Variable Selection for Linear Regression Model with General Variance

Nuchjaree Lenghoe[1*], and Jiraphan Suntornchost[2]

[1] *Mathematics and Computer Science, Chulalongkorn University, Bangkok, 10330, Thailand, nuchjaree.lh@gmail.com*

[2] *Mathematics and Computer Science, Chulalongkorn University, Bangkok, 10330, Thailand, jiraphan.s@chula.ac.th*

Abstract

Lahiri and Suntornchost (2015) proposed adjustments to variable selection criteria for the Fay-Herriot model by considering sampling errors. Their results were shown to improve traditional variable selection criteria. In this study, we extend their method to construct variable selection criteria for linear regression models with general variance assumption allowing for the possibility of correlated regression errors. Closed forms of statistics for variable selection criteria are provided with numerical studies. Simulation results show that our proposed variable selection criteria can reduce the approximation errors of the standard variable selection criterion.

Keywords: Variable selection, Bias-reduction, Adjusted R^2.

1. Introduction

Frequently, one of the most difficult problems in regression analysis is to study the relationship between the variable of interest $\mathbf{y}$ and other observed variables $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_p$ by examining sample observations. Data can often be modelled in many different ways attempting to use all covariates or only a subset of covariates when many covariates are measured. Variable selection procedures require a criterion to start with, well-known variable selection criteria in linear regression analysis such as $RMSE$ (root mean square error), Adjusted R^2, AIC (Akaike 1973), Schwarz's BIC (Schwarz 1978). In addition, considerable efforts have been made by many researchers to obtain better criteria, by either modifying the existing ones or creating innovative new measures.

Among those researches, Lahiri and Suntornchost (2015) proposed a variable selection criteria for the Fay-Herriot model, a standard linear regression model when response variable is subject to sampling error variability. They suggested ways to adjust variable selection methods that reduce the approximation errors for the Fay-Herriot model where model error terms are assumed to be independent and have the same variance. Their results were shown to improve traditional statistics for variable selection criteria. However, the assumptions of having the same variance and being independent of regression error terms could be too restricted and might not be applicable to some applications. Many models in applications may require relaxed assumptions of regression errors to allow regression errors of different individuals to have unequal variance and also be correlated. In real situation, we hardly know the true properties of response variables, therefore it is important to consider variable selection method for regression model with general assumptions. Hence, in this work, we will extend their method to adjust the original variable selection criteria for general linear regression model with possibly correlated regression errors;

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

where $\mathbf{y}$ is a $m \times 1$ vector of response variables of interest, $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_p)$ is a $m \times p$ design matrix of covariates with rank p, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)'$ is a $p \times 1$ vector of unknown regression coefficients and $\boldsymbol{\varepsilon}$ is a $m \times 1$ vector of regression errors such that $E(\boldsymbol{\varepsilon}) = \mathbf{0}$ and $cov(\boldsymbol{\varepsilon}) = \sigma^2\mathbf{V}$, with $\mathbf{V}$ is a $m \times m$ known positive definite matrix and σ^2 is a constant. We consider a situation when the values of the response variables are not observed due to sampling errors, but are approximated by direct survey values.

The remainder of this paper is organized as follows. In section 2, adjustments to statistics for variable selection criteria are presented. Section 3 presents simulation results. Conclusion and discussion are provided in Section 4.

2. Research Methodology

The model (1) can be reduced to the standard linear regression model with the assumption of uncorrelated regression errors having the same variance if $\mathbf{V} = \mathbf{I}_m$ is the $m \times m$ identity matrix. Since $\mathbf{V}$ is positive definite, there exists an $n \times n$ nonsingular matrix $\mathbf{P}$ such that $\mathbf{V} = \mathbf{PP}'$. Therefore, to obtain parameter estimation and the statistics needed in model selection criteria, we obtain a transformed model:

$$\mathbf{P}^{-1}\mathbf{y} = \mathbf{P}^{-1}\mathbf{X}\beta + \mathbf{P}^{-1}\varepsilon. \tag{2}$$

Therefore, the generalized least squares (GLS) estimator of $\boldsymbol{\beta}$ is then given by

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}. \tag{3}$$

Some commonly used variable selection criteria in regression analysis are;

Root mean square error: $RMSE = \sqrt{MSE}$,

Adjusted R^2: Adj $R^2 = 1 - \frac{MSE}{MST}$,

Akaike information criterion: $\text{AIC} = 2p + m \cdot \log\left(\frac{SSE}{m}\right)$,

Bayesian information criterion: $\text{BIC} = p \cdot \log(m) + m \cdot \log\left(\frac{SSE}{m}\right)$,

where $MSE = \frac{SSE}{m-p}$, $MST = \frac{SST}{m-1}$, SSE is sum of square error, SST is total sum of squares, m is sample size and p is the number of covariates in the regression model. We can notice that these variable selection criteria can be expressed as a smooth function of MSE and MST, $f(MSE, MST)$. The assumption of variable selection criteria is that the value of response variable is ideal value.

From model in (2) and GLS estimator, we can express the sum of squared error (SSE) and the total sum of squares (SST) for general linear regression model as quadratic forms,

$$SSE = \mathbf{y}'(\mathbf{V}^{-1} - \mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1})\mathbf{y}$$
$$SST = \mathbf{y}'\left(\mathbf{V}^{-1} - m^{-1}(\mathbf{P}^{-1})'\mathbf{J}\mathbf{P}^{-1}\right)\mathbf{y}$$

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

In this paper, we are interested in variable selection criteria for the following linear regression model with general variance,

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

where $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_m)'$, θ_i is the unobserved response variable for individual i and $\mathbf{v}$ is the vector of error terms with mean zero and covariance matrix $cov(\mathbf{v}) = \sigma^2\mathbf{V}$ when σ^2 is unknown and $\mathbf{V}$ is a known positive definite matrix. However, the true response variable $\boldsymbol{\theta}$ is unobserved, therefore the value of the response variables $\theta_i\,(i = 1, \ldots, m)$ is approximated by observed value $y_i\,(i = 1, \ldots, m)$. The relation between θ_i and y_i can be expressed as

$$\mathbf{y} = \boldsymbol{\theta} + e$$

where $\mathbf{y} = (y_1, \ldots, y_m)'$ and $\boldsymbol{e} = (e_1, \ldots, e_m)'$ is the vector of sampling errors, we assume that the $e_i\,(i = 1, \ldots, m)$ are independent over i having mean zero and known sampling variances D_i.

In practice, the response variable $\boldsymbol{\theta}$ is replaced by $\mathbf{y}$ in linear regression model that ignore sampling errors and use variable selection criteria which cause to error. So we want to see the effect of sampling error on variable selection criteria, that is examine the scope of error in approximating $f(MSE_\theta, MST_\theta)$, true variable selection criterion, by $f(MSE, MST)$, naive variable selection criterion, in presence of sampling errors. Since variable selection criteria can be expressed as function of mean squares, MSE and MST, so we want to see the effect of sampling error on MSE and MST. It seems reasonable to consider the properties of MSE and MST conditional on θ.

We are interested in obtaining a good approximation to ideal variable selection criterion. We propose a simple adjustment to reduce bias in traditional statistics as the following theorem.

Theorem 1. *The unbiased estimators for MSE_θ and MST_θ are defined as:*

$$\begin{aligned}\widehat{MSE}_\theta &= MSE - D_{w1},\\ \widehat{MST}_\theta &= MST - D_{w2},\end{aligned}$$

respectively, where $D_{w1} = \frac{1}{m-p}\,\mathrm{tr}\{(\mathbf{V}^{-1} - \mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1})\Sigma_e\}$,
$D_{w2} = \frac{1}{m-1}\,\mathrm{tr}\{(\mathbf{V}^{-1} - m^{-1}(\mathbf{P}^{-1})\mathbf{J}\mathbf{P}^{-1})\Sigma_e\}$ *and* $\Sigma_e = \mathrm{diag}(D_1, D_2, \ldots, D_m)$.

Proof. Consider conditional expectation with respect to $\boldsymbol{\theta}$,

$$\begin{aligned}E[\widehat{MSE}_\theta - MSE_\theta \,|\, \boldsymbol{\theta}] &= E[MSE - D_{w1} \,|\, \boldsymbol{\theta}] - MSE_\theta\\ &= E[MSE \,|\, \boldsymbol{\theta}] - D_{w1} - MSE_\theta\end{aligned}$$

and

$$\begin{aligned}E[\widehat{MST}_\theta - MST_\theta \,|\, \boldsymbol{\theta}] &= E[MST - D_{w2} \,|\, \boldsymbol{\theta}] - MST_\theta\\ &= E[MST \,|\, \boldsymbol{\theta}] - D_{w2} - MST_\theta.\end{aligned}$$

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

$$\begin{aligned}E\,[MSE \,|\, \boldsymbol{\theta}] &= \frac{1}{m-p}\,E\left[\mathbf{y}'(\mathbf{V}^{-1} - \mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1})\mathbf{y} \,\middle|\, \boldsymbol{\theta}\right]\\ &= \frac{1}{m-p}\,E\left[(\boldsymbol{\theta} + e)'(\mathbf{V}^{-1} - \mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1})(\boldsymbol{\theta} + e) \,\middle|\, \boldsymbol{\theta}\right]\\ &= \frac{1}{m-p}\Big[SSE_\theta - 2\boldsymbol{\theta}'(\mathbf{V}^{-1} - \mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1})E\,[\mathbf{e}]\\ &\qquad + E\left[\mathbf{e}'(\mathbf{V}^{-1} - \mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1})\mathbf{e}\right]\Big]\\ &= MSE_\theta + \frac{1}{m-p}\,E\left[\mathbf{e}'(\mathbf{V}^{-1} - \mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1})\mathbf{e}\right]\\ &= MSE_\theta + \frac{1}{m-p}\,\mathrm{tr}\left\{(\mathbf{V}^{-1} - \mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1})\Sigma_e\right\}\\ &= MSE_\theta + D_{w1}.\end{aligned}$$

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

$$E\,[MST \,|\, \boldsymbol{\theta}] = \frac{1}{m-1}\,E\left[\mathbf{y}'\left(\mathbf{V}^{-1} - \frac{1}{m}\,(\mathbf{P}^{-1})'\mathbf{J}\mathbf{P}^{-1}\right)\mathbf{y} \,\middle|\, \boldsymbol{\theta}\right]$$

$$
\begin{aligned}
&= \frac{1}{m-1} E\left[(\boldsymbol{\theta}+\mathbf{e})' \left(\mathbf{V}^{-1} - \frac{1}{m} (\mathbf{P}^{-1})'\mathbf{J}\mathbf{P}^{-1} \right) (\boldsymbol{\theta}+\mathbf{e}) \,\middle|\, \boldsymbol{\theta} \right] \\
&= \frac{1}{m-1} \left[SST_\theta - 2\boldsymbol{\theta}' \left(\mathbf{V}^{-1} - \frac{1}{m} (\mathbf{P}^{-1})'\mathbf{J}\mathbf{P}^{-1} \right) E[\mathbf{e}] + E\left[\mathbf{e}' \left(\mathbf{V}^{-1} - \frac{1}{m} (\mathbf{P}^{-1})'\mathbf{J}\mathbf{P}^{-1} \right) \mathbf{e} \right] \right] \\
&= MST_\theta + \frac{1}{m-1} E\left[\mathbf{e}' \left(\mathbf{V}^{-1} - \frac{1}{m} (\mathbf{P}^{-1})'\mathbf{J}\mathbf{P}^{-1} \right) \mathbf{e} \right] \\
&= MST_\theta + \frac{1}{m-1} \operatorname{\mathbf{tr}} \left\{ \left(\mathbf{V}^{-1} - \frac{1}{m} (\mathbf{P}^{-1})'\mathbf{J}\mathbf{P}^{-1} \right) \Sigma_e \right\} \\
&= MST_\theta + D_{w2}.
\end{aligned}
$$

Hence, $$E[\widehat{MSE}_\theta - MSE_\theta \,|\, \boldsymbol{\theta}] = MSE_\theta + D_{w1} - D_{w1} - MSE_\theta = 0$$

and $$E[\widehat{MST}_\theta - MST_\theta \,|\, \boldsymbol{\theta}] = MST_\theta + D_{w2} - D_{w2} - MST_\theta = 0.$$

Therefore, $\widehat{MSE}_\theta$ and $\widehat{MST}_\theta$ are unbiased estimators for MSE_θ and MST_θ, respectively. □

Theorem 2. *The estimators $\widehat{MSE}_\theta$ and $\widehat{MST}_\theta$ defined in Theorem 1 are consistent estimators.*

Proof. Consider variance of estimators conditional on $\boldsymbol{\theta}$,

$$
\begin{aligned}
Var[\widehat{MSE}_\theta - MSE_\theta \,|\, \boldsymbol{\theta}] &= Var[MSE - D_{w1} \,|\, \boldsymbol{\theta}] \\
&= Var[MSE \,|\, \boldsymbol{\theta}]
\end{aligned}
$$

and

$$
\begin{aligned}
Var[\widehat{MST}_\theta - MST_\theta \,|\, \boldsymbol{\theta}] &= Var[MST - D_{w2} \,|\, \boldsymbol{\theta}] \\
&= Var[MST \,|\, \boldsymbol{\theta}].
\end{aligned}
$$

It is obvious that $Var\,(MST \,|\, \boldsymbol{\theta}) = O(m^{-1})$. Therefore, $\widehat{MST}_\theta$ is a consistent estimate of MST_θ. To prove that $\widehat{MSE}_\theta$ is a consistent estimate of MSE_θ , note that

$$\frac{SSE}{\sigma^2} = \frac{\mathbf{y}'(\mathbf{V}^{-1} - \mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1})\mathbf{y}}{\sigma^2}$$

follows a chi-square distribution with parameter $m-p$. Therefore, the variance of MSE conditional on $\boldsymbol{\theta}$ is given by

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

Hence, $$Var[\widehat{MSE}_\theta - MSE_\theta \,|\, \boldsymbol{\theta}] = O(m^{-1}).$$

Therefore,

$$\widehat{MSE}_\theta - MSE_\theta \quad \xrightarrow{p} \quad 0,$$

where $\xrightarrow{p}$ denotes convergence in probability. Thus, $\widehat{MSE}_\theta$ and $\widehat{MST}_\theta$ are consistent estimators for MSE_θ and MST_θ, respectively. □

Based on results from the two theorems, we propose $f(\widehat{MSE}_\theta, \widehat{MST}_\theta)$ as a new variable selection criterion for linear regression model with general variance assumption and unobserved response variable. We obtain,

$$f(\widehat{MSE}_\theta, \widehat{MST}_\theta) - f(MSE_\theta, MST_\theta) \xrightarrow{p} 0.$$

However, in some situations, $\widehat{MSE}_\theta$ and/or $\widehat{MST}_\theta$ could be negative which makes $f(\widehat{MSE}_\theta, \widehat{MST}_\theta)$ go out of the admissible range. For example, adjusted R^2 should belong to the interval $\left(-\frac{p-1}{m-p}, 1\right)$. If either $\widehat{MSE}_\theta$ or $\widehat{MST}_\theta$ is negative, the proposed adjusted R^2 may be greater than one.

We advise an adjustment to our proposed variable selection criterion by suggest strictly positive approximations to MSE_θ and MST_θ. Chatterjee and Lahiri (2007) have provided

$$h(x,b) = \frac{2x}{1+\exp\{\frac{2b}{x}\}}.$$

Consider the following alternative approximations of MSE_θ and MST_θ,

Theorem 3. *The $\widehat{MSE}_{\theta,hfunc}$ and $\widehat{MST}_{\theta,hfunc}$ estimators defined as:*

$$\begin{aligned}\widehat{MSE}_{\theta,hfunc} &= h(MSE, D_{w1}),\\ \widehat{MST}_{\theta,hfunc} &= h(MSE, D_{w2}).\end{aligned}$$

are always positive, and they also satisfy

$$\begin{aligned}\widehat{MSE}_{\theta,hfunc} - MSE_\theta &\xrightarrow{p} 0,\\ \widehat{MST}_{\theta,hfunc} - MST_\theta &\xrightarrow{p} 0.\end{aligned}$$

Proof. Obviously that $\widehat{MSE}_{\theta,hfunc}$ and $\widehat{MST}_{\theta,hfunc}$ are both strictly positive and we have

$$\begin{aligned}\widehat{MSE}_{\theta,hfunc} - [MSE - D_{w1}] &\xrightarrow{p} 0,\\ \widehat{MST}_{\theta,hfunc} - [MST - D_{w2}] &\xrightarrow{p} 0.\end{aligned}$$

Follow from Theorem 1 and Theorem 2, the differences between the estimators and the corresponding statistics converge to zero in probability. □

Therefore, we propose $f(\widehat{MSE}_{\theta,hfunc}, \widehat{MST}_{\theta,hfunc})$ as an alternative variable selection criteria and so

$$f(\widehat{MSE}_{\theta,hfunc}, \widehat{MST}_{\theta,hfunc}) - f(MSE_\theta, MST_\theta) \xrightarrow{p} 0.$$

However, $f(\widehat{MSE}_\theta, \widehat{MST}_\theta)$ approximates true variable selection criterion $f(MSE_\theta, MST_\theta)$ better than $f(\widehat{MSE}_{\theta,hfunc}, \widehat{MST}_{\theta,hfunc})$ because $\widehat{MSE}_\theta$ and $\widehat{MST}_\theta$ are unbiased estimators of MSE_θ and MST_θ, whereas $\widehat{MSE}_{\theta,hfunc}$ and $\widehat{MST}_{\theta,hfunc}$ are approximations of $\widehat{MSE}_{\theta}$ and $\widehat{MST}_{\theta}$, respectively. Therefore, we suggest users to use $f(\widehat{MSE}_{\theta}, \widehat{MST}_{\theta})$ for variable selection but apply the $h-$transformation to $f(\widehat{MSE}_{\theta,hfunc}, \widehat{MST}_{\theta,hfunc})$ only if $\widehat{MSE}_{\theta}$ and/or $\widehat{MST}_{\theta}$ are/is negative.

Consequently, we can also propose an alternative variable selection criterion as $f(\widehat{MSE}_{\theta,trun}, \widehat{MST}_{\theta,trun})$ where $\widehat{MSE}_{\theta,trun}$ and $\widehat{MST}_{\theta,trun}$ are defined as follows

$$\widehat{MSE}_{\theta,trun} = \begin{cases} \widehat{MSE}_{\theta} & \text{if } \widehat{MSE}_{\theta} \geq 0 \\ \widehat{MSE}_{\theta,hfunc} & \text{otherwise,} \end{cases}$$

and

$$\widehat{MST}_{\theta,trun} = \begin{cases} \widehat{MST}_{\theta} & \text{if } \widehat{MST}_{\theta} \geq 0 \\ \widehat{MST}_{\theta,hfunc} & \text{otherwise.} \end{cases}$$

Note that the proposed variable selection criteria may choose the same model selected by variable selection criteria $f(MSE, MST)$, having the same covariates. Our proposed variable selection criteria are expected to approximate the corresponding true variable selection criterion better than the naive variable selection criteria.

3. Simulation study

In simulation experiment, we use 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 difference adjusted R^2. For details on SAIPE, the readers are referred to Bell (1999) and the website: http://www.census.gov/hhes/www/saipe.html.

For our simulation experiment, we use the following algorithm to generate data.

1. Generate positive definite matrix $\mathbf{V}$ where we consider in three cases: (1) $\mathbf{V}$ is diagonal matrix which elements were generated from the uniform distribution $U(0.01, 0.1)$, (2) $\mathbf{V}$ is diagonal matrix which elements were generated from the uniform distribution $U(0.1, 1)$, and (3) $\mathbf{V} = \mathbf{PP}'$ where $\mathbf{P}$ is any lower triangular with $p_{ii} \sim U(1, 2)$ and $p_{ij}(i > j) \sim U(0.9, 1.1)$.

2. Using $\boldsymbol{\beta} = (0.8738, 0.0204)'$, $\sigma^2 = 0.0351$, $\mathbf{V}$ and real x_i from SAIPE 2005 data, we generate $\boldsymbol{\theta}$ using linear regression model, $\boldsymbol{\theta} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ where $\boldsymbol{\varepsilon} \sim N_m(\mathbf{0}, \sigma^2\mathbf{V})$.

3. Using real D_i from SAIPE 2005 data, generate $\mathbf{y}$ from sampling model, $\mathbf{y} = \boldsymbol{\theta} + e$ where $e_i \sim N(0, D_i)$ for $i = 1, \dots, m$.

We first generate one sample to compute true adjusted R^2, naive adjusted R^2 and proposed adjusted R^2 and compare those R^2 among difference sample sizes by considering sample size in 4 cases: $m = 775$ (100%), $m = 580$ (75%), $m = 194$ (25%), $m = 78$ (10%). The results are presented using the following notations:

- Adj $R^2_{true} = 1 - \frac{MSE_{\theta}}{MST_{\theta}}$, the true adjusted R^2,
- Adj $R^2_{naive} = 1 - \frac{MSE}{MST}$, the standard adjusted R^2 that ignores the sampling errors in y,
- Adj $R^2_{hat} = 1 - \frac{\widehat{MSE}_{\theta}}{\widehat{MST}_{\theta}}$, an adjustment to naive adjusted R^2 that could go out of range,
- Adj $R^2_{hfunc} = 1 - \frac{\widehat{MSE}_{\theta,hfunc}}{\widehat{MST}_{\theta,hfunc}}$, an adjustment to naive adjusted R^2 that is constrained within an admissible range,

- Adj $R^2_{trun} = 1 - \frac{\widehat{MSE}_{\theta,trun}}{\widehat{MST}_{\theta,trun}}$, an adjustment to naive adjusted R^2 that is constrained within an admissible range.

m	775 (100%)	580 (75%)	194 (25%)	78 (10%)
case (1): $\mathbf{V} = \text{diag}(v_1, \ldots, v_m)$ where $v_i \sim U(0.01, 0.1)$				
Adj R^2_{true}	0.9984	0.9982	0.9972	0.9968
Adj R^2_{naive}	0.9778	0.9699	0.9355	0.9025
Adj R^2_{hat}	0.9982	0.9973	0.9967	0.9933
Adj R^2_{hfunc}	0.9938	0.9914	0.9822	0.9714
Adj R^2_{trun}	0.9982	0.9973	0.9967	0.9933
case (2): $\mathbf{V} = \text{diag}(v_1, \ldots, v_m)$ where $v_i \sim U(0.1, 1)$				
Adj R^2_{true}	0.9828	0.9824	0.9781	0.9842
Adj R^2_{naive}	0.9606	0.9532	0.9160	0.9044
Adj R^2_{hat}	0.9820	0.9807	0.9723	0.9772
Adj R^2_{hfunc}	0.9800	0.9777	0.9641	0.9641
Adj R^2_{trun}	0.9820	0.9807	0.9723	0.9772
case (3): $\mathbf{V} = \mathbf{PP}'$ where $p_{ii} \sim U(1,2)$ and $p_{ij}(i > j) \sim U(0.9, 1.1)$				
Adj R^2_{true}	0.9392	0.9253	0.8614	0.8396
Adj R^2_{naive}	0.8768	0.8736	0.7937	0.7222
Adj R^2_{hat}	0.9391	0.9251	0.8419	0.7659
Adj R^2_{hfunc}	0.9328	0.9215	0.8404	0.7651
Adj R^2_{trun}	0.9391	0.9251	0.8419	0.7659

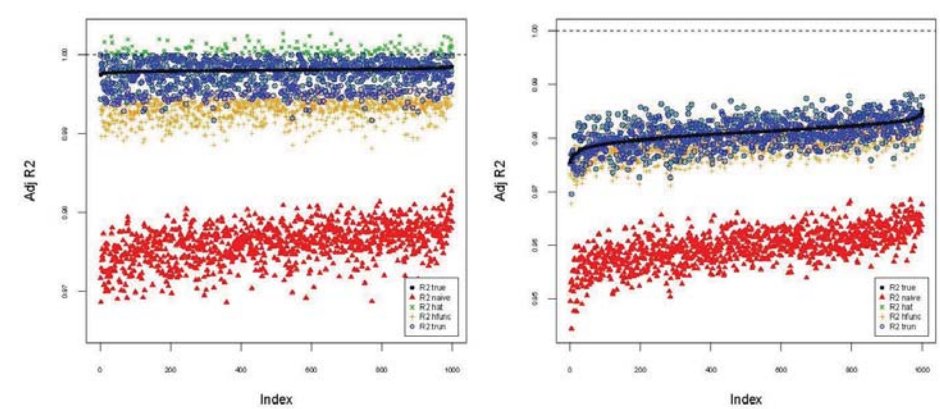
Table 1: Comparisons of different versions of adjusted R^2 varying by sample sizes

In all cases, adjusted R^2_{naive} underestimate the true adjusted R^2 and we can see that those values are much smaller than the true adjusted R^2 in particular in the cases of small sample sizes but our proposed adjusted R^2 (adjusted R^2_{hat}, adjusted R^2_{hfunc} and adjusted R^2_{trun}) can cut down the biases as we can see that our proposed adjusted R^2 are closer to the true adjusted R^2 than the naive ones in all cases. All versions of adjusted R^2 get closer to the true adjusted R^2 when sample size increases.

Next, we generate 1000 samples using data generation algorithm. For each sample, compute true adjusted R^2, naive adjusted R^2 and proposed adjusted R^2. Table 2 displays various percentiles of adjusted R^2_{true} and different approximations from 1000 experiments. Figures 1 and 2 show plots of different adjusted R^2 for all 1000 experiments. From Table 2 and Figures 1 and 2, our proposed adjusted R^2s approximate adjusted R^2_{true} better then adjusted R^2_{naive}. Since there is no situation when either $\widehat{MSE}_\theta$ or $\widehat{MST}_\theta$ is negative in cases (2) and (3), the adjusted R^2_{trun} are the same as adjusted R^2_{hat}. Therefore, the unbiased estimate adjusted R^2_{hat} seems to be the best estimate for the adjusted R^2_{true}. However, considering the left panel of Figure 1 for case (1) , there are some cases when the adjusted R^2_{hat} are greater than 1, therefore the $h-$ function transformation is needed and hence the adjusted R^2_{trun} is recommended for this case.

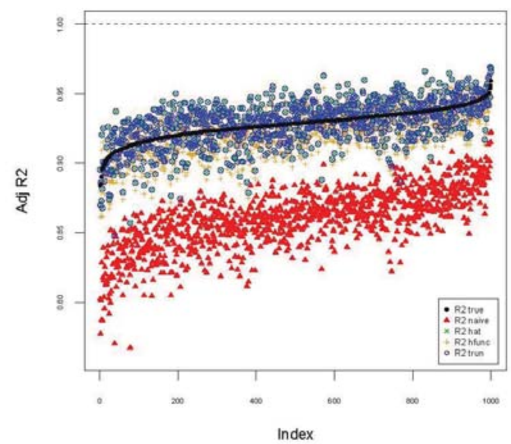
Percentiles	1	10	25	50	75	90	100
case (1): $\mathbf{V} = \text{diag}(v_1, \ldots, v_m)$ where $v_i \sim U(0.01, 0.1)$							
Adj R^2_{true}	0.9977	0.9978	0.9980	0.9981	0.9982	0.9983	0.9985
Adj R^2_{naive}	0.9708	0.9699	0.9770	0.9772	0.9775	0.9798	0.9788
Adj R^2_{hat}	0.9966	0.9948	0.9994	0.9985	0.9987	0.9990	0.9999
Adj R^2_{hfunc}	0.9913	0.9902	0.9941	0.9938	0.9939	0.9946	0.9948
Adj R^2_{trun}	0.9966	0.9948	0.9994	0.9985	0.9987	0.9990	0.9999
case (2): $\mathbf{V} = \text{diag}(v_1, \ldots, v_m)$ where $v_i \sim U(0.1, 1)$							
Adj R^2_{true}	0.9769	0.9789	0.9800	0.9811	0.9821	0.9828	0.9854
Adj R^2_{naive}	0.9513	0.9558	0.9628	0.9621	0.9666	0.9637	0.9676
Adj R^2_{hat}	0.9742	0.9776	0.9820	0.9831	0.9856	0.9839	0.9878
Adj R^2_{hfunc}	0.9725	0.9759	0.9803	0.9811	0.9837	0.9819	0.9854
Adj R^2_{trun}	0.9742	0.9776	0.9820	0.9831	0.9856	0.9839	0.9878
case (3): $\mathbf{V} = \mathbf{PP}'$ where $p_{ii} \sim U(1, 2)$ and $p_{ij}(i > j) \sim U(0.9, 1.1)$							
Adj R^2_{true}	0.8984	0.9146	0.9222	0.9288	0.9351	0.9404	0.9593
Adj R^2_{naive}	0.8154	0.8319	0.8498	0.8605	0.8771	0.9080	0.9224
Adj R^2_{hat}	0.8994	0.9139	0.9181	0.9238	0.9421	0.9603	0.9684
Adj R^2_{hfunc}	0.8914	0.9054	0.9121	0.9184	0.9352	0.9544	0.9630
Adj R^2_{trun}	0.8994	0.9139	0.9181	0.9238	0.9421	0.9603	0.9684

Table 2: Various percentile of different adjusted R^2 from 1000 experiments.



case $v_i \sim U(0.01, 0.1)$ case $v_i \sim U(0.1, 1)$

Figure 1: Plot of different adjusted R^2 for all 1000 experiments.



case $\mathbf{V} = \mathbf{PP}'$ where $p_{ii} \sim U(1,2)$ and $p_{ij}(i > j) \sim U(0.9, 1.1)$

Figure 2: Plot of different adjusted R^2 for all 1000 experiments.

4. Conclusion

In this research, we have examined the possibility of generalizing the bias reduction of variable selection criteria proposed in Lahiri and Suntornchost (2015) to general variance assumptions allowing for correlated regression errors. Several forms of adjusted R^2 are provided and asymptotic properties are justified. Simulation results show that the naive adjusted R^2 always underestimate the true adjusted R^2, our proposed adjusted R^2, R^2_{hat} reduces this underestimation. This simple adjustment works well except in some cases it exceeds the suitable range in which case adjustment by h-function, R^2_{hfunc}, is helpful. However, to accommodate all situations, the adjusted R^2_{trun} are recommended.

5. Reference

[1] Burnham KP, Anderson DR. Model selection and multimodel inference: a practical information-theoretic approach. 2nd ed. New York: Springer; 2002.

[2] Casella G, Berger RL. Statistical inference. 2nd ed. Pacific Grove, CA: Duxbury; 2002.

[3] Chatterjee S, Lahiri P. A simple computational method for estimating mean squared prediction error in general small-area model: Proceedings of the section on survey research methods; American Statistical Association Alexandria VA; 2007:3486-3493.

[4] Hurvich CM, Tsai CL. Regression and time series model selection in small samples. Biometrika. 1989:297-307.

[5] Lahiri P, Suntornchost J. Variable selection for linear mixed models with applications in small area estimation. Sankhya B. 2015; 77(2):312-20.

[6] Magnus JR, Neudecker H. Matrix differential calculus with applications in statistics and econometrics. 3nd Netherlands: John Wiley & Sons; 2007.
[7] Neath AA, Cavanaugh JE. The Bayesian information criterion: background, derivation, and applications. Wiley Interdisciplinary Reviews: Computational Statistics. 2012; 4(2):199-203.
[8] Rencher AC, Schaalje GB. Linear models in statistics. 2nd ed. Canada: John Wiley & Sons; 2008.