Spec. Matrices 2016; 4:218–224

DE GRUYTER OPEN

Research Article

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

Open Access

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