Journal of Multivariate Analysis 128 (2014) 154–164

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

Optimal and minimax prediction in multivariate normal populations under a balanced loss function

Guikai Hu [a,b], Qingguo Li [a,*], Shenghua Yu [c]

[a] School of Mathematics and Econometrics, Hunan University, Changsha 410082, China
[b] School of Science, East China Institute of Technology, Nanchang 330013, China
[c] School of Economics and Trade, Hunan University, Changsha 410079, China

ARTICLE INFO

Article history:
Received 23 September 2013
Available online 31 March 2014

AMS subject classifications:
62M20
62C15

Keywords:
Optimal predictor
Minimax predictor
Finite population regression coefficient
Balanced loss function

ABSTRACT

Under a balanced loss function, we investigate the optimal and minimax prediction of finite population regression coefficient in a general linear regression superpopulation model with normal errors. The best unbiased prediction (BUP) is obtained in the class of all unbiased predictors. The minimax predictor (MP) is also obtained in the class of all predictors. We prove that MP is unique in the class of all predictors and is better than BUP in a certain region of parameter space. Next, we give some conditions for optimality of the simple projection predictor (SPP) and prove that MP dominates SPP on certain occasions.

1. Introduction

Let us consider finite population $\mathscr{P} = \{1, \ldots, N\}$ as the collection of a known number N of identifiable units. Associated with the ith unit of $\mathscr{P}$, there are $p + 1$ quantities: $y_i, x_{i1}, \ldots, x_{ip}$, where all but y_i are known, $i = 1, \ldots, N$. Denote $y = (y_1, \ldots, y_N)'$ and $X = (X_1, \ldots, X_N)'$, where $X_i = (x_{i1}, \ldots, x_{ip})'$, $i = 1, \ldots, N$. We express the relationships among the variables by the linear model

$$y = X\beta + \varepsilon, \tag{1.1}$$

where β is a $p \times 1$ unknown parameter vector, ε is an $N \times 1$ normal random vector with mean vector 0 and covariance matrix $\sigma^2 V$, V is a known positive definite matrix, but the parameter $\sigma^2 > 0$ is unknown.

The finite population regression coefficient is denoted as $\beta_N = (X'V^{-1}X)^{-1}X'V^{-1}y$. In order to predict the finite population regression coefficient β_N, let us select a sample $\mathbf{s}$ of size n from $\mathscr{P}$ according to some specified sampling plan so as to obtain information on β_N. Let $\mathbf{r} = \mathscr{P} - \mathbf{s}$ be the unobserved part of $\mathscr{P}$. After the sample $\mathbf{s}$ has been selected, we may reorder the elements of y such that we have the corresponding partitions of y, X and V, that is

$$y = \begin{pmatrix} y_s \\ y_r \end{pmatrix}, \qquad X = \begin{pmatrix} X_s \\ X_r \end{pmatrix}, \qquad V = \begin{pmatrix} V_s & V_{sr} \\ V_{rs} & V_r \end{pmatrix}.$$

* Corresponding author.
E-mail addresses: huguikai97@163.com (G. Hu), liqingguoli@yahoo.com.cn (Q. Li), yshh1966@163.com (S. Yu).

http://dx.doi.org/10.1016/j.jmva.2014.03.014
0047-259X/

Following Bolfarine et al. [8], we can write the finite population regression coefficient β_N as

$$\begin{aligned}\beta_N &= (X'V^{-1}X)^{-1}X'V^{-1}y \\ &= \left[(X_s'\,X_r')\begin{pmatrix}V_s & V_{sr}\\ V_{rs} & V_r\end{pmatrix}^{-1}\begin{pmatrix}X_s\\ X_r\end{pmatrix}\right]^{-1}(X_s'\,X_r')\begin{pmatrix}V_s & V_{sr}\\ V_{rs} & V_r\end{pmatrix}^{-1}\begin{pmatrix}y_s\\ y_r\end{pmatrix} \\ &= Q_s y_s + Q_r y_r,\end{aligned}$$

where

$$\begin{aligned}&Q_s = G^{-1}MC^{-1}, \qquad Q_r = G^{-1}DE^{-1}, \\ &M = X_s' - X_r'V_r^{-1}V_{rs}, \qquad C = V_s - V_{sr}V_r^{-1}V_{rs}, \\ &D = X_r' - X_s'V_s^{-1}V_{sr}, \qquad E = V_r - V_{rs}V_s^{-1}V_{sr},\end{aligned}$$

and

$$G = MC^{-1}X_s + DE^{-1}X_r.$$

In the literature, a lot of predictions for the finite population regression coefficient have been given. For example, Bolfarine and Zacks [6,7] studied Bayes and minimax prediction under square error loss function. Bolfarine et al. [8] obtained the optimal prediction under the generalized prediction mean squared error. In their studies, the quadratic function and its variants were usually considered as the loss functions. However, in regression analysis, we are often interested in using an estimator which has high precision of estimation and high goodness of fit. In this situation, Zellner [16] proposed a balanced loss function which takes account of both precision of estimation and goodness of fit. Balanced loss function is a more comprehensive and reasonable standard than quadratic loss and residual sum of squares. It has attracted considerable attention under different setups in the literature. For more details, the readers are referred to Arashi [1], Xu and Wu [15], Hu and Peng [10,11], and Cao [9].

Therefore, the problem of prediction in a superpopulation regression model under balanced loss function arises naturally. Recently, Bansal and Aggarwal [2–4] have considered Bayes prediction of finite population regression coefficient under balanced loss function. It is interesting for us to consider the optimal prediction and minimax prediction of finite population regression coefficient in the normal populations under a balanced loss function. In next section, we will give the best predictor in the class of all unbiased predictors. In Section 3, we will discuss the minimax predictor in the class of all predictors. In Section 4, we will discuss the risk comparison of BUP and MP. In Section 5, we compare SPP with BUP and MP. Concluding remarks are given in Section 6.

2. Optimal predictor

In this section, we will discuss the best unbiased prediction of finite population regression coefficient in the class of all predictors. For every $\beta \in R^p$ and $\sigma^2 > 0$, where R^p stands for the set composed of all $p \times 1$ real vectors, we define the loss function as

$$L(\delta(y_s), \beta_N) = \frac{\theta(y_s - X_s\delta(y_s))'V_s^{-1}(y_s - X_s\delta(y_s)) + (1-\theta)(\delta(y_s) - \beta_N)'X_s'V_s^{-1}X_s(\delta(y_s) - \beta_N)}{\sigma^2 + \beta'X_s'V_s^{-1}X_s\beta}, \tag{2.1}$$

where $\theta \in [0, 1]$ is a weight coefficient. The numerator of the loss function is proposed using the idea of Zellner's balanced loss and the theory of generalized least squares estimator. It can be seen that the best unbiased prediction under loss (2.1) is equivalent to the best unbiased prediction under the numerator. We add to the denominator $\sigma^2 + \beta'X_s'V_s^{-1}X_s\beta$ in the loss function (2.1) such that the maximum risk function of $\delta(y_s)$ does not rely on parameters σ^2 and β. This make us discuss the minimax prediction under the same loss function better.

We denote $\mathscr{D}$ by the space of all predictors $\delta(y_s)$ of β_N such that the expected value of the loss $L(\delta(y_s), \beta_N)$ is finite. For every $\beta \in R^p$ and $\sigma^2 > 0$, we define the risk function of $\delta(y_s)$ as

$$R(\delta(y_s), \beta_N) = E(L(\delta(y_s), \beta_N)).$$

If the element is finite, thus the optimality of a predictor $\delta(y_s) \in \mathscr{D}$, such as domination, admissibility, minimaxity and so on, can be evaluated by its risk in the range spaces of the risk function. This section deals with the optimal predictor of β_N, which is defined as follows.

Definition 2.1. An unbiased predictor $\delta^*(y_s)$ is said to be the best unbiased predictor of β_N, if for any unbiased predictor $\delta(y_s)$,

$$R(\delta^*(y_s), \beta_N) \le R(\delta(y_s), \beta_N)$$

holds for all $\beta \in R^p$ and $\sigma^2 > 0$.