Chamchuri Journal of Mathematics
Volume 10(2018), 28–52
http://www.math.sc.chula.ac.th/cjm

Variable Selection in Multivariate Linear Regression Models Subject to Sampling Errors

Annop Angkunsit[†] and Jiraphan Suntornchost[*]

Received 26 April 2018
Revised 14 August 2018
Accepted 30 August 2018

Abstract: In 2015, Lahiri and Suntornchost showed that sampling errors could cause bias in the variable selection criteria statistics in univariate linear regression models. In their study, they suggested ways to adjust those variable selection statistics to reduce the biasedness for the Fay-Herriot model when regression error terms are assumed to be independent and identically distributed. In this study, we will extend their methods to adjust the original variable selection criteria for multivariate linear regression model subject to sampling errors. Simulation results show that our proposed variable selection criteria can reduce the approximation errors of the standard variable selection criterion.

Keywords: Variable selection, Multivariate linear regression, Adjusted R^2, Sampling errors

2000 Mathematics Subject Classification: 62F07, 62F12, 62J05, 62H20

[†] *The author is supported by the His Royal Highness Crown Prince Maha Vajiralongkorn Fund.*

[*] *Corresponding author*

1 Introduction

Linear regression is a statistical technique to determine the linear relationship between variable of interest and other observed variables, known as covariates, by examining sample observations. The linear relationship can often be modelled in many different ways attempting to use all covariates or only a subset of covariates when many covariates are measured. There are a variety of variable selection criteria available in literatures such as RMSE (root mean square error), Adjusted R^2, AIC, BIC. In general, these variable selection criteria are used without concerning sampling errors.

In 2015, Lahiri and Suntornchost [7] showed that sampling errors could cause bias in the variable selection criteria statistics in univariate linear regression models. In their study, they suggested ways to adjust those variable selection statistics to reduce the biasedness for the Fay-Herriot model when regression error terms are assumed to be independent and identically distributed. Later in 2017, Lenghoe and Suntornchost [8] extended the methods to derive variable selection criteria statistics for a general case when the regression error terms are not assumed to be independent and identically distributed allowing for the possibility of correlated regression errors.

In real world applications, multivariate linear regression models have brought interests from researchers worldwide due to availabilities of data and technologies and the flexibility in allowing correlations between different variables of interest. There are several applications of multivariate linear regression models, for example, Aleixandre-Tud and Alvarez [1] applied the model to predict wine quality based on the definition of chemical and phenolic parameters of grapes and wines harvested at different ripening levels; Cserhti and Szgyi [5] applied the model to extract of maximal information of large data sets of evaluation of chromatographic retention data measured under different conditions; and Seidou, Asselin and Ouarda [11] applied the model to explain key variables in hydrology and climate sciences.

In multivariate linear regression models, many variables of interest will be included in the model. Therefore, the effect from sampling errors to the regression analysis could be more severe than those of univariate models. Therefore, in this study, we extend the techniques proposed in Lahiri and Suntornchost [7] to adjust variable selection criteria statistics for multivariate linear regression models subject to sampling errors. The organization of this paper is as follows. In section 2, we introduce the multivariate linear regression models and variable selection criteria.

In section 3, we establish some adjustments for variable selection criteria statistics for multivariate linear regression models. Simulation results and discussions are provided in Sections 4 and 5, respectively.

2 The Multivariate Linear Regression Models

The multivariate linear regression considered in this paper is in the form

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

where $\mathbf{Y}$ is an $m \times q$ matrix of response variables, $\mathbf{X}$ is an $m \times p$ matrix of covariates $(m > p)$, $\text{rank}(\mathbf{X}) = p$, $\boldsymbol{\beta}$ is a $p \times q$ matrix of unknown regression coefficients and $\boldsymbol{\varepsilon} = [\, \boldsymbol{\varepsilon}_1\ \boldsymbol{\varepsilon}_2\ \ldots\ \boldsymbol{\varepsilon}_q \,]$ is an $m \times q$ matrix of error terms with mean zero and covariance matrix $\text{Cov}\,(\boldsymbol{\varepsilon}_i) = \sigma^2\mathbf{I}$ for $i = 1, 2, \ldots, q$, $\{\boldsymbol{\varepsilon}_i, i = 1, \ldots, q\}$ are not assumed to be uncorrelated.

Therefore, following [6] and [9], the least squares estimator $\hat{\boldsymbol{\beta}}$ is then given by

$$\hat{\boldsymbol{\beta}} = \left[\, \hat{\boldsymbol{\beta}}_1 \mid \hat{\boldsymbol{\beta}}_2 \mid \ \ldots\ \mid \hat{\boldsymbol{\beta}}_q \,\right] = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\,[\, \mathbf{y}_1 \mid \mathbf{y}_2 \mid \ \ldots\ \mid \mathbf{y}_q \,],$$

or equivalently

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

From model in (1) and the least squares estimator (2), we can express the sum of squared errors and cross product matrix **SSE** and the total sum of squares and cross product matrix **SST** as quadratic forms in $\mathbf{Y}$ as follows

$$\mathbf{SSE} = \mathbf{Y}'(\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')\mathbf{Y},$$
$$\mathbf{SST} = \mathbf{Y}'(\mathbf{I} - m^{-1}\mathbf{J})\mathbf{Y},$$

where $\mathbf{J}$ is an $m \times m$ matrix of ones.

Following [3] and [12] to define variable selection criteria for multivariate linear regression models, we can minimize some quadratic forms of **SSE** and **SST**, for example, the trace, the determinant, or the largest eigenvalue. In this study, we consider the trace function. In particular, the variable selection criterion considered in this study is the Adjusted R^2 defined as

$$Adj\ R^2 = 1 - \frac{\text{tr}(\mathbf{MSE})}{\text{tr}(\mathbf{MST})} = 1 - \frac{\text{tr}(\mathbf{SSE})/(q(m-p))}{\text{tr}(\mathbf{SST})/(q(m-1))},$$

where m is the sample size, q is the number of response variables and p is the number of covariates.

3 Methodology

In this work, we consider the multivariate linear regression model defined as

$$\boldsymbol{\theta} = \mathbf{X}\boldsymbol{\beta} + \mathbf{V},$$

where $\boldsymbol{\theta}$ is an $m \times q$ matrix of q unobserved response variables of m individuals and $\mathbf{V} = [\mathbf{v}_1\ \mathbf{v}_2\ \ldots\ \mathbf{v}_q]$ is the matrix of error terms, where $\mathbf{v}_k \sim N_m(\mathbf{0}, \sigma^2\mathbf{I})$ $(k = 1, \ldots, q)$ are random vectors.

We are interested in the situation where the response variable $\boldsymbol{\theta}$ is not observed but it is approximated by the observed value $\mathbf{Y}$, where the relation between $\boldsymbol{\theta}$ and $\mathbf{Y}$ is described by the sampling model

$$\mathbf{Y} = \boldsymbol{\theta} + \mathbf{E},$$

where $\mathbf{E} = (e_{ij})$ is an $m \times q$ matrix of sampling errors assuming that the $e_{ij} \sim N(0, D_{ij})$ $(i = 1, \ldots, m,\ j = 1, \ldots, q)$ are independent over i, j. Moreover, we assume that $\boldsymbol{\theta}$ and $\mathbf{E}$ are independent.

In many situations, the response variable $\boldsymbol{\theta}_i (i = 1, \ldots, q)$ is replaced by $\mathbf{y}_i$ that ignores sampling errors and use variable selection criteria. This causes unbiased in estimating variable selection criteria statistics which is shown in Lahiri and Suntornchost [7] and Lenghoe and Suntornchost [8] for the univariate linear regression models. In this study, we extend the two studies to examine errors in approximating the true variable selection criterion, *Adj* R^2, by the naive variable selection criterion in presence of sampling errors for multivariate linear regression models. Moreover, we adapt their methods to derive some variable selection criteria for multivariate linear regression models. Following their techniques, we first examine the effect of sampling errors on the naive estimates $\mathrm{tr}(\mathbf{MSE})$ and $\mathrm{tr}(\mathbf{MST})$ conditional on $\boldsymbol{\theta}$.

Theorem 3.1. *The conditional expectations of* $\mathrm{tr}(\mathbf{MSE})$ *and* $\mathrm{tr}(\mathbf{MST})$ *given* $\boldsymbol{\theta}$ *are respectively defined as:*

$$\mathrm{E}\left[\mathrm{tr}(\mathbf{MSE}) | \boldsymbol{\theta}\right] = \mathrm{tr}(\mathbf{MSE}_{\boldsymbol{\theta}}) + \mathrm{tr}(\mathbf{D}_{w1}),$$
$$\mathrm{E}\left[\mathrm{tr}(\mathbf{MST}) | \boldsymbol{\theta}\right] = \mathrm{tr}(\mathbf{MST}_{\boldsymbol{\theta}}) + \mathrm{tr}(\mathbf{D}_{w2}),$$

where $\mathrm{tr}(\mathbf{D}_{w1}) = \dfrac{1}{q(m-p)} \displaystyle\sum_{j=1}^{q}\sum_{i=1}^{m}(1 - \mathbf{x}_i'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_i)D_{ij}$ *and* $\mathrm{tr}(\mathbf{D}_{w2}) = \dfrac{1}{qm}\displaystyle\sum_{j=1}^{q}\sum_{i=1}^{m} D_{ij}$.

Proof. The conditional expectation of **MSE** given $\boldsymbol{\theta}$ can be computed as

$$\begin{aligned}
\mathrm{E}[\mathbf{MSE}|\boldsymbol{\theta}] &= \mathrm{E}\left[\frac{\mathbf{SSE}}{q(m-p)}\Big|\boldsymbol{\theta}\right] \\
&= \frac{1}{q(m-p)}\mathrm{E}[\mathbf{Y}'(\mathbf{I}-\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')\mathbf{Y}|\boldsymbol{\theta}] \\
&= \frac{1}{q(m-p)}\mathrm{E}[(\boldsymbol{\theta}+\mathbf{E})'(\mathbf{I}-\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')(\boldsymbol{\theta}+\mathbf{E})|\boldsymbol{\theta}] \\
&= \frac{1}{q(m-p)}\mathrm{E}[\boldsymbol{\theta}'(\mathbf{I}-\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')\boldsymbol{\theta}+\boldsymbol{\theta}'(\mathbf{I}-\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')\mathbf{E} \\
&\qquad +\mathbf{E}'(\mathbf{I}-\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')\boldsymbol{\theta}+\mathbf{E}'(\mathbf{I}-\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')\mathbf{E}|\boldsymbol{\theta}] \\
&= \frac{1}{q(m-p)}\left[\mathbf{SSE}_{\boldsymbol{\theta}}+\mathrm{E}[\mathbf{E}'(\mathbf{I}-\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')\mathbf{E}]\right] \\
&= \mathbf{MSE}_{\boldsymbol{\theta}}+\frac{1}{q(m-p)}\mathrm{E}\left[\left(\sum_{i=1}^{m}e_{ij}e_{ik}-\sum_{i=1}^{m}\sum_{l=1}^{m}e_{ij}e_{lk}\mathbf{x}_i'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_l\right)_{1\le j,k\le q}\right],
\end{aligned}$$

where we use the facts that $\boldsymbol{\theta}$ and $\mathbf{E}$ are independent, $\mathrm{E}[e_{ij}]=0$ for $i=1,\ldots,m$, $j=1,\ldots,q$ and $\mathbf{E}'(\mathbf{I}-\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')\mathbf{E}=\left(\sum_{i=1}^{m}e_{ij}e_{ik}-\sum_{i=1}^{m}\sum_{l=1}^{m}e_{ij}e_{lk}\mathbf{x}_i'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_l\right)_{1\le j,k\le q}$ to obtain the last two equations.

Consequently, the conditional expectation of $\mathrm{tr}(\mathbf{MSE})$ given $\boldsymbol{\theta}$ can be computed as

$$\begin{aligned}
\mathrm{E}[\mathrm{tr}(\mathbf{MSE})|\boldsymbol{\theta}] &= \mathrm{tr}(\mathbf{MSE}_{\boldsymbol{\theta}}) \\
&\quad+\frac{1}{q(m-p)}\mathrm{E}\left[\mathrm{tr}\left(\left(\sum_{i=1}^{m}e_{ij}e_{ik}-\sum_{i=1}^{m}\sum_{l=1}^{m}e_{ij}e_{lk}\mathbf{x}_i'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_l\right)_{1\le j,k\le q}\right)\right] \\
&= \mathrm{tr}(\mathbf{MSE}_{\boldsymbol{\theta}}) \\
&\quad+\frac{1}{q(m-p)}\sum_{j=1}^{q}\left(\sum_{i=1}^{m}\mathrm{E}\left[e_{ij}^2\right]-\sum_{i=1}^{m}\sum_{l=1}^{m}\mathrm{E}\left[e_{ij}e_{lj}\right]\mathbf{x}_i'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_l\right) \\
&= \mathrm{tr}(\mathbf{MSE}_{\boldsymbol{\theta}})+\frac{1}{q(m-p)}\sum_{j=1}^{q}\sum_{i=1}^{m}(1-\mathbf{x}_i'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_i)\,\mathrm{E}\left[e_{ij}^2\right] \\
&= \mathrm{tr}(\mathbf{MSE}_{\boldsymbol{\theta}})+\frac{1}{q(m-p)}\sum_{j=1}^{q}\sum_{i=1}^{m}(1-\mathbf{x}_i'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_i)D_{ij} \\
&= \mathrm{tr}(\mathbf{MSE}_{\boldsymbol{\theta}})+\mathrm{tr}(\mathbf{D}_{w1}),
\end{aligned}$$

where we use the fact that $\mathrm{E}\left[e_{ij}^2\right]=D_{ij}$ for $i=1,\ldots,m$ and $j=1,\ldots,q$ to

obtain the second to last equation.

Similarly, the conditional expectation of $\mathbf{MST}$ given $\boldsymbol{\theta}$ can be computed as

$$\begin{aligned}
\mathrm{E}[\mathbf{MST}|\boldsymbol{\theta}] &= \mathrm{E}\left[\frac{\mathbf{SST}}{q(m-1)}\Big|\boldsymbol{\theta}\right] \\
&= \frac{1}{q(m-1)}\,\mathrm{E}[\mathbf{Y}'(\mathbf{I}-m^{-1}\mathbf{J})\mathbf{Y}|\boldsymbol{\theta}] \\
&= \frac{1}{q(m-1)}\,\mathrm{E}[(\boldsymbol{\theta}+\mathbf{E})'(\mathbf{I}-m^{-1}\mathbf{J})(\boldsymbol{\theta}+\mathbf{E})|\boldsymbol{\theta}] \\
&= \frac{1}{q(m-1)}\,\mathrm{E}[\boldsymbol{\theta}'(\mathbf{I}-m^{-1}\mathbf{J})\boldsymbol{\theta}+\boldsymbol{\theta}'(\mathbf{I}-m^{-1}\mathbf{J})\mathbf{E}+\mathbf{E}'(\mathbf{I}-m^{-1}\mathbf{J})\boldsymbol{\theta} \\
&\qquad\qquad +\mathbf{E}'(\mathbf{I}-m^{-1}\mathbf{J})\mathbf{E}|\boldsymbol{\theta}] \\
&= \frac{1}{q(m-1)}\left[\mathbf{SST}_{\boldsymbol{\theta}}+\mathrm{E}[\mathbf{E}'(\mathbf{I}-m^{-1}\mathbf{J})\mathbf{E}]\right] \\
&= \mathbf{MST}_{\boldsymbol{\theta}}+\frac{1}{q(m-1)}\,\mathrm{E}\left[\left(\sum_{i=1}^{m}e_{ij}e_{ik}-\frac{1}{m}\left(\sum_{i=1}^{m}e_{ij}\right)\left(\sum_{i=1}^{m}e_{ik}\right)\right)_{1\le j,k\le q}\right],
\end{aligned}$$

where we use the facts that $\boldsymbol{\theta}$ and $\mathbf{E}$ are independent, $\mathrm{E}[e_{ij}]=0$ for $i=1,\ldots,m$, $j=1,\ldots,q$ and $\mathbf{E}'(\mathbf{I}-m^{-1}\mathbf{J})\mathbf{E}=\left(\sum_{i=1}^{m}e_{ij}e_{ik}-\frac{1}{m}\left(\sum_{i=1}^{m}e_{ij}\right)\left(\sum_{i=1}^{m}e_{ik}\right)\right)_{1\le j,k\le q}$ to obtain the last two equations.

Therefore, the conditional expectation of $\mathrm{tr}(\mathbf{MST})$ given $\boldsymbol{\theta}$ can be computed as

$$\begin{aligned}
\mathrm{E}[\mathrm{tr}(\mathbf{MST})|\boldsymbol{\theta}] &= \mathrm{tr}(\mathbf{MST}_{\boldsymbol{\theta}}) \\
&\quad+\frac{1}{q(m-1)}\,\mathrm{E}\left[\mathrm{tr}\left(\left(\sum_{i=1}^{m}e_{ij}e_{ik}-\frac{1}{m}\left(\sum_{i=1}^{m}e_{ij}\right)\left(\sum_{i=1}^{m}e_{ik}\right)\right)_{1\le j,k\le q}\right)\right] \\
&= \mathrm{tr}(\mathbf{MST}_{\boldsymbol{\theta}})+\frac{1}{q(m-1)}\sum_{j=1}^{q}\left(\sum_{i=1}^{m}\mathrm{E}\left[e_{ij}^2\right]-\frac{1}{m}\,\mathrm{E}\left[\left(\sum_{i=1}^{m}e_{ij}\right)^2\right]\right) \\
&= \mathrm{tr}(\mathbf{MST}_{\boldsymbol{\theta}})+\frac{1}{q(m-1)}\sum_{j=1}^{q}\left(\sum_{i=1}^{m}D_{ij}-\frac{1}{m}\sum_{i=1}^{m}D_{ij}\right) \qquad (3) \\
&= \mathrm{tr}(\mathbf{MST}_{\boldsymbol{\theta}})+\frac{1}{q(m-1)}\sum_{j=1}^{q}\left(\frac{m-1}{m}\right)\sum_{i=1}^{qm}D_{ij} \\
&= \mathrm{tr}(\mathbf{MST}_{\boldsymbol{\theta}})+\frac{1}{qm}\sum_{j=1}^{q}\sum_{i=1}^{m}D_{ij} \\
&= \mathrm{tr}(\mathbf{MST}_{\boldsymbol{\theta}})+\mathrm{tr}(\mathbf{D}_{w2}),
\end{aligned}$$

where we use the facts that $\mathrm{E}\left[e_{ij}^2\right] = D_{ij}$ for $i = 1, \ldots, m,\ j = 1, \ldots, q$ and $\mathrm{E}\left[\left(\sum_{i=1}^{m} e_{ij}\right)^2\right] = \sum_{i=1}^{m} D_{ij}$ to obtain (3).
Therefore, the theorem is proved. □

3.1 Variable Selection Criteria based on Unbiased Estimators

Note from Theorem 3.1 that the naive estimators $\mathrm{tr}(\mathbf{MSE})$ and $\mathrm{tr}(\mathbf{MST})$ are biased estimators of $\mathrm{tr}(\mathbf{MSE}_{\boldsymbol{\theta}})$ and $\mathrm{tr}(\mathbf{MST}_{\boldsymbol{\theta}})$, respectively. Therefore, in this paper, following Lahiri and Suntornchost [7], we propose an adjustment of the adjusted R^2 by replacing the naive estimators of $\mathrm{tr}(\mathbf{MSE}_{\boldsymbol{\theta}})$ and $\mathrm{tr}(\mathbf{MST}_{\boldsymbol{\theta}})$ by their unbiased and consistent estimators defined in the following theorem.

Theorem 3.2. *Define* $\mathrm{tr}(\widehat{\mathbf{MSE}}_{\boldsymbol{\theta}})$ *and* $\mathrm{tr}(\widehat{\mathbf{MST}}_{\boldsymbol{\theta}})$ *as*

$$\mathrm{tr}(\widehat{\mathbf{MSE}}_{\boldsymbol{\theta}}) = \mathrm{tr}(\mathbf{MSE}) - \mathrm{tr}(\mathbf{D}_{w1})$$

and

$$\mathrm{tr}(\widehat{\mathbf{MST}}_{\boldsymbol{\theta}}) = \mathrm{tr}(\mathbf{MST}) - \mathrm{tr}(\mathbf{D}_{w2}),$$

respectively, where $\mathrm{tr}(\mathbf{D}_{w1}) = \dfrac{1}{q(m-p)} \sum_{j=1}^{q} \sum_{i=1}^{m} (1 - \mathbf{x}_i'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_i) D_{ij}$ *and* $\mathrm{tr}(\mathbf{D}_{w2}) = \dfrac{1}{qm} \sum_{j=1}^{q} \sum_{i=1}^{m} D_{ij}$. *Then*

1) $\mathrm{tr}(\widehat{\mathbf{MSE}}_{\boldsymbol{\theta}})$ *is an unbiased and consistent estimator of* $\mathrm{tr}(\mathbf{MSE}_{\boldsymbol{\theta}})$, *and*

2) $\mathrm{tr}(\widehat{\mathbf{MST}}_{\boldsymbol{\theta}})$ *is an unbiased and consistent estimator of* $\mathrm{tr}(\mathbf{MST}_{\boldsymbol{\theta}})$.

Proof. The bias of $\mathrm{tr}(\widehat{\mathbf{MSE}}_{\boldsymbol{\theta}})$ given $\boldsymbol{\theta}$ can be computed as

$$\begin{aligned}
\mathrm{E}[\mathrm{tr}(\widehat{\mathbf{MSE}}_{\boldsymbol{\theta}}) - \mathrm{tr}(\mathbf{MSE}_{\boldsymbol{\theta}})|\boldsymbol{\theta}] &= \mathrm{E}[\mathrm{tr}(\mathbf{MSE}) - \mathrm{tr}(\mathbf{D}_{w1})|\boldsymbol{\theta}] - \mathrm{tr}(\mathbf{MSE}_{\boldsymbol{\theta}}) \\
&= \mathrm{E}[\mathrm{tr}(\mathbf{MSE})|\boldsymbol{\theta}] - \mathrm{tr}(\mathbf{D}_{w1}) - \mathrm{tr}(\mathbf{MSE}_{\boldsymbol{\theta}}) \\
&= \mathrm{tr}(\mathbf{MSE}_{\boldsymbol{\theta}}) + \mathrm{tr}(\mathbf{D}_{w1}) - \mathrm{tr}(\mathbf{D}_{w1}) - \mathrm{tr}(\mathbf{MSE}_{\boldsymbol{\theta}}) \\
&= 0.
\end{aligned}$$

To show that $\mathrm{tr}(\widehat{\mathbf{MSE}}_{\boldsymbol{\theta}})$ is consistent, we define, for $k = 1, 2, \ldots, q$,

$$\mathrm{SSE}_k = \mathbf{y}_k'(\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')\mathbf{y}_k,$$

and

$$\mathrm{MSE}_k = \frac{1}{q(m-p)} \mathbf{y}_k'(\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')\mathbf{y}_k.$$

Note that SSE_k/σ^2 is distributed as a chi-square distribution with the degree of freedom of $m-p$. Therefore, the conditional variance of MSE_k given $\boldsymbol{\theta}$ can be computed as

$$\begin{aligned}
\mathrm{Var}[\mathrm{MSE}_k|\boldsymbol{\theta}] &= \mathrm{Var}\left[\frac{\mathrm{SSE}_k}{q(m-p)}\Big|\boldsymbol{\theta}\right] \\
&= \frac{\sigma^4}{(q(m-p))^2}\,\mathrm{Var}[\mathrm{SSE}_k/\sigma^2|\boldsymbol{\theta}] \\
&= \frac{\sigma^4}{(q(m-p))^2}2(m-p) \\
&= \frac{2\sigma^4}{q^2(m-p)} \\
&= O(m^{-1}).
\end{aligned}$$

Consequently, the conditional variance of $\mathrm{tr}(\widehat{\mathbf{MSE}}_{\boldsymbol{\theta}})$ given $\boldsymbol{\theta}$ can be computed as

$$\begin{aligned}
\mathrm{Var}[\mathrm{tr}(\widehat{\mathbf{MSE}}_{\boldsymbol{\theta}})|\boldsymbol{\theta}] &= \mathrm{Var}[\mathrm{tr}(\mathbf{MSE}) - \mathrm{tr}(\mathbf{D}_{w1})|\boldsymbol{\theta}] \\
&= \mathrm{Var}\left[\sum_{k=1}^{q}\mathrm{MSE}_k\Big|\boldsymbol{\theta}\right] \\
&= \sum_{k=1}^{q}\mathrm{Var}[\mathrm{MSE}_k|\boldsymbol{\theta}] + \sum_{\substack{i,j=1\\ i\neq j}}^{q}\mathrm{Cov}[\mathrm{MSE}_i,\mathrm{MSE}_j|\boldsymbol{\theta}] \\
&\leq \sum_{k=1}^{q}\mathrm{Var}[\mathrm{MSE}_k|\boldsymbol{\theta}] + \sum_{\substack{i,j=1\\ i\neq j}}^{q}\sqrt{\mathrm{Var}[\mathrm{MSE}_i|\boldsymbol{\theta}]\,\mathrm{Var}[\mathrm{MSE}_j|\boldsymbol{\theta}]} \\
&= O(m^{-1}).
\end{aligned}$$

Therefore, we can show that $\mathrm{tr}(\widehat{\mathbf{MSE}}_{\boldsymbol{\theta}})$ is an unbiased and consistent estimator of $\mathrm{tr}(\mathbf{MSE}_{\boldsymbol{\theta}})$.

By the same technique, we can show that $\mathrm{tr}(\widehat{\mathbf{MST}}_{\boldsymbol{\theta}})$ is an unbiased and consistent estimator of $\mathrm{tr}(\mathbf{MST}_{\boldsymbol{\theta}})$. □

From the above theorem, we propose the $Adj\ R^2_{hat} = 1 - \dfrac{\mathrm{tr}(\widehat{\mathbf{MSE}}_{\boldsymbol{\theta}})}{\mathrm{tr}(\widehat{\mathbf{MST}}_{\boldsymbol{\theta}})}$ as a variable selection criterion.

3.2 Positive Adjustments to the Unbiased Estimator

From the definitions of $\text{tr}(\widehat{\mathbf{MSE}}_{\boldsymbol{\theta}})$ and $\text{tr}(\widehat{\mathbf{MST}}_{\boldsymbol{\theta}})$ defined in Theorem 3.2, we can see that in some situations, $\text{tr}(\widehat{\mathbf{MSE}}_{\boldsymbol{\theta}})$ and/or $\text{tr}(\widehat{\mathbf{MST}}_{\boldsymbol{\theta}})$ could be negative. In this case, the *Adj* R^2_{hat} may go out of the admissible range if either $\text{tr}(\widehat{\mathbf{MSE}}_{\boldsymbol{\theta}})$ or $\text{tr}(\widehat{\mathbf{MST}}_{\boldsymbol{\theta}})$ is negative. Then the proposed adjusted R^2 may be greater than one.

Therefore, in this section, we propose a positive adjustment to the unbiased estimators obtained in Theorem 3.2 by constructing a positive approximation of $x - y$ defined as follows.

Theorem 3.3. *The g-function defined as*

$$g(x,y) = x + \frac{2x^3\left(1 - \exp\left(\left(\frac{y}{x}\right)^3\right)\right)}{y^2\left(1 + \exp\left(\left(\frac{y}{x}\right)^3\right)\right)}$$

is a positive approximation of $x - y$ when $x, y > 0$ such that $y < \sqrt[3]{\pi}x$.

Proof. We will first show that the g-function is an approximation of $x - y$.

$$\begin{aligned} g(x,y) &= x + \frac{2x^3\left(1 - \exp\left(\left(\frac{y}{x}\right)^3\right)\right)}{y^2\left(1 + \exp\left(\left(\frac{y}{x}\right)^3\right)\right)} \\ &= x - \frac{2x^3}{y^2}\tanh\left(\frac{1}{2}\left(\frac{y}{x}\right)^3\right). \end{aligned} \tag{4}$$

Note that the first order Taylor Polynomial approximation of $\tanh\left(\frac{1}{2}\left(\frac{y}{x}\right)^3\right)$ is $\frac{1}{2}\left(\frac{y}{x}\right)^3$. Applying the approximation to (4), we can show that $g(x,y)$ approximates $x - \left(\frac{2x^3}{y^2}\left(\frac{1}{2}\left(\frac{y}{x}\right)^3\right)\right) = x - y$, where the error term is bounded by $C\sum_{k=1}^{\infty}\frac{y^{6k+1}}{x^{6k}}$ for a constant C. Moreover, we can also show that the error term approaches to 0 as $\frac{y}{x}$ and y approach to 0.

To prove that the g-function is positive, we consider the function f defined as $f(z) = \frac{1}{2}z^2 - \tanh\left(\frac{1}{2}z^3\right)$. Since f is nonnegative for all $z \geq 0$, we can show that the g-function is positive by substituting $z = \frac{y}{x}$. □

Note from Theorem 3.3 that the error in the approximation is small when $\frac{y}{x}$ approaches to 0 which is the case when $x - y$ is already positive. However, our main goal in this section is to find an approximation of $x - y$ such that (1) it is the closet positive approximation when the original function is negative, and (2) it is also a good approximation when the original function is positive. While there are other positive approximations of $x - y$ considered in literature, for example, the h-function, $h(x, y) = \dfrac{2x}{1 + \exp(\frac{2y}{x})}$, established in Chatterjee and Lahiri [4], we propose the g-function as an alternative positive approximation. Comparing errors in approximations of the two functions, we can show that the error in approximation of h-function is $R_h = -x\left(\tanh\left(\frac{y}{x}\right) - \frac{y}{x}\right)$ and the error in approximation of g-function is $R_g = -\frac{2x^3}{y^2}\left(\tanh\left(\frac{1}{2}\left(\frac{y}{x}\right)^3\right) - \frac{1}{2}\left(\frac{y}{x}\right)^3\right)$. Using basic algebra techniques, we can mathematically prove that $R_h > R_g$ when $\frac{y}{x} < \sqrt{2}$ but the inequality is reversed when $\sqrt{2} < \frac{y}{x} < \sqrt[3]{\pi}$. By considering the domain of approximations, the g-function seems to be a better approximation than the h-function.

Next, we apply Theorem 3.3 to obtain positive approximations to the $\mathrm{tr}(\widehat{\mathbf{MSE}}_{\boldsymbol{\theta}})$ and the $\mathrm{tr}(\widehat{\mathbf{MST}}_{\boldsymbol{\theta}})$ defined in Theorem 3.2. The approximations are stated in the following theorem.

Theorem 3.4. *Define* $\mathrm{tr}(\widehat{\mathbf{MSE}}_{\boldsymbol{\theta},gfunc})$ *and* $\mathrm{tr}(\widehat{\mathbf{MST}}_{\boldsymbol{\theta},gfunc})$ *as*

$$\mathrm{tr}(\widehat{\mathbf{MSE}}_{\boldsymbol{\theta},gfunc}) = g(\mathrm{tr}(\mathbf{MSE}), \mathrm{tr}(\mathbf{D}_{w1}))$$

and

$$\mathrm{tr}(\widehat{\mathbf{MST}}_{\boldsymbol{\theta},gfunc}) = g(\mathrm{tr}(\mathbf{MST}), \mathrm{tr}(\mathbf{D}_{w2})),$$

respectively. Then the following statements hold.

1) *If* $\mathrm{tr}(\mathbf{D}_{w1}) < \sqrt[3]{\pi}\,\mathrm{tr}(\mathbf{MSE})$, *then* $\mathrm{tr}(\widehat{\mathbf{MSE}}_{\boldsymbol{\theta},gfunc})$ *is a positive approximation of* $\mathrm{tr}(\widehat{\mathbf{MSE}}_{\boldsymbol{\theta}})$.

2) *If* $\mathrm{tr}(\mathbf{D}_{w2}) < \sqrt[3]{\pi}\,\mathrm{tr}(\mathbf{MST})$, *then* $\mathrm{tr}(\widehat{\mathbf{MST}}_{\boldsymbol{\theta},gfunc})$ *is a positive approximation of* $\mathrm{tr}(\widehat{\mathbf{MST}}_{\boldsymbol{\theta}})$.

Proof. It is obvious by its definition that $\mathrm{tr}(\mathbf{MSE}) \geq 0$. To prove that $\mathrm{tr}(\mathbf{D}_{w1}) \geq 0$, we use the fact that $\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ is symmetric and idempotent. Then each of its main diagonal elements, $1 - \mathbf{x}_i'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_i$, is non-negative. Consequently, $\mathrm{tr}(\mathbf{D}_{w1}) \geq 0$.

However, the cases when either $\text{tr}(\mathbf{MSE}) = 0$ or $\text{tr}(\mathbf{D}_{w1}) = 0$ are trivial cases where there is no regression error or no sampling error which are not of interest. Therefore, we can consider only the case $\text{tr}(\mathbf{MSE}) > 0$ and $\text{tr}(\mathbf{D}_{w1}) > 0$. Hence, under the condition that $\text{tr}(\mathbf{D}_{w1}) < \sqrt[3]{\pi}\,\text{tr}(\mathbf{MSE})$, $\text{tr}(\widehat{\mathbf{MSE}}_{\boldsymbol{\theta},gfunc})$ is a positive approximation of $\text{tr}(\widehat{\mathbf{MSE}}_{\boldsymbol{\theta}})$.

The same concept can be applied to the $\text{tr}(\widehat{\mathbf{MST}}_{\boldsymbol{\theta},gfunc})$. In particular, $\text{tr}(\mathbf{MST}) \geq 0$ and $\text{tr}(\mathbf{D}_{w2}) \geq 0$ by their definitions. The trivial cases where either $\text{tr}(\mathbf{MST}) = 0$ or $\text{tr}(\mathbf{D}_{w2}) = 0$ are not of interest in the context of our paper and we consider only the case where $\text{tr}(\mathbf{MST}) > 0$ and $\text{tr}(\mathbf{D}_{w2}) > 0$. Hence, under the condition that $\text{tr}(\mathbf{D}_{w2}) < \sqrt[3]{\pi}\,\text{tr}(\mathbf{MST})$, $\text{tr}(\widehat{\mathbf{MST}}_{\boldsymbol{\theta},gfunc})$ is a positive approximation of $\text{tr}(\widehat{\mathbf{MST}}_{\boldsymbol{\theta}})$. □

Therefore, we can use $Adj\ R^2_{gfunc} = 1 - \dfrac{\text{tr}(\widehat{\mathbf{MSE}}_{\boldsymbol{\theta},gfunc})}{\text{tr}(\widehat{\mathbf{MST}}_{\boldsymbol{\theta},gfunc})}$ as an alternative variable selection criterion.

3.3 Truncated Variable Selection Criteria

Since the $Adj\ R^2_{hat}$ is constructed from unbiased estimators of $\text{tr}(\mathbf{MSE}_{\boldsymbol{\theta}})$ and $\text{tr}(\mathbf{MST}_{\boldsymbol{\theta}})$, it approximates the $Adj\ R^2_{true}$ better than the positive adjustment $Adj\ R^2_{gfunc}$ if there is no cases such that either $\text{tr}(\widehat{\mathbf{MSE}}_{\boldsymbol{\theta}})$ or $\text{tr}(\widehat{\mathbf{MST}}_{\boldsymbol{\theta}})$ is negative. However, if such cases occur, the positive adjustments outperform the unbiased estimators. Therefore, we suggest users to use $\text{tr}(\widehat{\mathbf{MSE}}_{\boldsymbol{\theta}})$ and $\text{tr}(\widehat{\mathbf{MST}}_{\boldsymbol{\theta}})$ for variable selection criteria but apply the g-transformation only if $\text{tr}(\widehat{\mathbf{MSE}}_{\boldsymbol{\theta}})$ and/or $\text{tr}(\widehat{\mathbf{MST}}_{\boldsymbol{\theta}})$ are/is negative. In particular, consider new estimators of $\text{tr}(\mathbf{MSE}_{\boldsymbol{\theta}})$ and $\text{tr}(\mathbf{MST}_{\boldsymbol{\theta}})$ as follows:

$$\text{tr}(\widehat{\mathbf{MSE}}_{\boldsymbol{\theta},gtrunc}) = \begin{cases} \text{tr}(\widehat{\mathbf{MSE}}_{\boldsymbol{\theta}}) & \text{if } \text{tr}(\widehat{\mathbf{MSE}}_{\boldsymbol{\theta}}) \geq 0 \\ \text{tr}(\widehat{\mathbf{MSE}}_{\boldsymbol{\theta},gfunc}) & \text{otherwise,} \end{cases}$$

and

$$\text{tr}(\widehat{\mathbf{MST}}_{\boldsymbol{\theta},gtrunc}) = \begin{cases} \text{tr}(\widehat{\mathbf{MST}}_{\boldsymbol{\theta}}) & \text{if } \text{tr}(\widehat{\mathbf{MST}}_{\boldsymbol{\theta}}) \geq 0 \\ \text{tr}(\widehat{\mathbf{MST}}_{\boldsymbol{\theta},gfunc}) & \text{otherwise.} \end{cases}$$

Consequently, we propose $Adj\ R^2_{gtrunc} = 1 - \dfrac{\text{tr}(\widehat{\mathbf{MSE}}_{\boldsymbol{\theta},gtrunc})}{\text{tr}(\widehat{\mathbf{MST}}_{\boldsymbol{\theta},gtrunc})}$ as an alternative variable selection criterion.

4 Numerical Simulations

In simulation experiment, we compare different approximations for the true variable selection criterion in multivariate linear regression in many different cases of the covariance matrix presenting different correlation structures between different response variables. In our simulation, we use covariates from the public-use data for 775 U.S. largest counties from the 2005 Small Area Income and Poverty Estimates (SAIPE) program of the U.S. Census Bureau to compare different adjusted R^2 and simulate other variables by the following algorithm, multivariate linear regression model:

$$\boldsymbol{\theta} = \mathbf{X}\boldsymbol{\beta} + \mathbf{V}, \tag{5}$$

where $\mathbf{V} = [\mathbf{v}_1\ \mathbf{v}_2\ \ldots\ \mathbf{v}_q]$, and $[v_{i1}, v_{i2}, \ldots, v_{iq}] \sim N_q(\mathbf{0}, \mathbf{C})$ for $i = 1, \ldots, m$.

1. Generate a positive definite matrix $\mathbf{C} = [\mathbf{c}_1\ \mathbf{c}_2\ \ldots\ \mathbf{c}_q]$ in four cases:

   (a) $\mathbf{C}$ is a diagonal matrix which elements were σ^2, i.e.

   $$\mathbf{C} = \begin{bmatrix} \sigma^2 & 0 & \cdots & 0 \\ 0 & \sigma^2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \sigma^2 \end{bmatrix},$$

   where $\sigma^2 = 0.3511$ is variance of covariates.

   (b) $\mathbf{C}$ is a matrix with $\mathrm{Var}(\mathbf{c}_i) = \sigma^2$ and $\mathrm{Cov}(\mathbf{c}_i, \mathbf{c}_j) = \gamma^2$ for $i \neq j$, i.e.

   $$\mathbf{C} = \begin{bmatrix} \sigma^2 & \gamma^2 & \cdots & \gamma^2 \\ \gamma^2 & \sigma^2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \gamma^2 \\ \gamma^2 & \cdots & \gamma^2 & \sigma^2 \end{bmatrix},$$

   where $\sigma^2 = 0.3511$ and $\gamma^2 = 0.225$.

   (c) $\mathbf{C}$ is a matrix with $\mathrm{Var}(\mathbf{c}_i) = \sigma^2$ and $\mathrm{Cov}(\mathbf{c}_i, \mathbf{c}_j) = \gamma_{ij}^2$ for $i \neq j$, i.e.

   $$\mathbf{C} = \begin{bmatrix} \sigma^2 & \gamma_{12}^2 & \cdots & \gamma_{1q}^2 \\ \gamma_{21}^2 & \sigma^2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \gamma_{(q-1)q}^2 \\ \gamma_{q1}^2 & \cdots & \gamma_{q(q-1)}^2 & \sigma^2 \end{bmatrix},$$

where $\sigma^2 = 0.3511$ and $\gamma_{ij}^2 (i \neq j) \sim U(0.01, 0.3511)$.

(d) $\mathbf{C} = \sigma^2 \mathbf{\Gamma} = \sigma^2 \mathbf{P}\mathbf{P}'$ is a positive definite matrix, where $\mathbf{P}$ is any lower triangular matrix. In particular,

$$\mathbf{C} = \sigma^2 \begin{bmatrix} p_{11} & 0 & \cdots & 0 \\ p_{21} & p_{22} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ p_{q1} & \cdots & p_{q(q-1)} & p_{qq} \end{bmatrix} \begin{bmatrix} p_{11} & p_{21} & \cdots & p_{q1} \\ 0 & p_{22} & \ddots & \vdots \\ \vdots & \ddots & \ddots & p_{q(q-1)} \\ 0 & \cdots & 0 & p_{qq} \end{bmatrix},$$

where $\sigma^2 = 0.3511$, $p_{ii} \sim U(1.2, 2)$ and $p_{ij} (i > j) \sim U(0.9, 1.1)$

2. Using real covariates $\mathbf{X}$ from the SAIPE 2005 data.

3. Using $\boldsymbol{\beta} = \begin{bmatrix} 0.1449968 & 0.949318 & 0.2822066 \\ 0.4029203 & 0.699723 & 0.9549865 \end{bmatrix}$, we generate $\boldsymbol{\theta}$ using multivariate linear regression model in (5).

4. Generate sampling variance $D_{ij} \sim U(\sigma^2, \sigma^2 + r)$, where r is the range of variance from SAIPE 2005 data, generate $\mathbf{Y}$ from the sampling model

$$\mathbf{Y} = \boldsymbol{\theta} + \mathbf{E},$$

where $e_{ij} \sim N(0, D_{ij})$ for $i = 1, \ldots, m$ and $j = 1, \ldots, q$.

The results shown in the following tables and figures are presented using the following notations:

- $Adj\ R^2_{true} = 1 - \dfrac{\text{tr}(\mathbf{MSE}_{\boldsymbol{\theta}})}{\text{tr}(\mathbf{MST}_{\boldsymbol{\theta}})}$, the true adjusted R^2 ,
- $Adj\ R^2_{naive} = 1 - \dfrac{\text{tr}(\mathbf{MSE})}{\text{tr}(\mathbf{MST})}$, the naive adjusted R^2 that ignores the sampling errors in y,
- $Adj\ R^2_{hat} = 1 - \dfrac{\text{tr}(\widehat{\mathbf{MSE}}_{\boldsymbol{\theta}})}{\text{tr}(\widehat{\mathbf{MST}}_{\boldsymbol{\theta}})}$, an adjustment to naive adjusted R^2 that could go out of range,
- $Adj\ R^2_{hfunc} = 1 - \dfrac{\text{tr}(\widehat{\mathbf{MSE}}_{\boldsymbol{\theta},hfunc})}{\text{tr}(\widehat{\mathbf{MST}}_{\boldsymbol{\theta},hfunc})}$, an adjustment to naive adjusted R^2 by the h-function,
- $Adj\ R^2_{gfunc} = 1 - \dfrac{\text{tr}(\widehat{\mathbf{MSE}}_{\boldsymbol{\theta},gfunc})}{\text{tr}(\widehat{\mathbf{MST}}_{\boldsymbol{\theta},gfunc})}$, an adjustment to naive adjusted R^2 by the g-function,
- $Adj\ R^2_{htrunc} = 1 - \dfrac{\text{tr}(\widehat{\mathbf{MSE}}_{\boldsymbol{\theta},htrunc})}{\text{tr}(\widehat{\mathbf{MST}}_{\boldsymbol{\theta},htrunc})}$, the truncation version of the h-function approximation,
- $Adj\ R^2_{gtrunc} = 1 - \dfrac{\text{tr}(\widehat{\mathbf{MSE}}_{\boldsymbol{\theta},gtrunc})}{\text{tr}(\widehat{\mathbf{MST}}_{\boldsymbol{\theta},gtrunc})}$, the truncation version of the g-function approximation.

In our simulation, we generate 1000 samples using the data generation algorithm explained above and different versions of proposed adjusted R^2. For each sample, we compute true adjusted R^2, naive adjusted R^2 and different proposed adjusted R^2's.

Tables and figures show comparisons of different adjusted R^2 for four different versions of the covariance matrix in the multivariate linear regression models. The numbers presented in the tables are the 1^{st}, 10^{th}, 25^{th}, 50^{th}, 75^{th}, 90^{th} and 100^{th} percentiles, respectively. Figures 1b - 8b show plots of $Adj\ R^2_{hfunc}$ and $Adj\ R^2_{gfunc}$ in different cases.

Percentiles	1	10	25	50	75	90	100
Case (a.1):	$\mathbf{C} = \sigma^2\mathbf{I}$, where $\sigma^2 = 0.3511$						
$Adj\ R^2_{true}$	0.9497	0.9514	0.9524	0.9534	0.9542	0.9550	0.9573
$Adj\ R^2_{naive}$	0.8285	0.8437	0.8438	0.8548	0.8509	0.8453	0.8499
$Adj\ R^2_{hat}$	0.9319	0.9512	0.9512	0.9611	0.9567	0.9552	0.9610
$Adj\ R^2_{hfunc}$	0.9170	0.9328	0.9329	0.9416	0.9380	0.9356	0.9401
$Adj\ R^2_{gfunc}$	0.9312	0.9497	0.9497	0.9591	0.9550	0.9534	0.9588
$Adj\ R^2_{htrunc}$	0.9319	0.9512	0.9512	0.9611	0.9567	0.9552	0.9610
$Adj\ R^2_{gtrunc}$	0.9319	0.9512	0.9512	0.9611	0.9567	0.9552	0.9610
Case (a.2):	$\mathbf{C} = \sigma^2\mathbf{I}$, where $\sigma^2 = 0.3511 \times 0.04$						
$Adj\ R^2_{true}$	0.9979	0.9980	0.9980	0.9980	0.9981	0.9981	0.9982
$Adj\ R^2_{naive}$	0.9185	0.9201	0.9224	0.9253	0.9195	0.9159	0.9181
$Adj\ R^2_{hat}$	0.9950	0.9992	1.0000	1.0012	0.9982	0.9943	0.9943
$Adj\ R^2_{hfunc}$	0.9768	0.9790	0.9800	0.9812	0.9784	0.9757	0.9763
$Adj\ R^2_{gfunc}$	0.9904	0.9930	0.9936	0.9945	0.9924	0.9898	0.9899
$Adj\ R^2_{htrunc}$	0.9950	0.9992	1.0000	0.9812	0.9982	0.9943	0.9943
$Adj\ R^2_{gtrunc}$	0.9950	0.9992	1.0000	0.9945	0.9982	0.9943	0.9943

Table 1: Various percentiles of different adjusted R^2's from 1000 experiments in case (a)

Percentiles	1	10	25	50	75	90	100
Case (b.1):	$\mathbf{C} = (c_{ij})$, where $c_{ii} = \sigma^2 = 0.3511$, $c_{ij}(i \neq j) = \gamma^2 = 0.225$						
$Adj\ R^2_{true}$	0.9484	0.9507	0.9521	0.9533	0.9545	0.9555	0.9599
$Adj\ R^2_{naive}$	0.8505	0.8488	0.8485	0.8564	0.8470	0.8497	0.8563
$Adj\ R^2_{hat}$	0.9549	0.9506	0.9518	0.9619	0.9515	0.9552	0.9596
$Adj\ R^2_{hfunc}$	0.9369	0.9339	0.9345	0.9425	0.9339	0.9369	0.9412
$Adj\ R^2_{gfunc}$	0.9534	0.9493	0.9504	0.9599	0.9501	0.9536	0.9579
$Adj\ R^2_{htrunc}$	0.9549	0.9506	0.9518	0.9619	0.9515	0.9552	0.9596
$Adj\ R^2_{gtrunc}$	0.9549	0.9506	0.9518	0.9619	0.9515	0.9552	0.9596
Case (b.2):	$\mathbf{C} = (c_{ij})$, where $c_{ii} = \sigma^2 = 0.3511 \times 0.04$, $c_{ij}(i \neq j) = \gamma^2 = 0.225 \times 0.04$						
$Adj\ R^2_{true}$	0.9978	0.9979	0.9980	0.9980	0.9981	0.9981	0.9983
$Adj\ R^2_{naive}$	0.9227	0.9204	0.9241	0.9239	0.9259	0.9221	0.9252
$Adj\ R^2_{hat}$	0.9985	0.9962	0.9982	1.0012	0.9994	0.9980	1.0002
$Adj\ R^2_{hfunc}$	0.9794	0.9778	0.9797	0.9809	0.9807	0.9791	0.9808
$Adj\ R^2_{gfunc}$	0.9929	0.9914	0.9928	0.9944	0.9936	0.9925	0.9940
$Adj\ R^2_{htrunc}$	0.9985	0.9962	0.9982	0.9809	0.9994	0.9980	0.9808
$Adj\ R^2_{gtrunc}$	0.9985	0.9962	0.9982	0.9944	0.9994	0.9980	0.9940

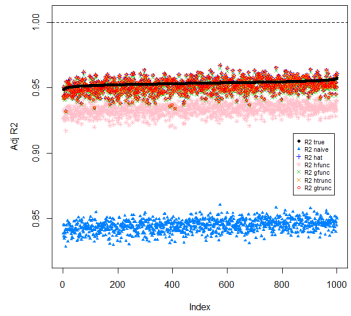
Table 2: Various percentiles of different adjusted R^2's from 1000 experiments in case (b)

Percentiles	1	10	25	50	75	90	100
Case (c.1):	$\mathbf{C} = (c_{ij})$, where $c_{ii} = \sigma^2 = 0.3511$, $c_{ij}(i \neq j) = \gamma_{ij}^2 \sim U(0.01, 0.3511)$						
Adj R^2_{true}	0.9487	0.9509	0.9520	0.9532	0.9544	0.9555	0.9584
Adj R^2_{naive}	0.8408	0.8477	0.8547	0.8512	0.8465	0.8501	0.8498
Adj R^2_{hat}	0.9461	0.9518	0.9592	0.9596	0.9463	0.9581	0.9523
Adj R^2_{hfunc}	0.9291	0.9343	0.9405	0.9397	0.9307	0.9386	0.9352
Adj R^2_{gfunc}	0.9449	0.9504	0.9574	0.9576	0.9452	0.9562	0.9509
Adj R^2_{htrunc}	0.9461	0.9518	0.9592	0.9596	0.9463	0.9581	0.9523
Adj R^2_{gtrunc}	0.9461	0.9518	0.9592	0.9596	0.9463	0.9581	0.9523
Case (c.2):	$\mathbf{C} = (c_{ij})$, where $c_{ii} = \sigma^2 = 0.3511 \times 0.04$, $c_{ij}(i \neq j) = \gamma_{ij}^2 = 0.04 \times g$, where $g \sim U(0.01, 0.3511)$						
Adj R^2_{true}	0.9979	0.9979	0.9980	0.9980	0.9981	0.9981	0.9983
Adj R^2_{naive}	0.9228	0.9231	0.9229	0.9251	0.9291	0.9288	0.9256
Adj R^2_{hat}	0.9967	0.9979	0.9962	0.9979	1.0038	1.0022	1.0024
Adj R^2_{hfunc}	0.9787	0.9793	0.9785	0.9799	0.9833	0.9826	0.9818
Adj R^2_{gfunc}	0.9918	0.9926	0.9915	0.9928	0.9959	0.9952	0.9950
Adj R^2_{htrunc}	0.9967	0.9979	0.9962	0.9979	0.9833	0.9826	0.9818
Adj R^2_{gtrunc}	0.9967	0.9979	0.9962	0.9979	0.9959	0.9952	0.9950

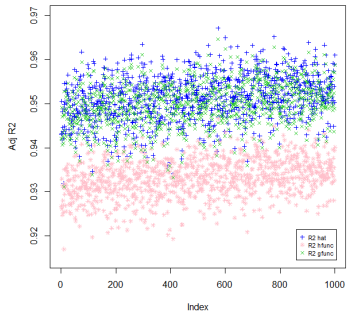
Table 3: Various percentiles of different adjusted R^2's from 1000 experiments in case (c)

Percentiles	1	10	25	50	75	90	100
case (d.1):	$\mathbf{C} = \sigma^2\mathbf{\Gamma} = \sigma^2\mathbf{PP'}$, where $\sigma^2 = 0.3511$, $p_{ii} \sim U(1.2, 2)$, $p_{ij}(i > j) \sim U(0.9, 1.1)$						
$Adj\ R^2_{true}$	0.8215	0.8280	0.8320	0.8362	0.8400	0.8439	0.8552
$Adj\ R^2_{naive}$	0.7444	0.7413	0.7533	0.7492	0.7464	0.7653	0.7732
$Adj\ R^2_{hat}$	0.8237	0.8196	0.8326	0.8328	0.8290	0.8465	0.8525
$Adj\ R^2_{hfunc}$	0.8190	0.8151	0.8277	0.8273	0.8237	0.8410	0.8471
$Adj\ R^2_{gfunc}$	0.8237	0.8196	0.8325	0.8328	0.8290	0.8465	0.8524
$Adj\ R^2_{htrunc}$	0.8237	0.8196	0.8326	0.8328	0.8290	0.8465	0.8525
$Adj\ R^2_{gtrunc}$	0.8237	0.8196	0.8326	0.8328	0.8290	0.8465	0.8525
case (d.2):	$\mathbf{C} = \sigma^2\mathbf{\Gamma} = \sigma^2\mathbf{PP'}$, where $\sigma^2 = 0.3511 \times 0.04$, $p_{ii} \sim U(1.2, 2)$, $p_{ij}(i > j) \sim U(0.9, 1.1)$						
$Adj\ R^2_{true}$	0.9915	0.9918	0.9920	0.9922	0.9924	0.9926	0.9930
$Adj\ R^2_{naive}$	0.9210	0.9203	0.9171	0.9193	0.9175	0.9157	0.9198
$Adj\ R^2_{hat}$	0.9913	0.9945	0.9907	0.9959	0.9918	0.9918	0.9956
$Adj\ R^2_{hfunc}$	0.9758	0.9771	0.9744	0.9774	0.9750	0.9746	0.9774
$Adj\ R^2_{gfunc}$	0.9881	0.9903	0.9874	0.9911	0.9882	0.9881	0.9909
$Adj\ R^2_{htrunc}$	0.9913	0.9945	0.9907	0.9959	0.9918	0.9918	0.9956
$Adj\ R^2_{gtrunc}$	0.9913	0.9945	0.9907	0.9959	0.9918	0.9918	0.9956

Table 4: Various percentiles of different adjusted R^2's from 1000 experiments in case (d)

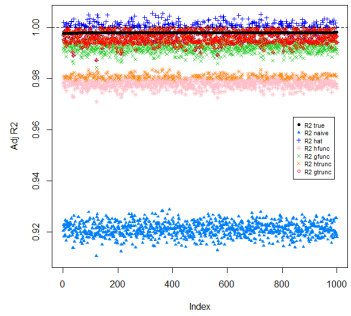


(a) Plot of different adjusted R^2's.

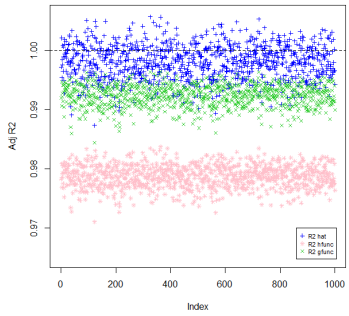


(b) Plot of comparison of $Adj\ R^2_{hfunc}$ and $Adj\ R^2_{gfunc}$.

Figure 1: Plot of different adjusted R^2's in case (a.1).

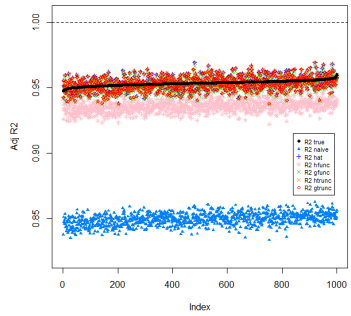


(a) Plot of different adjusted R^2's.

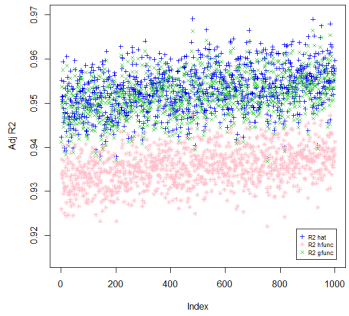


(b) Plot of comparison of $Adj\ R^2_{hfunc}$ and $Adj\ R^2_{gfunc}$.

Figure 2: Plot of different adjusted R^2's in case (a.2).

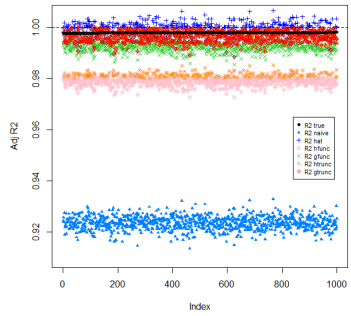


(a) Plot of different adjusted R^2's.

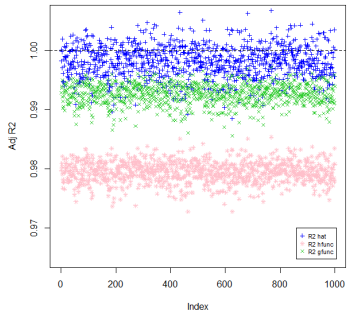


(b) Plot of comparison of $Adj\ R^2_{hfunc}$ and $Adj\ R^2_{gfunc}$.

Figure 3: Plot of different adjusted R^2's in case (b.1).

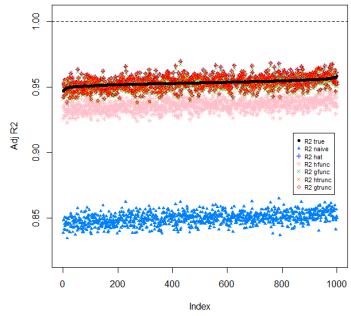


(a) Plot of different adjusted R^2's.

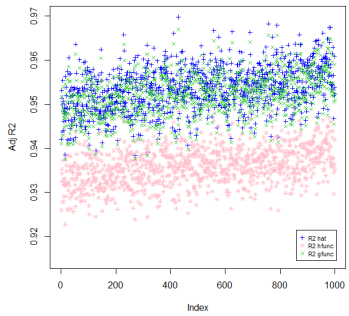


(b) Plot of comparison of $Adj\ R^2_{hfunc}$ and $Adj\ R^2_{gfunc}$.

Figure 4: Plot of different adjusted R^2's in case (b.2).

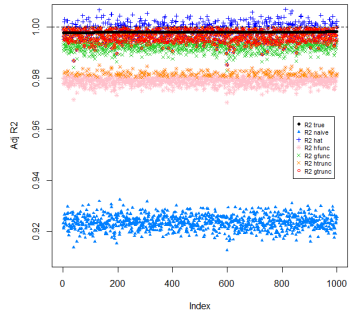


(a) Plot of different adjusted R^2's.

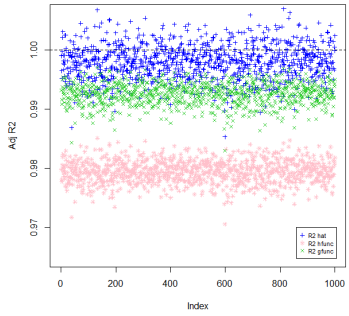


(b) Plot of comparison of $Adj\ R^2_{hfunc}$ and $Adj\ R^2_{gfunc}$.

Figure 5: Plot of different adjusted R^2's in case (c.1).

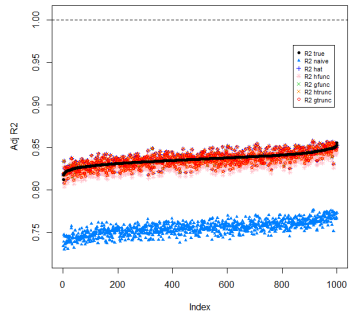


(a) Plot of different adjusted R^2's.

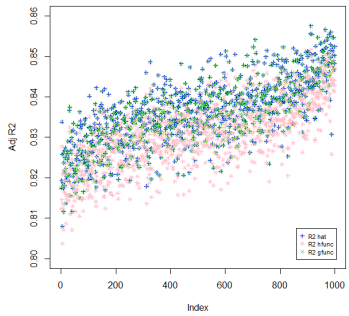


(b) Plot of comparison of $Adj\ R^2_{hfunc}$ and $Adj\ R^2_{gfunc}$.

Figure 6: Plot of different adjusted R^2's in case (c.2).

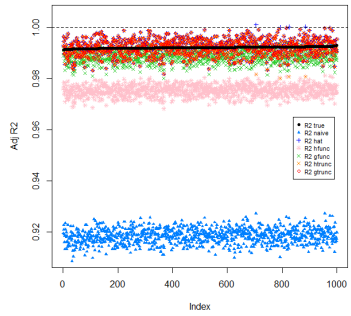


(a) Plot of different adjusted R^2's.

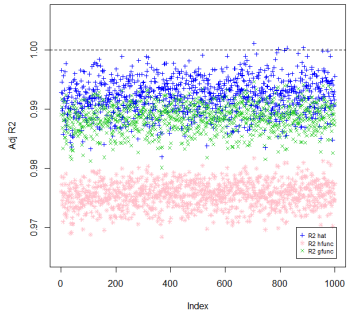


(b) Plot of comparison of $Adj\ R^2_{hfunc}$ and $Adj\ R^2_{gfunc}$.

Figure 7: Plot of different adjusted R^2's in case (d.1).



(a) Plot of different adjusted R^2's.



(b) Plot of comparison of $Adj\ R^2_{hfunc}$ and $Adj\ R^2_{gfunc}$.

Figure 8: Plot of different adjusted R^2's in case (d.2).

From Tables 1 – 4, the naive adjusted R^2's are smaller than the true adjusted R^2 but our proposed adjusted R^2's can cut down those biases. Comparing performances between the two positive approximations, *Adj* R^2_{hfunc} and *Adj* R^2_{gfunc}, we can see that the *Adj* R^2_{gfunc} is closer to the *Adj* R^2_{hat} than the *Adj* R^2_{hfunc}. Considering the situations when either $\mathrm{tr}(\widehat{\mathbf{MSE}}_{\boldsymbol{\theta}})$ or $\mathrm{tr}(\widehat{\mathbf{MST}}_{\boldsymbol{\theta}})$ could be negative, we can see that there is no cases such that at least one of the two values is negative in cases (a.1) – (d.1). However, when we reduce the variances of regression error terms by a factor of 0.04, we can see some negative values as follows. In case (a.2), there are 276 cases such that either $\mathrm{tr}(\widehat{\mathbf{MSE}}_{\boldsymbol{\theta}})$ or $\mathrm{tr}(\widehat{\mathbf{MST}}_{\boldsymbol{\theta}})$ is negative. In case (b.2), there are 247 cases such that either $\mathrm{tr}(\widehat{\mathbf{MSE}}_{\boldsymbol{\theta}})$ or $\mathrm{tr}(\widehat{\mathbf{MST}}_{\boldsymbol{\theta}})$ is negative. In case (c.2), there are 235 cases such that either $\mathrm{tr}(\widehat{\mathbf{MSE}}_{\boldsymbol{\theta}})$ or $\mathrm{tr}(\widehat{\mathbf{MST}}_{\boldsymbol{\theta}})$ is negative. In case (d.2), there are 4 cases such that either $\mathrm{tr}(\widehat{\mathbf{MSE}}_{\boldsymbol{\theta}})$ or $\mathrm{tr}(\widehat{\mathbf{MST}}_{\boldsymbol{\theta}})$ is negative. Therefore, the *Adj* R^2_{gtrunc} is recommended for these cases.

5 Conclusion

In this study, we have examined the possibility of extending the bias reduction of variable selection criteria proposed in Lahiri and Suntornchost [7] to multivariate linear regression models subject to sampling errors. Moreover, we have proposed a new positive approximation function, g-function, which is shown to be a better approximation than the h-function used in their paper. From our study, we found that the naive adjusted R^2 always underestimate the true adjusted R^2, but our proposed adjusted R^2, *Adj* R^2_{hat} reduce this underestimation. This simple adjustment works well except in some cases when it exceeds the suitable range. In these cases, adjustments by the g-function, *Adj* R^2_{gfunc} is helpful. To accommodate all situations, the *Adj* R^2_{gtrunc} is recommended.

Acknowledgements: The first author would like to thank the Department of Mathematics and Computer Science, Faculty of Science, Chulalongkorn University, and the His Royal Highness Crown Prince Maha Vajiralongkorn Fund for financial support.

References

[1] J.L. Aleixandre-Tud and I. Alvarez, Application of Multivariate Regression Methods to Predict Sensory Quality of Red Wines, *Czech J. Food Sci.*, **33**(2015), 217–227.

[2] T.W. Anderson, *An Introduction to Multivariate Statistical Analysis*, (3rd ed.), John Wiley & Sons, Inc, 2003.

[3] B.E. Barrett and J.B. Gray, A computational framework for variable selection in multivariate regression, *Statistics and Computing*, **4**(1994), 203–212.

[4] S. Chatterjee and P. Lahiri, A Simple Computational Method for Estimating Mean Squared Prediction Error in General Small-Area Model, *JSM Proceedings*, (2007), 3486–3493.

[5] T. Cserhti and M. Szőgyi, Application of Multivariate Regression Models in Chromatography, *Eur. Chem. Bull.*, **1**(2012), 274–279.

[6] R.A. Johnson and D.W. Wichern, *Applied Multivariate Statistical Analysis*, (6th ed.), Pearson Education, Inc, 2007.

[7] P. Lahiri and J. Suntornchost, Variable Selection for Linear Mixed Models with Applications in Small Area Estimation, *Sankhya B*, **77**(2015), 312–320.

[8] N. Lenghoe, and J. Suntornchost, Variable Selection for linear Regression Model with General Variance, *Journal of Applied Statistics and Information Technology*, **2**(2017), 15–24.

[9] A.C. Rencher, *Methods of Multivariate Analysiss*, (2nd ed.), John Wiley & Sons, Inc, 2002.

[10] A.C. Rencher and G.B. Schaalje, *Linear Models in Statistics*, (2nd ed.), John Wiley & Sons, Inc, 2008.

[11] O. Seidou and J.J. Asselin, Bayesian Multivariate Linear Regression with Application to change point models in Hydrometeorological variables, *Water Resour. Res.*, (2007), 43.

[12] D.W. Smith and D.S. Gill, Variable Selection in Multivariate Multiple Regression, *J. Statist. Comput. Simul.*, **22**(1985), 217–227.

Annop Angkunsit
Department of Mathematics and Computer Science, Faculty of Science,
Chulalongkorn University, Bangkok 10330, Thailand
Email: annop_ngoun@hotmail.com

Jiraphan Suntornchost
Department of Mathematics and Computer Science, Faculty of Science,
Chulalongkorn University, Bangkok 10330, Thailand
Email: jiraphan.s@chula.ac.th