf{X}^T & \mathbf{X}^T_{\sim s} \end{pmatrix} \begin{pmatrix} \boldsymbol{V}^{-1}_{\mathbf{e}_P,s,n} & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \mathbf{X} \\ \mathbf{X}_{\sim s} \end{pmatrix} = \mathbf{X}^T \boldsymbol{V}^{-1}_{\mathbf{e}_P,s,n}\mathbf{X} \tag{8}$$

and

$$\mathbf{X}_P^T V^{+}_{\mathbf{e}_P,s}\mathbf{Y}_P = \begin{pmatrix} \mathbf{X}^T & \mathbf{X}^T_{\sim s} \end{pmatrix} \begin{pmatrix} \boldsymbol{V}^{-1}_{\mathbf{e}_P,s,n} & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \mathbf{Y} \\ \mathbf{Y}_{\sim s} \end{pmatrix} = \mathbf{X}^T \boldsymbol{V}^{-1}_{\mathbf{e}_P,s,n}\mathbf{Y}, \tag{9}$$

where $\mathbf{Y}_{\sim s}$ denotes the y-values and $\mathbf{X}_{\sim s}$ the auxiliary variables for the non-sampled elements.

Thus, by extension of (2) and (3), the approximate BLUE based on the sampled elements only is:

$$\hat{\beta} = (\mathbf{X}^T \boldsymbol{V}^{-1}_{\mathbf{e}_P,s,n}\mathbf{X})^{-1}\mathbf{X}^T \boldsymbol{V}^{-1}_{\mathbf{e}_P,s,n}\mathbf{Y} \tag{10}$$

with

$$\hat{\mathbf{V}}(\hat{\boldsymbol{\beta}}) = [(\mathbf{X}^T \boldsymbol{V}^{-1}_{\mathbf{e}_P,s,n}\mathbf{X})^{-1}\mathbf{X}^T \boldsymbol{V}^{-1}_{\mathbf{e}_P,s,n}]\ \boldsymbol{V}_{\mathbf{e}}\,[\boldsymbol{V}^{-1}_{\mathbf{e}_P,s,n}\mathbf{X}(\mathbf{X}^T \boldsymbol{V}^{-1}_{\mathbf{e}_P,s,n}\mathbf{X})^{-1}], \tag{11}$$

where, as in Section 2.1, $\mathbf{V}(\mathbf{e}) = \mathbf{V_e}$ and $\mathbf{e}$ denotes that part of $\mathbf{e}_P$ that corresponds to the n sampled elements.

When $\mathbf{V}_{\mathbf{e}_P}$ is diagonal, i.e., $\mathbf{V}_{\mathbf{e}_P} = \sigma^2_{\mathbf{e}_P}\mathbf{I}$, then $\mathbf{V}_{\mathbf{e}_P,s,n}$ is also diagonal and (10) and (11) reduce to (2) and (3) respectively.

One major advantage of (10) and (11) over (2) and (3) is that they can be applied to estimation of fixed effects in mixed linear models, where incorporation of the random effects into $\mathbf{V}_{\mathbf{e}_P}$ means that $\mathbf{V}_{\mathbf{e}_P}$ is no longer a diagonal matrix, so that (2) and (3) cannot be applied.

3.3 Is $\mathbf{V}_{\mathbf{e}_P,s}$ positive semidefinite, and is this necessary for BLUE?

Perhaps surprisingly, following from [10] and the earlier results of [14], to produce viable estimates of $\boldsymbol{\beta}$ from (9), $\mathbf{V}_{\mathbf{e}_P,\boldsymbol{\Pi}_P,s} = \mathbf{V}_{\mathbf{e}_P,s} = \boldsymbol{\chi}_P \odot \mathbf{V}_{\mathbf{e}_P} \odot \boldsymbol{\Pi}_P^{\odot -}$ need not be positive semidefinite, and its submatrix $\mathbf{V}_{\mathbf{e}_P,s,n}$ need not be positive definite.

This can be seen for fixed effect linear models from an extension to [14]. Given a linear model of the form (1) with error variance $\mathbf{V}_1$, then for any $\mathbf{V}_2$ of the form

$$\mathbf{V}_2 = \lambda\mathbf{V}_1 + \mathbf{X}\mathbf{K}_{\mathbf{X}}\mathbf{X}^T + \mathbf{V}_1\mathbf{X}_\perp\mathbf{K}_{\mathbf{X}_\perp}\mathbf{X}_\perp^T\mathbf{V}_1, \tag{12}$$

where $\lambda \neq 0$, $\mathbf{X}_\perp$ is a matrix orthogonal to $\mathbf{X}$ such that $(\mathbf{X} : \mathbf{X}_\perp)$ is of full column rank, and $\mathbf{K}_{\mathbf{X}_\perp}$ and $\mathbf{K}_{\mathbf{X}_\perp}$ are arbitrary, the BLUE of β is unchanged. If $\lambda = -1$, $\mathbf{K}_{\mathbf{X}_\perp} = \mathbf{0}$ and $\mathbf{K}_{\mathbf{X}_\perp} = \mathbf{0}$, for example, then clearly $\mathbf{V}_2$ is not positive semidefinite.

Rao's and Haslett and Puntanen's results are relevant here because the diagonal elements of the $n{\times}n$ submatrix $\mathbf{V}_{\mathbf{e}_P,s,n}$ are v_{ii}/π_i and its ijth element is v_{ij}/π_{ij}. Of course, $v_{ij}/\sqrt{v_{ii}v_{jj}} \le 1$, but generally $1/\pi_i \ll 1/\pi_{ij}$, because the joint selection probabilities are such that (unless the survey design is clustered) $\pi_{ij} \approx \pi_i\pi_j$. So the diagonal elements of $\mathbf{V}_{\mathbf{e}_P,s,n}$ may be much smaller than its off-diagonal elements and $\mathbf{V}_{\mathbf{e}_P,s,n}$ may have at least some negative eigenvalues.

The core problem parallels issues of negative estimates of variance for the Horvitz-Thompson estimator.

4 Properties of $\mathbf{V}_{\mathbf{e}_P,s}$

4.1 Adjusting $\mathbf{V}_{\mathbf{e}_P,s}$ to be positive semidefinite

Recall that $\mathbf{V}_{\mathbf{e}_P,s} = \boldsymbol{\chi}_P \odot \mathbf{V}_{\mathbf{e}_P} \odot \boldsymbol{\Pi}_P^{\odot -}$, where all three component matrices are $N{\times}N$, $\mathbf{V}_{\mathbf{e}_P}$ is positive semidefinite and known or estimable, $\boldsymbol{\Pi}_P$ and hence $\boldsymbol{\Pi}_P^{\odot -}$is positive but not necessarily positive semidefinite, and $\boldsymbol{\chi}_P$ is also a non-negative matrix dependent on the sample s and containing n^2 ones, n of them on the diagonal (corresponding to the sampled elements), and N^2-n^2 zeros.

Now $\boldsymbol{V}$ positive semidefinite if $\mathbf{x}^T\mathbf{V}\mathbf{x} \geq 0$, for all $\boldsymbol{x}$, so that after any choice of permutation of the rows and columns of $\boldsymbol{V}$, and choosing $\mathbf{x}_0 = \begin{pmatrix} \mathbf{x}_1 & \mathbf{0} \end{pmatrix}$, the submatrix $\mathbf{V}_{11}$ corresponding to $\mathbf{x}_1$ has the property that $\mathbf{x}_1^T\mathbf{V}_{11}\mathbf{x}_1 \geq 0$ and so is positive semidefinite.

Nevertheless, $\mathbf{V}_{\mathbf{e}_P,s}$ not being positive semidefinite is an undesirable property so, to avoid complications entirely, for survey designs where the intention is to fit linear models, when $\mathbf{V}_{\mathbf{e}_P}$ (positive semidefinite) is known, $\boldsymbol{\Pi}_P$ (nonnegative, but not necessarily positive semidefinite) should be chosen via the survey design so that $\mathbf{V}_{\mathbf{e}_P} \odot \boldsymbol{\Pi}_P^{\odot -}$ and hence $\mathbf{V}_{\mathbf{e}_P,s} = \boldsymbol{\chi}_P \odot \mathbf{V}_{\mathbf{e}_P} \odot \boldsymbol{\Pi}_P^{\odot -}$ is positive semidefinite, because then, for any sample s, $\mathbf{V}_{\mathbf{e}_P,s,n}$ must also be positive semidefinite.

An alternative may be to consider using (12) to generate an equivalent matrix for the error covariance structure in BLUE so that the equivalent matrix is positive semidefinite. For example, it may suffice to use $(\mathbf{V}_{\mathbf{e}_P} \odot \boldsymbol{\Pi}_P^{\odot -}) + k_1\mathbf{X}_P\mathbf{X}_P^T + k_2(\mathbf{V}_{\mathbf{e}_P} \odot \boldsymbol{\Pi}_P^{\odot -})\mathbf{X}_{P\perp}\mathbf{X}_{P\perp}^T(\mathbf{V}_{\mathbf{e}_P} \odot \boldsymbol{\Pi}_P^{\odot -})$, where $\mathbf{X}_{P\perp}$ spans the space orthogonal to $\mathbf{X}_P$ and k_1 and k_2 are suitable scalars. Provided the second and third terms taken together span the same space as $(\mathbf{V}_{\mathbf{e}_P} \odot \boldsymbol{\Pi}_P^{\odot -})$, k_1 and k_2 can be increased until they dominate and the resulting covariance structure is positive semidefinite. This is however not always possible; as the sum of the second and third terms imply some restrictions. Matrices formed from the sum of these two terms do not span the same space as $(\mathbf{V}_{\mathbf{e}_P} \odot \boldsymbol{\Pi}_P^{\odot -})$, because vectors generated from them are either in the space spanned by $\mathbf{X}_P$ or the space spanned by $(\mathbf{V}_{\mathbf{e}_P} \odot \boldsymbol{\Pi}_P^{\odot -})\mathbf{X}_{P\perp}$, but cannot be in both. More formally, this result follows from noting that

$$\begin{aligned}
&\mathbf{X}_P\mathbf{K}_\mathbf{X}\mathbf{X}_P^T + (\boldsymbol{V}_{\mathbf{e}_P} \odot \boldsymbol{\Pi}_P^{\odot -})\mathbf{X}_{P\perp}\mathbf{K}_{\mathbf{X}_\perp}\mathbf{X}_{P\perp}^T(V_{\mathbf{e}_P} \odot \boldsymbol{\Pi}_P^{\odot -}) \\
&= \begin{pmatrix} \mathbf{X}_P & (\boldsymbol{V}_{\mathbf{e}_P} \odot \boldsymbol{\Pi}_P^{\odot -})\mathbf{X}_{P\perp} \end{pmatrix} \begin{pmatrix} \mathbf{K}_\mathbf{X} & \mathbf{0} \\ \mathbf{0} & \mathbf{K}_{\mathbf{X}_\perp} \end{pmatrix} \begin{pmatrix} \mathbf{X}_P^T \\ \mathbf{X}_{P\perp}^T(V_{\mathbf{e}_P} \odot \boldsymbol{\Pi}_P^{\odot -}) \end{pmatrix}.
\end{aligned}$$

A simple example in two dimensions would be vectors that can be along the x-axis or the y-axis only, when compared with whole of two dimensional space.

4.2 Cauchy's interlace theorem and $\mathbf{V}_{\mathbf{e}_P,s}$

An alternative approach is to use Cauchy's interlace theorem which, as stated in [6], says that the characteristic polynomial of a Hermitian (or, if all entries are real, symmetric) matrix is interlaced by the characteristic polynomial of any of its principal submatrices. By interlacing is meant that if two polynomials of order n and n-1 have roots $q_1, q_2, \ldots, q_n$ and $r_1, r_2, \ldots r_{n-1}$ respectively then $q_1 \leq r_1 \leq q_2 \leq r_2 \ldots \leq r_{n-1} \leq q_n$. An alternative statement of the theorem is that the eigenvalues of a Hermitian matrix of order n are interlaced with those of any principal submatrix of order n−1, where a principal submatrix is obtained by compression, i.e., by removing any $n - m$ rows and the same $n - m$ columns. See also [13]. Of course, Cauchy's interlace theorem may be re-applied for any of n-2, n-3,…,2, 1 so that the eigenvalues q_1 and q_n constitute upper and lower bounds respectively of the eigenvalues of any 2×2 principal submatrix.

As a corollary, a necessary (but not sufficient) condition for all the eigenvalues of a matrix (with real eigenvalues) to be non-negative is that the eigenvalues of all its 2×2 principal submatrices are non-negative. Consider then a 2×2 symmetric matrix $\boldsymbol{A}$ with real entries, the two diagonal entries equal to one, and the two off-diagonal entries equal to a.Then $\boldsymbol{A}$ has eigenvalues (1+a) and (1-a) and corresponding eigenvectors $(1 \quad 1)^T/\sqrt{2}$ and $(1 \quad -1)^T/\sqrt{2}$. For $|a| \leq 1$, $\boldsymbol{A}$ is positive semi definite, and for $|a| < 1$, $\boldsymbol{A}$ is positive definite.

Applying this result to the 2×2 principal submatrices of $\mathbf{V}_{\mathbf{e}_P,s} = \boldsymbol{\chi}_P \odot \mathbf{V}_{\mathbf{e}_P} \odot \boldsymbol{\Pi}_P^{\odot-}$, rescaled to be a correlation matrix rather than covariance matrix (so that all diagonal entries are zeros or ones), yields matrices of three types:

– a matrix of zeros,
– a matrix with either the 1st or 2nd diagonal entry is one (and the other zero), and both off diagonal entries are zero, and
– a matrix of the form of $\boldsymbol{A}$ with $v_{ii}>0$, $v_{jj}>0$ and

$$a = \frac{v_{ij}}{\sqrt{v_{ii}v_{jj}}}\frac{\sqrt{\pi_{ii}\pi_{jj}}}{\pi_{ij}}.$$

The first two types are positive semidefinite, so impose no constraints via the necessary condition that $\mathbf{V}_{\mathbf{e}_P,s}$ be positive semidefinite. For the third type, the necessary condition imposed on $\mathbf{V}_{\mathbf{e}_P,s}$ for positive semidefinite is that:

$$\left|\frac{v_{ij}}{\sqrt{v_{ii}v_{jj}}}\frac{\sqrt{\pi_{ii}\pi_{jj}}}{\pi_{ij}}\right| \le 1$$

or, since v_{ii}, v_{jj}, π_{ii}, π_{jj} and π_{ij} are all necessarily positive for $i = 1,2,\ldots,N$ and $j = 1,2,\ldots,N$:

$$\frac{|v_{ij}|}{\sqrt{v_{ii}v_{jj}}} \le \frac{\pi_{ij}}{\sqrt{\pi_{ii}\pi_{jj}}}. \tag{13}$$

For example, for simple random sampling without replacement, $\pi_{ii} = \pi_{jj} = n/N$ and $\pi_{ij} = n(n-1)/N(N-1)$ for $i = 1,2,\ldots,N$, and $j = 1,2,\ldots,N$, and (13) reduces to $|corr(y_i, y_j)| \le (n-1)/(N-1)$, where *corr* denotes correlation. The lower bound is usually met, but the upper bound can often be exceeded when there are neighbourhood effects, for example in surveys linked to socio-economic status. For random samples that are not simple random, the results of [9] (where selection and joint selection probabilities are specified for a range of sample designs can be utilised.

Nevertheless, whether for simple random sampling or other random sampling schemes, the necessary condition (13) for positive semidefiniteness of $\mathbf{V}_{\mathbf{e}_P,s}$ is not generally met.

However from (12), the superpopulation variance structure can be modified without changing the BLUE of $\boldsymbol{\beta}$ in (4) for example via $\mathbf{V}_2 = \lambda\mathbf{V}_1 + \mathbf{X}\mathbf{A_0}\mathbf{X}^T$. If $\mathbf{X}\mathbf{A_0}\mathbf{X}^T$can set to have all elements equal to a_0, (as when (4) contains a common intercept term so that the vector of ones $\mathbf{1} \in C(\mathbf{X})$ where C denotes column space, and the corresponding diagonal element of is set to a_0), then (13) becomes:

$$\frac{|v_{ij} + a_0|}{\sqrt{(v_{ii} + a_0)(v_{jj} + a_0)}} \le \frac{\pi_{ij}}{\sqrt{\pi_{ii}\pi_{jj}}}. \tag{14}$$

In the situation where v_{ii} for $i = 1,2,\ldots,N$, are equal, setting $a_0 = -v_{ij}$ suffices to ensure (14) is true. Interestingly, although this is the stronger result, it has a parallel with (2) and (3) where it is effectively assumed that $v_{ij} = 0$.

A common structure for $\mathbf{V}_{\mathbf{e}_P}$, the error covariance structure for (4), is block diagonal with blocks all of the form $\sigma^2[(1-\rho)\mathbf{I} + \rho\mathbf{1}\mathbf{1}^T)$ where $\mathbf{I}$ is the identity matrix and $\mathbf{1}$ is a vector of ones. Such matrices occur, for example, in cluster sampling where there is intra-cluster correlation for sampled elements within clusters but no correlation between clusters. Setting $a_0 = -\sigma^2\rho$ then ensures that the necessary condition for positive semidefiniteness of $\mathbf{V}_{\mathbf{e}_P,s} = \boldsymbol{\chi}_P \odot \mathbf{V}_{\mathbf{e}_P} \odot \boldsymbol{\Pi}_P^{\odot-}$ is met for any noninformative survey design.

Situations where there is not a common variance σ^2 are more difficult, since a suitable form of a_0 may not exist (for example, if there is a v_{ij} such that $v_{ij} \ge v_{i'i'}$ for some i, j, and i', where the range of all three indices is $1, 2, \ldots, N$). This complication reflects the issues raised for the necessary and sufficient conditions given in Section 4.1.

5 Conclusions

Whether the results of [14] or [10] are used, or the Cauchy interlace theorem is utilised to give necessary conditions, the positive semidefiniteness of some possibly augmented form of covariance structure for the error in a linear model constructed from survey data cannot be guaranteed, except if joint selection probabilities are ignored or equivalently the covariances between population elements are ignored.

For particular types of covariance structure often used in linear models for survey data however, where the covariance matrix is block diagonal with common correlation and scale, it is possible to ensure the necessary condition for positive definiteness is met by using a suitable transformation.

Whether more general and/or stronger results can be obtained is an open question. So is the extent to which structures used in linear models as covariances can depart from positive semidefiniteness, without affecting the numerical results of least squares algorithms used to determine BLUEs.

References

[1] D. Basu. An essay on the logical foundations of survey sampling, Part One, In Foundations of Statistical Inference, V.P. Godambe and D.A. Sprott (eds). Holt, Rinehart & Winston, 203-242, 1971.

[2] C-M. Cassel, C.E Särndal, J.H. Wretman. Foundations of Inference in Survey Sampling, Wiley, 1977.

[3] R.L. Chambers, C.J. Skinner (eds.). Analysis of Survey Data, Wiley, 2003.

[4] W.G. Cochran. Sampling Techniques, Wiley, 1953.

[5] W.G. Cochran. Sampling Techniques, 3rd edition, Wiley, 1977.

[6] S. Fisk. A very short proof of Cauchy's interlace theorem for eigenvalues of Hermitian matrices, arXiv:math/0502408v1 [math.CA] 18 Feb 2005, http://arxiv.org/pdf/math/0502408v1.pdf, 2005.

[7] W.A. Fuller. Sampling Statistics, Wiley, 2009.

[8] M.H. Hansen, W.N. Hurwitz, W.G. Madow. Sample Survey Methods and Theory, Volumes 1 and 2, Wiley, 1953.

[9] S. Haslett. The linear non-homogeneous estimator in sample surveys, Sankhyä Ser. B, 47(1985), 101-117.

[10] S. Haslett, S. Puntanen. Equality of the BLUEs and/or BLUPs under two linear models using stochastic restrictions, Statistical Papers, 51(2010), 465-475.

[11] D.G. Horvitz, D.J. Thompson. A generalization of sampling without replacement from a finite universe, Journal of the American Statistical Association, 47(1952) 663-685.

[12] R. Lehtonen, E. Pahkinen. Practical Methods for Design and Analysis of Complex Surveys, 2nd Edition, Wiley, 2003.

[13] A. Mercer, P. Mercer. Cauchy's interlace theorem and lower bounds for the spectral radius, International Journal of Mathematics and Mathematical Sciences, 23(2000), 563-566.

[14] C.R. Rao. A note on a previous lemma in the theory of least squares and some further results, Sankhyä Ser. A 30(1968), 259-266.

[15] R.M. Royal, W.G.Cumberland. Variance estimation in finite population sampling, Journal of the American Statistical Association, 73(1978) 351-358.

[16] C-E. Särndal, B. Swensson, J. Wretman. Model Assisted Survey Sampling, Springer, 1992.

[17] C.J. Skinner, D. Holt, T.M.F. Smith (eds.). Analysis of Complex Surveys, Wiley, 1989.