Open Mathematics

Research Article

Chao Bai* and Haiqi Li

Simultaneous prediction in the generalized linear model

https://doi.org/10.1515/math-2018-0087
Received November 28, 2017; accepted July 17, 2018.

Abstract: This paper studies the prediction based on a composite target function that allows to simultaneously predict the actual and the mean values of the unobserved regressand in the generalized linear model. The best linear unbiased prediction (BLUP) of the target function is derived. Studies show that our BLUP has better properties than some other predictions. Simulations confirm its better finite sample performance.

Keywords: Generalized linear model, Simultaneous prediction, Best linear unbiased prediction

MSC: 62M20, 62J12

1 Introduction

Generalized linear models have a long history in the statistical literature and have been used to analyze data from various branches of science on account of both mathematical and practical convenience. Consider the following generalized linear model:

$$\begin{pmatrix} y \\ y_0 \end{pmatrix} = \begin{pmatrix} X \\ X_0 \end{pmatrix} \beta + \begin{pmatrix} \varepsilon \\ \varepsilon_0 \end{pmatrix} \tag{1}$$

where
y is the n-dimensional vector of observed data;
y_0 is the m-dimensional vector of unobserved values that is to be predicted;
X and X_0 are $n \times p$ and $m \times p$ known matrices of explanatory variables. Let $\mathrm{rk}(A)$ denote the rank of matrix A and suppose $\mathrm{rk}(X) \leq p$;
β is the $p \times 1$ unknown vector of regression coefficients, and
ε and ε_0 are random errors with zero mean and covariance matrix

$$\mathrm{Cov}(\varepsilon, \varepsilon_0') = \begin{pmatrix} \Sigma & V' \\ V & \Sigma_0 \end{pmatrix},$$

where $\Sigma \geq 0$ and $\Sigma_0 \geq 0$ are known positive semi-definite matrices of arbitrary ranks.

The problem of predicting unobserved variables plays an important role in decision making and has received much attention in recent years. For the prediction of y_0 in model (1), [1] obtained the best linear unbiased predictor (BLUP) when $\Sigma > 0$. The Bayes and minimax prediction were obtained by [2] when random errors were normally distributed. [3] and [4] derived the linear minimax prediction under a modified quadratic loss function. [5] considered the optimal Stein-rule prediction. [6] reviewed the existing theory of minimum mean squared error loss predictors and made an extension based on the principle of equivariance.

***Corresponding Author: Chao Bai:** College of Finance and Statistics, Hunan University, Changsha, Hunan, 410006, China and College of Science, Central South University of Forestry and Technology, Changsha, Hunan, 410000, China, E-mail: bbc144@163.com
Haiqi Li: College of Finance and Statistics, Hunan University, Changsha, Hunan, 410006, China, E-mail: lihaiqi00@hnu.edu.cn

Open Access. © 2018 Bai and Li, published by De Gruyter. This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 License.

[7] investigated the admissibility of linear predictors with inequality constraints under the mean squared error loss function. Another interested subject of prediction relates to the mean of y_0, since [8] figured out that the best predictor of y_0 is the conditional mean under the criterion of minimum mean squared error. In model (1), prediction of the mean value of y_0 (namely = $X_0\beta$) relates naturally to the plug-in estimators of parameter β. [9] proposed the simple projection predictor (SPP) of $X_0\beta$ by plugging in the best linear unbiased estimator (BLUE) of β. [10, 11] considered plugging in the prediction of β under the balanced loss function. The plug-in approach spawned a large literature for the derivation of combined prediction, see [12–14].

Generally, predictions are investigated either for y_0 or for $\mathrm{E}y_0$ at a time. However, sometimes in the fields of medicine and economics, people would like to know the actual value of y_0 and its mean value $\mathrm{E}y_0$ simultaneously. For example, in the financial markets, some investors may want to know the actual profit while others would be more interested in the mean profit. Therefore, in order to meet different requirements, the market manager should acquire both the prediction of the actual profit and the prediction of the mean profit simultaneously. Let aside investors' demands and from the point of view of a decision maker, the market manager needs to determine which prediction should be preferred or provides another comprehensive combined prediction both of the actual and the mean profit based on empirical data. [15] gave other examples of practical situations where one is required to predict both the mean and the actual values of a variable. Under these circumstances, we consider predictions of the following target function

$$\delta = \lambda y_0 + (1-\lambda)\mathrm{E}y_0, \tag{2}$$

where $\lambda \in [0, 1]$ is a non-stochastic weight scalar representing the preference to the prediction of actual and the mean value of the studied variable. Note that, $\delta = y_0$ if $\lambda = 1$ and $\delta = \mathrm{E}y_0$ if $\lambda = 0$, which means predicting δ can achieve the prediction of y_0 and $\mathrm{E}y_0$ simultaneously. If $0 < \lambda < 1$, then prediction of δ balances the prediction of actual and the average value of y_0. Besides, the unbiased prediction of δ is also the unbiased prediction of y_0 or $\mathrm{E}y_0$. Therefore, δ is more sensitive and inclusive to be studied.

Studies on the prediction of δ have been carried out in the literature from various perspective. The properties of the predictors by plugging in Stein-rule estimators have been concerned by [16–18]. [19] investigated the Stein-rule prediction for δ in linear regression model when the error covariance matrix was positive definite yet unknown. [20] studied the admissible prediction of δ. [21, 22] and [23] considered predictors for δ in linear regression models with stochastic or non-stochastic linear constraints on the regression coefficients. The issues of simultaneous prediction in measurement error models have been addressed in [24] and [25]. [26] considered a scalar multiple of the classical prediction vector for the prediction of δ and discussed the performance properties.

For model (1), most former work concerned about biased prediction under $\Sigma > 0$ (including the special case $\Sigma = I$), and did not discuss the value of the weight scalar λ in (2). In this paper, supposing $\Sigma \geq 0$, we studied the best linear unbiased prediction (BLUP) of δ and make some comparisons to the usual BLUPs of y_0 and $\mathrm{E}y_0$. We also propose a method to choose the value of λ in (2), which can give the way to determine which prediction of δ or y_0 or $\mathrm{E}y_0$ should be provided by finite sample data.

The rest of the paper is organized as follows. In Section 2, we derive the BLUPs of the target function (2) in the generalized linear model, and discuss the efficiency of our BLUP comparing to the usual BLUP and SPP. Simulation studies are provided in Section 3 to illustrate the determination of the weight scalar in our BLUP and the performance of our proposed BLUP comparing to the other two predictors. Concluding remarks are given in Section 4.

2 The BLUP of δ and its efficiency

Denote $\mathcal{LH} = \{Cy \mid C \text{ is an } m \times n \text{ matrix}\}$ as the set of all the homogeneous linear predictor of y_0. Denote $\hat{\delta}_{BLUP}$ as the best linear unbiased predictor of δ in model (1). In this section, we first derive the expressions of $\hat{\delta}_{BLUP}$ in $\mathcal{LH}$, and then study its performance comparing to the BLUP of y_0 and the SPP of $\mathrm{E}y_0$. All of the predictors

discussed in this paper are derived under the criterion of minimum mean squared error. Some preliminaries and basic results are given as follows:

Definition 2.1. *The predictor $\hat{\delta}$ of δ is unbiased if $E\hat{\delta} = E\delta$.*

Definition 2.2. *δ is linearly predictable if there exists a linear predictor Cy in $\mathcal{LH}$ such that Cy is an unbiased predictor of d.*

Lemma 2.3. *In model (1), δ is linearly predictable if there exists a matrix C such that $CX = X_0$, or $\mathcal{M}(X_0') \subseteq \mathcal{M}(X')$.*

Proof. From Definition 2.1 and 2.2, there exists a matrix C such that $\mathrm{E}(Cy) = \mathrm{E}\delta$ for any β, namely $CX = X_0$ or $X'C' = X_0'$, which is equivalent to $\mathcal{M}(X_0') \subseteq \mathcal{M}(X')$. □

If not specified otherwise, the variables we aim to predict in this paper are all linearly predictable.

Lemma 2.4 ([27])**.** *Suppose the $n \times n$ matrix $\Sigma \geq 0$ and let X be an $n \times p$ matrix, then*

$$\begin{pmatrix} \Sigma & X \\ X' & 0 \end{pmatrix}^- = \begin{pmatrix} T^+ - T^+X(X'T^+X)^-X'T^+ & T^+X(X'T^+X)^- \\ (X'T^+X)^-X'T^+ & (X'T^+X)^-(X'T^+X) - (X'T^+X)^- \end{pmatrix},$$

where $T = \Sigma + XX'$. Especially, if $\Sigma > 0$, then

$$\begin{pmatrix} \Sigma & X \\ X' & 0 \end{pmatrix}^- = \begin{pmatrix} \Sigma^{-1} - \Sigma^{-1}X(X'\Sigma^{-1}X)^-X'\Sigma^{-1} & \Sigma^{-1}X(X'\Sigma^{-1}X)^- \\ (X'\Sigma^{-1}X)^-X'\Sigma^{-1} & -(X'\Sigma^{-1}X)^- \end{pmatrix}.$$

Lemma 2.5. *In model (1), the BLUP of y_0 and the SPP of Ey_0 are respectively*

$$\begin{aligned} \widetilde{y}_{0_{BLUP}} &= X_0\widetilde{\beta} + VT^+(y - X\widetilde{\beta}), \quad \textit{and} \\ \widetilde{y}_{0_{SPP}} &= X_0\widetilde{\beta}, \end{aligned}$$

where $T = \Sigma + XX'$ and $\widetilde{\beta} = (X'T^+X)^-X'T^+y$ is the best linear unbiased estimator (BLUE) of β in model (1).

If $\Sigma > 0$ and $rk(X) = p$ in model (1), the BLUP of y_0 and the SPP of Ey_0 are respectively

$$\begin{aligned} \hat{y}_{0_{BLUP}} &= X_0\hat{\beta}_{BLUE} + V\Sigma^{-1}(y - X\hat{\beta}_{BLUE}), \quad \textit{and} \\ \hat{y}_{0_{SPP}} &= X_0\hat{\beta}_{BLUE}, \end{aligned}$$

where $\hat{\beta}_{BLUE} = (X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}y$ is the BLUE of β.

Proof. BLUPs of y_0 in Lemma 2.5 were derived by [1] and [28]. The SPPs of $\mathrm{E}y_0$ were derived by [9]. □

The BLUPs and SPPs are presented here for further comparisons.

2.1 The best linear unbiased predictor of δ

Theorem 2.6. *In model (1), the BLUP of δ in $\mathcal{LH}$ is*

$$\hat{\delta}_{BLUP} = X_0\widetilde{\beta} + \lambda VT^+(y - X\widetilde{\beta}),$$

where $T = \Sigma + XX'$, $\widetilde{\beta} = (X'T^+X)^-X'T^+y$.

Proof. Suppose $\hat{\delta} = Cy \in \mathcal{LH}$ and is unbiased, then by Lemma 2.3, $CX = X_0$. Denote $\mathrm{R}(\hat{\delta};\beta)$ as the risk of $\hat{\delta}$ and $\mathrm{tr}(A)$ as the trace of squared matrix A, we have

$$\mathrm{R}(\hat{\delta};\beta) = \mathrm{E}[(\hat{\delta} - \delta)'(\hat{\delta} - \delta)]$$

$$
\begin{aligned}
&=\mathrm{E}[Cy-\lambda y_0-(1-\lambda)X_0\beta]'[Cy-\lambda y_0-(1-\lambda)X_0\beta]\\
&=\mathrm{E}(Cy)'(Cy)+\lambda\mathrm{E}(Cy)'(Cy_0)-(1-\lambda)\mathrm{E}(Cy)'X_0\beta-\lambda\mathrm{E}(y_0)'(Cy)+\lambda^2\mathrm{E}y_0'y_0\\
&\quad+\lambda(1-\lambda)\mathrm{E}y_0'X_0\beta-(1-\lambda)\mathrm{E}(X_0\beta)'Cy+\lambda(1-\lambda)\mathrm{E}(X_0\beta)'y_0+(1-\lambda)^2\mathrm{E}(X_0\beta)'X_0\beta\\
&=\mathrm{tr}(C\Sigma C')+\lambda^2\mathrm{tr}\Sigma_0-2\lambda\mathrm{tr}(CV')+\beta'(CX-X_0)'(CX-X_0)\beta.
\end{aligned}
$$

Minimizing $\mathrm{R}(\hat{\delta};\beta)$ is equivalent to solve the following optimization problem to obtain C such that

$$
\begin{cases}
\arg\min\left[\mathrm{tr}(C\Sigma C')+\lambda\mathrm{tr}\Sigma_0-2\lambda\mathrm{tr}(CV')\right]=C\\
CX-X_0=0.
\end{cases}
$$

Let Λ be a $p\times m$ Lagrange multiplier and construct the Lagrange function as

$$
L(C,\Lambda)=\mathrm{tr}(C\Sigma C')+\lambda\mathrm{tr}\Sigma_0-2\lambda\mathrm{tr}(CV')+2\mathrm{tr}[(CX-X_0)\Lambda].
$$

Let $\partial L/\partial C=0$ and $\partial L/\partial\Lambda=0$, we have

$$
\begin{cases}
C\Sigma-\lambda V+\Lambda'X'=0,\\
X'C'=X_0',
\end{cases}
$$

namely

$$
\begin{pmatrix}\Sigma & X\\ X' & 0\end{pmatrix}\begin{pmatrix}C'\\ \Lambda\end{pmatrix}=\begin{pmatrix}\lambda V'\\ X_0'\end{pmatrix},\tag{3}
$$

and

$$
\begin{pmatrix}C'\\ \Lambda\end{pmatrix}=\begin{pmatrix}\Sigma & X\\ X' & 0\end{pmatrix}^{-}\begin{pmatrix}\lambda V'\\ X_0'\end{pmatrix}.
$$

By Lemma 2.4, we obtain $C=X_0(X'T^{+}X)^{-}X'T^{+}+\lambda VT^{+}(I-X(X'T^{+}X)^{-}X'T^{+})$. Let $\widetilde{\beta}=(X'T^{+}X)^{-}X'T^{+}y$, thus $\widetilde{\delta}_{BLUP}=Cy=X_0\widetilde{\beta}+\lambda VT^{+}(y-X\widetilde{\beta})$. □

Corollary 2.7. *If $\Sigma>0$ and $\mathrm{rk}(X)=p$ in model (1), then the BLUP of δ is*

$$
\hat{\delta}_{BLUP}=X_0\hat{\beta}_{BLUE}+\lambda V\Sigma^{-1}(y-X\hat{\beta}_{BLUE}),
$$

where $\hat{\beta}_{BLUE}=(X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}y$.

Proof. If $\Sigma>0$ and $\mathrm{rk}(X)=p$, then $X'\Sigma^{-1}X$ is nonsingular. Since

$$
\begin{vmatrix}\Sigma & X\\ X' & 0\end{vmatrix}=\begin{vmatrix}\Sigma & X\\ 0 & -X'\Sigma^{-1}X\end{vmatrix}=-|\Sigma||X'\Sigma^{-1}X|\neq 0,
$$

then $\begin{pmatrix}\Sigma & X\\ X' & 0\end{pmatrix}$ is nonsingular. By Lemma 2.4,

$$
\begin{pmatrix}\Sigma & X\\ X' & 0\end{pmatrix}^{-1}=\begin{pmatrix}\Sigma^{-1}-\Sigma^{-1}X(X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1} & \Sigma^{-1}X(X'\Sigma^{-1}X)^{-1}\\ (X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1} & -(X'\Sigma^{-1}X)^{-1}\end{pmatrix}.
$$

With similar calculations as in the proof of Theorem 2.6, the solution of (3) gives that

$$
C=X_0(X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}+\lambda V\Sigma^{-1}(I-X(X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}),
$$

and therefore $\hat{\delta}_{BLUP}=X_0\hat{\beta}_{BLUE}+\lambda V\Sigma^{-1}(y-X\hat{\beta}_{BLUE})$. □

Theorem 2.8. *For the prediction of (2) in model (1), $E\hat{\delta}_{BLUP}=E\widetilde{y}_{0_{BLUP}}=\widetilde{y}_{0_{SPP}}=Ey_0=X_0\beta$.*

Proof. By Theorem 2.6, $\mathrm{E}\hat{\delta}_{BLUP}=\mathrm{E}[X_0\widetilde{\beta}+\lambda VT^{+}(y-X\widetilde{\beta})]=X_0\beta=\mathrm{E}y_0$. From Lemma 2.5, it is easy to prove that $\mathrm{E}\widetilde{y}_{0_{BLUP}}=\widetilde{y}_{0_{SPP}}=\mathrm{E}y_0=X_0\beta$. □

Remark 2.9. *According to Definition 2.1 and Theorem 2.8, $\hat{\delta}_{BLUP}$, $\widetilde{y}_{0_{BLUP}}$ and $\widetilde{y}_{0_{SPP}}$ are all unbiased predictors of y_0 or Ey_0. Let $\lambda = 1$, $\hat{\delta}_{BLUP} = \widetilde{y}_{0_{BLUP}}$ is the BLUP of y_0; Let $\lambda = 0$, $\hat{\delta}_{BLUP} = X_0\widetilde{\beta}$ is the SPP of Ey_0. It shows that the function (2) can simultaneously predict the actual value of y_0 and its mean value. Since $\hat{\delta}_{BLUP} = \lambda\widetilde{y}_{0_{BLUP}} + (1-\lambda)\widetilde{y}_{0_{SPP}}$, then $\hat{\delta}_{BLUP}$ can be viewed as a tradeoff between the BLUP of y_0 and the SPP of Ey_0. By using $\hat{\delta}_{BLUP}$ in practical applications, forecasters can provide a more comprehensive predictor by assigning different weights in $\hat{\delta}_{BLUP}$.*

As for the choice of λ, usually the weight scalar should be given before predicting. Since λ represents the weight to the prediction of y_0 and is not a parameter, then there is no “true" but suitable value of it. One method to select λ is by forecasters' subjective preferences. For example, if the prediction of y_0 and Ey_0 are treated equally, then $\lambda = 0.5$. Another method to determine λ is by using observed data of (y, X) in model (1). In this paper we recommend to use the *leave-one-out cross-validation* technique. In order to determine λ, we take $\hat{\delta}_{BLUP}$ as the predictor of y_0 by Theorem 2.8 since the true β in $Ey_0 = X_0\beta$ is unknown. Define $\hat{\delta}_{(-j)}(\lambda)$ to be the predictor of y_j when the jth case of (y, X) in (1) is deleted. Denote $\mathcal{T} = \{\lambda_i | 0 \le \lambda_i \le 1, i = 1, 2, \cdots\}$. The predicted residual sum of squares is defined as

$$\mathrm{CV}(\lambda) = \sum_{j=1}^{n}[y_j - \hat{\delta}_{(-j)}(\lambda)]^2.$$

For each $\lambda_i \in \mathcal{T}$, compute $\sum_{j=1}^{n}[y_j - \hat{\delta}_{(-j)}(\lambda_i)]^2$. The choice of λ is the one that minimizes $\mathrm{CV}(\lambda)$ over $\mathcal{T}$. Simulations in Section 3 indicate the *leave-one-out cross-validation* technique for the selection of λ is feasible. Forecasters can determine which one of $\hat{\delta}_{BLUP}$, $\widetilde{y}_{0_{BLUP}}$ and $\widetilde{y}_{0_{SPP}}$ is more “suitable" to be afforded through the selection of λ by observed data.

2.2 Efficiency of $\hat{\delta}_{BLUP}$

According to Theorem 2.8, $\hat{\delta}_{BLUP}$, $\widetilde{y}_{0_{BLUP}}$ and $\widetilde{y}_{0_{SPP}}$ are all unbiased predictors of y_0 or Ey_0. From the point of view of the linearity and unbiasedness of the prediction, we mainly discuss the performance of $\hat{\delta}_{BLUP}$ comparing to $\widetilde{y}_{0_{BLUP}}$ and $\widetilde{y}_{0_{SPP}}$ in what follows.

Theorem 2.10. *For model (1),*

$$Cov(\hat{\delta}_{BLUP}) \le Cov(\widetilde{y}_{0_{BLUP}}),$$

and the equality holds if and only if $(1-\lambda^2)VT^+[I - T^+X(X'T^+X)^-X']T^+V' = 0$.

Proof. Denote $\widetilde{\varepsilon}_0 = \lambda VT^+(y - X\widetilde{\beta})$ as the predictor of ε_0, we have

$$\begin{aligned}\mathrm{Cov}(\hat{\delta}_{BLUP}) =& \mathrm{Cov}(X_0\widetilde{\beta} + \lambda\widetilde{\varepsilon}_0),\\ \mathrm{Cov}(\widetilde{y}_{0_{BLUP}}) =& \mathrm{Cov}(X_0\widetilde{\beta} + \widetilde{\varepsilon}_0).\end{aligned}$$

Since $\Sigma = T - XX'$ and $X'[I - T^+X(X'T^+X)^-X'] = 0$, then

$$\begin{aligned}\mathrm{Cov}(X_0\widetilde{\beta}, \widetilde{\varepsilon}_0) =& X_0(X'T^+X)^-X'T^+\Sigma[I - T^+X(X'T^+X)^-X']T^+V'\\ =& X_0(X'T^+X)^-X'T^+(T - XX')[I - T^+X(X'T^+X)^-X']T^+V'\\ =& 0.\end{aligned}$$

Therefore, $\mathrm{Cov}(\hat{\delta}_{BLUP}) - \mathrm{Cov}(\widetilde{y}_{0_{BLUP}}) = (1-\lambda^2)\mathrm{Cov}(\widetilde{\varepsilon}_0) \le 0$, and

$$\mathrm{Cov}(\hat{\delta}_{BLUP}) \le \mathrm{Cov}(\widetilde{y}_{0_{BLUP}}),$$

and the equality holds if and only if $(1-\lambda^2)\mathrm{Cov}(\widetilde{\varepsilon}_0) = (1-\lambda^2)VT^+[I - T^+X(X'T^+X)^-X']T^+V' = 0$. □

Corollary 2.11. *If $\Sigma > 0$ and $rk(X) = p$ in model (1), then*

$$Cov(\hat{\delta}_{BLUP}) \le Cov(\hat{y}_{0_{BLUP}}),$$

and the equality holds if and only if $(1-\lambda^2)V\Sigma^{-1}[I - \Sigma^{-1}X(X'\Sigma^{-1}X)^{-1}X']\Sigma^{-1}V' = 0$.

Proof. Corollary 2.11 is easily proved by Lemma 2.4 and Theorem 2.10. □

Remark 2.12. *Theorem 2.10 and Corollary 2.11 show that $\hat{\delta}_{BLUP}$ is better than $\widetilde{y}_{0_{BLUP}}$ under the criterion of covariance.*

Theorem 2.13. *For model (1), if $DT^+V'X_0(X'T^+X)^-X'T^+ + T^+X(X'T^+X)^-X_0'VT^+D \geq 0$, where $D = I - X(X'T^+X)^-X'T^+$, then*

$$E(\widetilde{y}_{0_{SPP}} - X_0\beta)'(\widetilde{y}_{0_{SPP}} - X_0\beta) \leq E(\hat{\delta}_{BLUP} - X_0\beta)'(\hat{\delta}_{BLUP} - X_0\beta) \leq E(\widetilde{y}_{0_{BLUP}} - X_0\beta)'(\widetilde{y}_{0_{BLUP}} - X_0\beta).$$

Proof. Denote

$$\begin{aligned} C_1 &= X_0(X'T^+X)^-X'T^+ + \lambda VT^+[I - X(X'T^+X)^{-1}X'T^+], \\ C_2 &= X_0(X'T^+X)^-X'T^+ + VT^+[I - X(X'T^+X)^{-1}X'T^+], \end{aligned}$$

then $\hat{\delta}_{BLUP} = C_1y$ and $\widetilde{y}_{0_{BLUP}} = C_2y$. By the unbiasedness, $C_1X = X_0$ and $C_2X = X_0$. Therefore,

$$\begin{aligned} &\mathrm{E}(\hat{\delta}_{BLUP} - X_0\beta)'(\hat{\delta}_{BLUP} - X_0\beta) - \mathrm{E}(\widetilde{y}_{0_{BLUP}} - X_0\beta)'(\widetilde{y}_{0_{BLUP}} - X_0\beta) \\ =&(X\beta)'(C_1'C_1 - C_2'C_2)X\beta + \mathrm{tr}(C_1\Sigma C_1' - C_2\Sigma C_2'). \end{aligned} \tag{4}$$

Note that D is a symmetric idempotent matrix and

$$\begin{aligned} C_1\Sigma C_1' =&C_1(T - XX')C_1' \\ =&X_0(X'T^+X)^-X_0' + \lambda^2VT^+DV' - X_0X_0', \\ C_2\Sigma C_2' =&C_2(T - XX')C_2' \\ =&X_0(X'T^+X)^-X_0' + VT^+DV' - X_0X_0', \end{aligned}$$

then we have

$$\begin{aligned} &C_1\Sigma C_1' - C_2\Sigma C_2' = -(1 - \lambda^2)VT^+DV' \leq 0, \text{ and} \\ &\mathrm{tr}(C_1\Sigma C_1' - C_2\Sigma C_2') \leq 0. \end{aligned} \tag{5}$$

Besides,

$$\begin{aligned} &C_1'C_1 - C_2'C_2 \\ =&(\lambda - 1)[DT^+V'X_0(X'T^+X)^-X'T^+ + T^+X(X'T^+X)^-X_0'VT^+D] \\ &+ (\lambda^2 - 1)DT^+V'VT^+D \\ \leq&(\lambda - 1)[DT^+V'X_0(X'T^+X)^-X'T^+ + T^+X(X'T^+X)^-X_0'VT^+D] \\ \leq&0. \end{aligned} \tag{6}$$

Substituting (5) and (6) into (4), we have

$$\mathrm{E}(\hat{\delta}_{BLUP} - X_0\beta)'(\hat{\delta}_{BLUP} - X_0\beta) \leq \mathrm{E}(\widetilde{y}_{0_{BLUP}} - X_0\beta)'(\widetilde{y}_{0_{BLUP}} - X_0\beta).$$

Let $\lambda = 0$ in (2), then $\widetilde{y}_{0_{SPP}} = X_0\widetilde{\beta} = \arg\min_{\hat{y}_0 \in \mathcal{L}\mathcal{J}} \mathrm{E}(\hat{y}_0 - X_0\beta)'(\hat{y}_0 - X_0\beta)$ by Theorem 2.6. It is obvious that

$$\mathrm{E}(\widetilde{y}_{0_{SPP}} - X_0\beta)'(\widetilde{y}_{0_{SPP}} - X_0\beta) \leq \mathrm{E}(\hat{\delta}_{BLUP} - X_0\beta)'(\hat{\delta}_{BLUP} - X_0\beta).$$ □

By Lemma 2.4 and Theorem 2.13, we have

Corollary 2.14. *In model (1), if $\Sigma > 0$, $rk(X) = p$ and $D\Sigma^{-1}V'X_0(X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1} + \Sigma^{-1}X(X'\Sigma^{-1}X)^{-1}X_0'V\Sigma^{-1}D \geq 0$, where $D = I - X(X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}$, then*

$$E(\hat{y}_{0_{SPP}} - X_0\beta)'(\hat{y}_{0_{SPP}} - X_0\beta) \leq E(\hat{\delta}_{BLUP} - X_0\beta)'(\hat{\delta}_{BLUP} - X_0\beta) \leq E(\hat{y}_{0_{BLUP}} - X_0\beta)'(\hat{y}_{0_{BLUP}} - X_0\beta).$$

Remark 2.15. *Theorem 2.13 and Corollary 2.14 show that $\hat{\delta}_{BLUP}$ is better than $\widetilde{y}_{0_{BLUP}}$ under the squared loss function as the predictor of Ey_0.*

Theorem 2.16. *For model (1),*

$$E(\widetilde{y}_{0_{BLUP}} - y_0)'(\widetilde{y}_{0_{BLUP}} - y_0) \le E(\hat{\delta}_{BLUP} - y_0)'(\hat{\delta}_{BLUP} - y_0) \le E(\widetilde{y}_{0_{SPP}} - y_0)'(\widetilde{y}_{0_{SPP}} - y_0).$$

Proof. Denote

$$\begin{aligned} C_1 &= X_0(X'T^+X)^-X'T^+ + \lambda VT^+[I - X(X'T^+X)^-X'T^+], \\ C_2 &= X_0(X'T^+X)^-X'T^+ + VT^+[I - X(X'T^+X)^-X'T^+], \\ C_3 &= X_0(X'T^+X)^-X'T^+, \end{aligned}$$

then $\hat{\delta}_{BLUP} = C_1y$, $\widetilde{y}_{0_{BLUP}} = C_2y$ and $\widetilde{y}_{0_{SPP}} = X_0\widetilde{\beta} = C_3y$. By Lemma 2.3, $C_1X = X_0$, $C_2X = X_0$ and $C_3X = X_0$. Since

$$\begin{aligned} &\mathrm{E}(C_iy - y_0)'(C_iy - y_0) = \mathrm{tr}C_i\Sigma C_i' - 2\mathrm{tr}(C_iV') + \mathrm{tr}\Sigma_0, \\ &\mathrm{E}(C_iy - y_0)'(C_iy - y_0) - \mathrm{E}(C_jy - y_0)'(C_jy - y_0) \\ =&\mathrm{tr}(C_i\Sigma C_i' - C_j\Sigma C_j') - 2\mathrm{tr}(C_i - C_j)V', 1 \le i, j \le 3, \text{ and } 0 \le \lambda \le 1, \end{aligned}$$

we have

$$\begin{aligned} &\mathrm{E}(C_1y - y_0)'(C_1y - y_0) - \mathrm{E}(C_2y - y_0)'(C_2y - y_0) = (\lambda - 1)^2\mathrm{tr}VT^+DV' \ge 0, \\ &\mathrm{E}(C_1y - y_0)'(C_1y - y_0) - \mathrm{E}(C_3y - y_0)'(C_3y - y_0) = [(\lambda - 1)^2 - 1]\mathrm{tr}VT^+DV' \le 0, \end{aligned}$$

which give that

$$\mathrm{E}(\widetilde{y}_{0_{BLUP}} - y_0)'(\widetilde{y}_{0_{BLUP}} - y_0) \le E(\hat{\delta}_{BLUP} - y_0)'(\hat{\delta}_{BLUP} - y_0) \le \mathrm{E}(\widetilde{y}_{0_{SPP}} - y_0)'(\widetilde{y}_{0_{SPP}} - y_0). \qquad \square$$

By Lemma 2.4 and Theorem 2.16, we have

Corollary 2.17. *In model (1), if $\Sigma > 0$ and $rk(X) = p$, then*

$$E(\hat{y}_{0_{BLUP}} - y_0)'(\hat{y}_{0_{BLUP}} - y_0) \le E(\hat{\delta}_{BLUP} - y_0)'(\hat{\delta}_{BLUP} - y_0) \le E(\hat{y}_{0_{SPP}} - y_0)'(\hat{y}_{0_{SPP}} - y_0).$$

Remark 2.18. *Theorem 2.16 and Corollary 2.17 show that $\hat{\delta}_{BLUP}$ is better than $\widetilde{y}_{0_{SPP}}$ under the squared loss function as the predictor of y_0.*

3 Simulation studies

In this section, we conduct simulations to illustrate the selection of λ in $\hat{\delta}_{0_{BLUP}}$ and the finite sample performance of our simultaneous prediction comparing to $\hat{y}_{0_{BLUP}}$ and $\hat{y}_{0_{SPP}}$.

The data are generated from the following model:

$$\begin{pmatrix} y \\ y_0 \end{pmatrix} = \begin{pmatrix} X \\ X_0 \end{pmatrix}\beta + \begin{pmatrix} \varepsilon \\ \varepsilon_0 \end{pmatrix}, \quad \begin{pmatrix} \varepsilon \\ \varepsilon_0 \end{pmatrix} \sim \mathrm{N}(0, \Sigma), \tag{7}$$

where $\Sigma = \begin{pmatrix} 50 & 2 & \cdots & 2 \\ 2 & 50 & \cdots & 2 \\ \vdots & \vdots & \ddots & \vdots \\ 2 & 2 & \cdots & 50 \end{pmatrix}$.

We assume y is the observation with sample size $n = 200$ and y_0 is to be predicted with sample size $m = 1$. In Section 3.1 we only need the sample data of y to determine λ, while in Section 3.2 we use all the sample data of y and y_0 for comparison with various λ. Elements in corresponding matrices X and X_0 are generated from the Uniform distribution $[1.1, 30.7]$.

3.1 Selection of λ in $\hat{\delta}_{BLUP}$

We set β to be the one-dimensional parameter with the true value 0.8. The number of simulated realizations for choosing λ is 1000. In each simulation, let λ vary from 0 to 1 with step size 0.001. We use the *leave-one-out cross-validation* technique (see Section 2.1) to determine λ. Let λ^* be the selected value of λ, then

$$\lambda^* = \arg\min\ CV(\lambda) = \arg\min \sum_{j=1}^{200} [y_j - \hat{\delta}_{-j}(\lambda)]^2, 0 \le \lambda \le 1.$$

Simulations show that the relationship between $\mathrm{CV}(\lambda)$ and λ is varying. Three of the simulations are presented to illustrate the relation between λ and $\log CV(\lambda)$ in Figure 1. Subfigure (a) tells that $\lambda = 1$ and $\hat{y}_{0_{BLUP}}$ should be provided when predicting; (b) tells that $\lambda = 0$ and $\hat{y}_{0_{SPP}}$ should be preferred; (c) tells that $\lambda = 0.315$ and $\hat{\delta}_{BLUP}$ should be provided when predicting. The relationship between $\mathrm{CV}(\lambda)$ and λ also tells us that there are three kinds of λ^* in our simulations. Table 1 shows that among 1000 simulations, 267 of them give that $\lambda = 0$, 332 of them determine $\lambda = 1$ and 401 of them give that $0 < \lambda < 1$. Simulation performance shows that the *leave-one-out cross-validation* technique for the selection of λ is feasible and give the way to solve the question " which one of $\hat{\delta}_{BLUP}$, $\hat{y}_{0_{BLUP}}$ and $\hat{y}_{0_{SPP}}$ is preferred from the observations ".

Fig. 1. Relationships between λ and $\log[\mathrm{CV}(\lambda)]$ in three simulations (a),(b) and (c) and the corresponding selection of λ

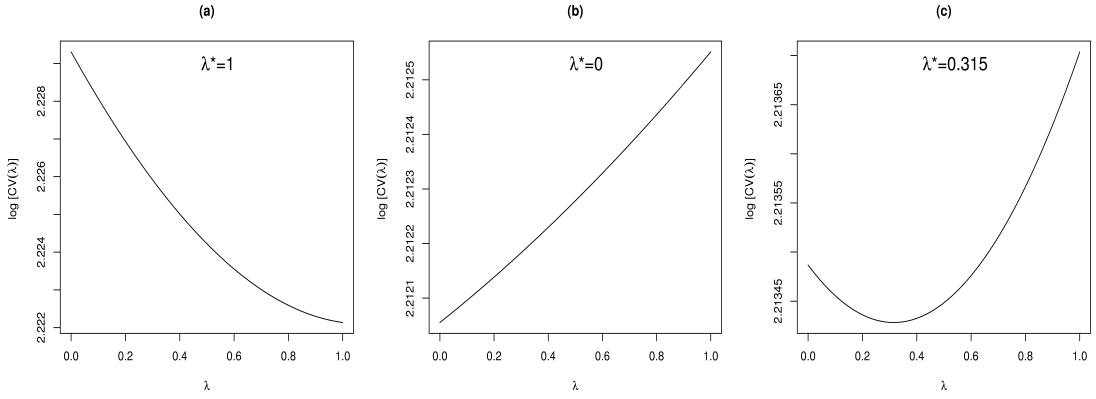

Table 1. Frequency of occurrences of three kinds of λ^* in 1000 simulations

	$\lambda^* = 0$	$\lambda^* = 1$	$0 < \lambda^* < 1$
Frequency	$\frac{267}{1000}$	$\frac{332}{1000}$	$\frac{401}{1000}$

3.2 Finite sample performance of the predictors

Let $n = 200$, $m = 1$, $p = 3$ and the true $\beta = (1, 0.8, 0.2)'$ in (7). λ in $\hat{\delta}_{BLUP}$ varies on a grid from 0.1 to 0.9. For each λ, the number of simulations is 1000. In each simulation, we make some comparisons about $\hat{\delta}_{BLUP}$, $\hat{y}_{0_{BLUP}}$ and $\hat{y}_{0_{SPP}}$. Regarding $\hat{\delta}_{BLUP} - y_0$, $\hat{y}_{0_{BLUP}} - y_0$ and $\hat{y}_{0_{SPP}} - y_0$, the sample means (**sm**s), the standard deviations

(**std**s) and the mean squares (**ms**s) of which are obtained in Table 2. Also, regarding $\hat{\delta}_{BLUP} - X_0\beta$, $\hat{y}_{0_{BLUP}} - X_0\beta$ and $\hat{y}_{0_{SPP}} - X_0\beta$, the **sm**s, the **std**s and the **ms**s of which are presented in Table 3.

From Table 2 and Table 3, we make the following observations:

(1) As for the prediction precision, no matter what λ is set to be, the sample means (**sm**s) of these prediction error of $\hat{y}_{0_{BLUP}}$, $\hat{\delta}_{BLUP}$ and $\hat{y}_{0_{SPP}}$ are all small. Comparisons of **sm**s can not tell which one of the three predictors is better, yet the standard deviations (**std**s) and the mean squares (**ms**s) of $\hat{\delta}_{BLUP} - y_0$ are less than that of $\hat{y}_{0_{SPP}} - y_0$.

(2) No matter what λ is set to be, the sample means (**sm**s) of $\hat{y}_{0_{BLUP}} - X_0\beta$, $\hat{\delta}_{BLUP} - X_0\beta$ and $\hat{y}_{0_{SPP}} - X_0\beta$ are all small. Comparisons of **sm**s can not determine which predictor is better, yet the standard deviations (**std**s) and the mean squares (**ms**s) of $\hat{\delta}_{BLUP} - X_0\beta$ are less than that of $\hat{y}_{0_{BLUP}} - X_0\beta$.

The above facts imply that for any $\lambda \in (0, 1)$, $\hat{\delta}_{BLUP}$, $\hat{y}_{0_{BLUP}}$ and $\hat{y}_{0_{SPP}}$ are all unbiased predictions of y_0 and Ey_0. $\hat{\delta}_{BLUP}$ is more efficient than $X_0\hat{\beta}_{BLUE}$ when predicting the actual value, and is more efficient than $\hat{y}_{0_{BLUP}}$ when predicting the mean value. Simulation performances verify the results in Section 2.2.

Table 2. Finite sample performance about forecast precision of $\hat{y}_{0_{BLUP}}$, $\hat{\delta}_{BLUP}$(with different λ) and $\hat{y}_{0_{SPP}}$

		$\lambda = 0.1$	$\lambda = 0.2$	$\lambda = 0.3$	$\lambda = 0.4$	$\lambda = 0.5$	$\lambda = 0.6$	$\lambda = 0.7$	$\lambda = 0.8$	$\lambda = 0.9$
sm	$\hat{y}_{0_{BLUP}} - y_0$	-0.2596	0.1509	-0.1056	0.3124	0.4998	-0.0703	0.1251	0.2432	0.2150
	$\hat{\delta}_{BLUP} - y_0$	0.2524	0.1790	-0.0662	0.3314	0.4911	-0.0783	0.1292	0.2358	0.2152
	$\hat{y}_{0_{SPP}} - y_0$	0.2516	0.1861	-0.0494	0.3341	0.4825	-0.0904	0.1389	0.2060	0.2172
std	$\hat{y}_{0_{BLUP}} - y_0$	7.2516	7.1930	6.9865	6.8545	6.9253	6.9844	6.8622	6.9193	6.9606
	$\hat{\delta}_{BLUP} - y_0$	7.2758	7.2239	7.0448	6.8462	6.9624	6.9791	6.8682	6.9277	6.9640
	$\hat{y}_{0_{SPP}} - y_0$	7.2843	7.2433	7.0890	6.8656	7.0298	7.0051	6.9230	7.0053	7.0463
ms	$\hat{y}_{0_{BLUP}} - y_0$	52.601	51.711	48.773	47.035	48.162	48.738	47.058	47.888	48.448
	$\hat{\delta}_{BLUP} - y_0$	52.948	52.163	49.584	46.933	48.668	48.665	47.141	48.000	48.496
	$\hat{y}_{0_{SPP}} - y_0$	53.072	52.447	50.206	47.207	49.601	49.030	47.899	49.068	49.648

4 Conclusion

In this paper, we study the prediction based on a composite target function that allows to simultaneously predict the actual and the mean values of the unobserved regressand in the generalized linear model. The BLUP of the target function is derived when the model error covariance is positive semi-definite. The BLUP is also the unbiased prediction of the actual and the mean values of the the unobserved regressand. We propose the *leave-one-out cross-validation* technique to determine the value of the weight scalar in our prediction, which can help to provide a suitable prediction. For the efficiency of the proposed BLUP, studies show that it is better than the usual BLUP under the criterion of covariance and dominates it as a prediction of the mean value of the regressand. Besides, the proposed BLUP is better than the SPP as a prediction of the actual value of the regressand. Simulation studies illustrate the selection of the weight scalar in the proposed BLUP and show that it has better finite sample performance. Further researches on simultaneous prediction are in progress.

Table 3. Finite sample performance about goodness fit of the model of $\hat{y}_{0_{BLUP}}$, $\hat{\delta}_{BLUP}$(with different λ) and $\hat{y}_{0_{SPP}}$

		$\lambda = 0.1$	$\lambda = 0.2$	$\lambda = 0.3$	$\lambda = 0.4$	$\lambda = 0.5$	$\lambda = 0.6$	$\lambda = 0.7$	$\lambda = 0.8$	$\lambda = 0.9$
sm	$\hat{y}_{0_{BLUP}}-X_0\beta$	0.0249	-0.0190	-0.0742	-0.0077	-0.0256	-0.0257	0.0047	-0.0543	-0.0340
	$\hat{\delta}_{BLUP}-X_0\beta$	0.0177	0.0091	-0.0349	0.0113	-0.0343	-0.0337	0.0089	-0.0618	-0.0338
	$\hat{y}_{0_{SPP}}-X_0\beta$	0.0169	0.0162	-0.0180	0.0240	-0.0429	0.0457	0.0186	-0.0915	-0.0318
std	$\hat{y}_{0_{BLUP}}-X_0\beta$	1.6389	1.6769	1.6415	1.6124	1.6121	1.6640	1.6401	1.5242	1.6445
	$\hat{\delta}_{BLUP}-X_0\beta$	1.3389	1.3831	1.3844	1.3774	1.3914	1.5010	1.5048	1.4389	1.5966
	$\hat{y}_{0_{SPP}}-X_0\beta$	1.3334	1.3629	1.3626	1.3308	1.3039	1.3983	1.3547	1.2949	1.3700
ms	$\hat{y}_{0_{BLUP}}-X_0\beta$	2.6841	2.8097	2.6974	2.5973	2.5968	2.7668	2.6872	2.3239	2.7028
	$\hat{\delta}_{BLUP}-X_0\beta$	1.7910	1.9112	1.9159	1.8955	1.9352	2.2518	2.2621	2.0721	2.5477
	$\hat{y}_{0_{SPP}}-X_0\beta$	1.7765	1.8558	1.8551	1.7697	1.7002	1.9554	1.8336	1.6835	1.8761

Acknowledgement: The authors are grateful to the responsible editor and the anonymous reviewers for their valuable comments and suggestions, which have greatly improved this paper. This research is supported by the Scientific Research Fund of Hunan Provincial Education Department (13C1139), the Youth Scientific Research Foundation of Central South University of Forestry and Technology of China (QJ2012013A) and the Natural Science Foundation of Hunan Province (2015JJ4090).

References

[1] Goldberger A.S., Best linear unbiased prediction in the generalized linear regression model, *Journal of the American Statistical Association*, 1962, 57(298), 369–375
[2] Bolfarine H., Zacks S., Bayes and minimax prediction in finite populations, *J. Statist. Plann. Infer.*, 1991, 28, 139–151
[3] Yu S. H., The linear minimax predictor in finite populations with arbitrary rank under quadratic loss function, *Chin. Ann. Math.*, 2004, 25, 485–496
[4] Xu L. W., Wang S. G., The minimax predictor in finite populations with arbitrary rank in normal distribution, *Chin. Ann. Math.*, 2006, 27, 405–416
[5] Gotway C. A., Cressie N., Improved multivariate prediction under a general linear model, *J. Multivariate Anal.*, 1993, 45, 56–72
[6] P J G Teunissen P. J. G., Best prediction in linear models with mixed integer/real unknowns: theory and application, *Journal of Geodesy*, 2007, 81(12),759–780
[7] Xu L. W., Admissible linear predictors in the superpopulation model with respect to inequality constraints, *Comm. Statist. Theory Methods*, 2009, 38, 2528–2540
[8] Searle S. R., Casella G., McCulloch C. E., Variance components, 1992, New York: Wiley.
[9] Bolfarine H., Rodrigues J., On the simple projection predictor in finite populations, *Australian Journal of Statistics*, 1988, 30(3), 338–341
[10] Hu G. K., Li Q. G., Yu S. H., Optimal and minimax prediction in multivariate normal populations under a balanced loss function, *J. Multivariate Anal.*, 2014, 128, 154–164
[11] Hu G. K., Peng P., Linear admissible predictor of finite population regression coefficient under a balanced loss function, *J. Math.*, 2014, 34, 820–828
[12] Diebold F. X., Lopez J. A., Forecast evaluation and combination, *Handbook of statistics*, 1996, 14, 241–268
[13] Hendry D. F., Clements M. P., Pooling of forecasts, *Econometrics Journal*, 2002, 5, 1–26
[14] Timmermann A., Forecast combinations, *Handbook of economic forecasting*, 2006, 1,135–196
[15] Shalabh, Performance of stein-rule procedure for simultaneous prediction of actual and average values of study variable in linear regression models, *Bull. Internat. Statist. Inst*, 1995, 56, 1357–1390
[16] Chaturvedi A., Singh S. P., Stein rule prediction of the composite target function in a general linear regression model, *Statist. Papers*, 2000, 41(3), 359–367

[17] Chaturvedi A., Kesarwani S., Chandra R., Simultaneous prediction based on shrinkage estimator, *in: Shalabh, C. Heumann (Eds.), Recent Advances in Linear Models and Related Areas, Springer*, 2008, 181–204
[18] Shalabh, Heumann C., Simultaneous prediction of actual and average values of study variable using stein-rule estimators, *in: K. Kumar, A. Chaturvedi (Eds.), Some Recent Developments in Statistical Theory and Application, Brown Walker Press, USA*, 2012, 68–81
[19] Chaturvedi A., Wan A. T. K., Singh S. P., Improved multivariate prediction in a general linear model with an unknown error covariance matrix, *J. Multivariate Anal.*, 2002, 83(1), 166–182
[20] Bai C., Li H., Admissibility of simultaneous prediction for actual and average values in finite population, *J. Inequal. Appl.*, 2018, 2018(1), 117
[21] Toutenburg H., Shalabh, Predictive performance of the methods of restricted and mixed regression estimators, *Biometrical J.*, 1996, 38(8), 951–959
[22] Toutenburg H., Shalabh, Improved predictions in linear regression models with stochastic linear constraints, *Biom. J.*, 2000, 42(1), 71–86
[23] Dubeand M., Manocha V., Simultaneous prediction in restricted regression models, *J. Appl. Statist. Sci.*, 2002, 11(4), 277–288
[24] Shalabh, Paudel C. M., Kumar N., Simultaneous prediction of actual and average values of response variable in replicated measurement error models, *in: Shalabh, C. Heumann (Eds.), Recent Advances in Linear Models and Related Areas, Springer*, 2008, 105–133
[25] Garg G., Shalabh, Simultaneous predictions under exact restrictions in ultrastructural model, *Journal of Statistical Research (in Special Volume on Measurement Error Models)*, 2011, 45(2), 139–154
[26] Shalabh, A revisit to efficient forecasting in linear regression models, *J. Multivariate Anal.*, 2013, 114, 161–170
[27] Wang S. G., Shi J. H., Introduction to the linear model, 2004, Science Press, Beijing.
[28] Yu S. H., Xu L. W., Admissibility of linear prediction under quadratic loss, *Acta Mathematicae Applicatae Sinica*, 2004, 27, 385–396