JOURNAL OF STATISTICAL COMPUTATION AND SIMULATION
https://doi.org/10.1080/00949655.2020.1838523

Optimum shrinkage parameter selection for ridge type estimator of Tobit model

Dursun Aydın, Öznur İşçi Güneri and Ersin Yılmaz

Department of Statistics, Faculty of Science, Mugla Sitki Kocman University, Mugla, Turkey

ABSTRACT

This paper presents different ridge type estimators based on maximum likelihood (*ML*) for parameters of a Tobit model. In this context, an algorithm is introduced to get the estimators based on *ML*. The most important issue in implementing these estimators is the selection of the optimum shrinkage parameter. Here attention is focused on the way in which the shrinkage parameter can be selected by six selection methods, including improved Akaike information criterion (AIC_c), Bayesian information criterion (*BIC*), generalized cross-validation (*GCV*), risk estimation using classical pilots (*RECP*), Mallows' (C_p) and $\hat{k}_{GM}$ proposed by Kibria [Performance of some new ridge regression estimators. Commun Stat Simul Comput. 2003;32:419–435]. Monte Carlo simulation experiments are performed and a real data example is presented to illustrate the ideas in the paper. Hence, an appropriate selection criterion or criteria are provided for optimum shrinkage parameter.

ARTICLE HISTORY
Received 1 May 2020
Accepted 14 October 2020

KEYWORDS
Maximum likelihood; censored data; ridge estimator; selection criteria

1. Introduction

Censored regression models are developed to describe the functional relationship between a dependent (or response) variable and a set of explanatory variables in which the response variable is subject to censoring. Formally, we assume that the basis of these models is a classical linear regression model with uncorrelated, normally distributed error terms,

$$z_i = \mathbf{x}_i'\boldsymbol{\beta} + \varepsilon_i,\ \varepsilon_i \sim N(0,\sigma^2),\ i = 1,2,\ldots,n \tag{1}$$

where z_is are the observations of the response variable, $\boldsymbol{x}_i = (x_{i1},\ldots,x_{ip})'$ is an $n \times 1$ vector containing the observations of p-dimensional explanatory variables, $\boldsymbol{\beta} = (\beta_1,\ldots,\beta_p)'$ is an $p \times 1$ vector of unknown regression parameters to be estimated, and ε_is should be normal random variables with a mean of zero and a common variance σ^2, as indicated in (1).

Note that the standard assumption of the error term implies that the values of the response variable in a regression model can be any real number; however, in many statistical applications, they are observed incompletely. In this case, the model (1) estimated

CONTACT Ersin Yılmaz yilmazersin13@hotmail.com Department of Statistics, Faculty of Science, Mugla Sitki Kocman University, Mugla, Turkey

by ordinary least squares (OLS) is improper since OLS regression leads to biased estimates. The error terms of this model are correlated and non-normally distributed. Consistent estimates in the case of censored data can be obtained by a censored regression model. Perhaps the most common example of such a regression model is the standard Tobit model [1]:

$$y_i = \max(0, z_i),\ z_i \sim N(\mathbf{x}_i' \boldsymbol{\beta}, \sigma^2 I_n) \tag{2}$$

where y_is are the observations of the censored response variable. One should note that, in practice, the values of z_i are unobserved, whereas the values of response y_i are observed due to left censoring. It should be emphasized that the model (2) is first discussed by Tobin [2] in economics. He analysed household expenditure (response variable) on durable goods (explanatory variables) across a year by considering that the values of the response variable could not be negative.

Note that if the explanatory variables are highly correlated, multi-collinearity becomes a serious problem, which can dramatically influence the effectiveness of a Tobit model estimated using the maximum likelihood (ML) method, as in the case of linear regression model. The collinearity results in a large variance and covariance of the parameter estimates and may lead to a lack of statistical significance of individual parameters. A common way to deal with this problem is to employ a ridge type regression estimator, originally proposed by Hoerl and Kennard [3]. In the literature, Tobit models are considered in different applications (see [1,4]). Note also that, there are important studies on different type of ridge estimators such as Roozbeh [5] studied about shrinkage ridge estimators under different error conditions Amini and Roozbeh [6] estimated partially linear model with ridge estimation under correlated errors. Roozbeh [7] proposed a modified estimator based on QR decomposition to overcome multicollinearity. In addition, Akdeniz and Roozbeh [8] and Roozbeh et al. [9] can be counted among them.

In this paper, various ridge type estimators are considered for estimating the parameters of a Tobit model with collinear data. The most important aspect of this problem is determining an optimum shrinkage parameter. The main objectives of this paper are therefore to select optimal ridge parameters, compare five selection methods that are AIC_c, BIC, GCV, $RECP$, C_p, and commonly used criterion $\hat{k}_{GM}$. Note that $\hat{k}_{GM}$ is a one of the commonly used plug-in method for estimating the ridge parameter proposed by Kibria [10]. As explained in Section 4, $\hat{k}_{GM}$ uses the geometric mean of $\hat{k}_i$ values calculated as the ratio of the model's error variance to the square of the estimates of the regression coefficients (i.e. $\hat{k}_i = \hat{\sigma}^2/\hat{\beta}_i^2$). Also, see the study of Kibria [10] for more detailed discussions. Thus, suitable ridge parameters can be found, and a good but parsimonious model fit can be obtained. For these purposes, the mentioned six criteria are inspected under simulated and real data settings. The value of the shrinkage parameter minimizing the information criteria corresponds to the optimum balance of model complexity and model fit. Furthermore, information criteria guide the process of making choices among various models. The basic idea is to find a useful selection criterion that provides a good estimation of a Tobit model based on the ML method which is given by Gajarado [11] and Khalaf et al. [12]. Due to shrinkage parameter selection criteria, a comparison of the different ridge type estimators is provided. In the literature, some selection methods for computing ridge parameters are discussed and compared by Mansson and Shukur [13] and also, Haq and Kibria [14], Kibria [10] can be counted as important studies to find optimal ridge parameter. However, the emphasis of this paper is on selection techniques based on information criteria rather than on selection methods. In the literature, Fang [15], Aydın et al. [16] and Yılmaz et al. [17] focused on using information criteria for the selection of the ridge parameter. Moreover, partial regression residual plots are used in evaluating whether we correctly specified the relationship between the dependent variable and the covariates. To the best of our knowledge, a study including ridge type estimators based on different selection criteria has not yet been conducted.

This paper is organized as follows. The estimation of a Tobit model based on maximum likelihood is examined, and an algorithm for calculating the Tobit ridge estimator is given in Section 2. In Section 3, statistical properties and characteristics of the estimator are discussed. Selection methods for finding optimum ridge parameter are explained in Section 4. Section 5 contains the Monte Carlo simulation study. In Section 6, gross domestic product (GDP) data is analysed with introduced method to see how it works using real-world data. Finally, conclusions and recommendations are presented in Section 7. Supplemental technical materials are found in the Appendix.

2. The ML estimation of a Tobit model

We consider the standard formulation for the Tobit model expressed in Equation (2). One should note that the response variable z can be considered as a partially latent variable whose values are concentrated at zero if they are negative. Hence, the model (2) can be rewritten as

$$y_i = \begin{cases} \boldsymbol{x}_i'\boldsymbol{\beta} + \varepsilon_i = z_i & \textit{if } z_i > 0 \\ 0 & \textit{if } z_i \leq 0 \end{cases} \tag{3}$$

Note that y_i and $\boldsymbol{x}_i$ are observed completely, but$z_i$is unobserved if it is not positive ($z_i \leq 0$) and is therefore a partially latent variable. It is clear from the definition of y_i in (3) that there are two cases to be considered: $y_i > 0$ and $y_i = 0$. Under the normality assumption of the error term ε, the first case shows that, if $y_i > 0$, we have the following conditional probability density function (pdf):

$$f(y_i|\boldsymbol{x}_i) = f(z_i|\boldsymbol{x}_i) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\frac{(y_i - \boldsymbol{x}_i'\boldsymbol{\beta})^2}{\sigma^2}\right] = \frac{1}{\sigma}\phi\left(\frac{y_i - \boldsymbol{x}_i'\boldsymbol{\beta}}{\sigma}\right) \tag{4}$$

It should be noted that the term $\phi(.)$ is the pdf of the standard normal distribution. However, the second case denotes that, if $y_i = 0$, we have the mass probability

$$\mathrm{P}(y_i = 0) = \mathrm{P}(z_i < 0) = \Phi\left(-\frac{\boldsymbol{x}_i'\boldsymbol{\beta}}{\sigma}\right) = 1 - \Phi\left(\frac{\boldsymbol{x}_i'\boldsymbol{\beta}}{\sigma}\right) = 1 - \Phi(v_i) \tag{5}$$

where $\Phi(.)$ shows the cumulative density function (cdf) of the standard normal distribution evaluated at $v_i = \mathbf{x}_i'\boldsymbol{\beta}/\sigma$, as defined above. According to the result of (4) and (5), we can then define the conditional pdf of the censored response variable y_i given $\boldsymbol{x}_i$ as

$$f(y_i|\boldsymbol{x}_i) = \{f(z_i|\boldsymbol{x}_i)\}^{d_i} \times \{P(y_i = 0|\boldsymbol{x}_i)\}^{1-d_i} = \left[\frac{1}{\sigma}\phi\left(\frac{y_i - \boldsymbol{x}_i'\boldsymbol{\beta}}{\sigma}\right)\right]^{d_i} \times \ [1 - \Phi(v_i)]^{1-d_i} \tag{6}$$

where d is a dummy variable equal to 1 if $z_i > 0$ and equal to zero otherwise. The likelihood function of the Tobit model expressed in (3) is then stated as

$$L(\boldsymbol{\beta},\sigma) = \prod_{i=1}^{n} f(y_i|\boldsymbol{x}_i) = \prod_{i=1}^{n}\left[\frac{1}{\sigma}\phi\left(\frac{y_i - \boldsymbol{x}_i'\boldsymbol{\beta}}{\sigma}\right)\right]^{d_i} \times [1-\Phi(v_i)]^{1-d_i} \tag{7}$$

This part of the paper focuses on the estimation of the unknown parameters $\boldsymbol{\beta}$ and σ in the standard Tobit (or censored) regression model. In the estimation sense, the ML method can be used to obtain consistent estimates of these parameters. Recall that response observations are censored on the right, that all cases discussed here fall into this framework, and that d is the indicator of censoring. In this case, for the standard Tobit model (3) with the normal error terms, the natural log-likelihood function is

$$l(\boldsymbol{\beta},\sigma) = lnL(\boldsymbol{\beta},\sigma) = \sum_{i=1}^{n}\left[\begin{array}{c} d_i ln\left[\frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{1}{2}\left(\frac{y_i - \boldsymbol{x}_i'\boldsymbol{\beta}}{\sigma}\right)^2\right)\right] \\ +(1-d_i)ln(1-\Phi(v_i)) \end{array}\right]. \tag{8}$$

The maximum likelihood estimators are the parameter values, say $\hat{\boldsymbol{\beta}}_{ML}$ and $\hat{\sigma}^2$, that maximize ln$L$ stated in (7) or, equivalently, $l(.)$ given in (8). Thus, the ML estimator $\hat{\boldsymbol{\beta}}_{ML}$ of the parameter vector $\boldsymbol{\beta}$ must satisfy

$$\frac{\partial lnL}{\partial\boldsymbol{\beta}} = \sum_{i=1}^{n}\left(d_i\left\{-\frac{1}{2}\frac{(y_i - \mathbf{x}_i'\hat{\boldsymbol{\beta}}_{ML})}{\sigma^2}\right\} + (1-d_i)\log\left(1-\Phi\left(\frac{\mathbf{x}_i'\hat{\boldsymbol{\beta}}_{ML}}{\sigma}\right)\right)\right)\mathbf{x}_i = 0 \tag{9}$$

The solution to Equation (9) gives the maximum likelihood estimator indicated by $\hat{\boldsymbol{\beta}}_{ML}$.

In most applications of regression, however, it seems that there is a nearly perfect linear relationship between the columns (covariates) of $\boldsymbol{X}$, and in such cases, the inferences based on the regression model can be misleading or erroneous. Moreover, where there is multi-collinearity, we know that the matrix $(\boldsymbol{X}'\boldsymbol{X})$ has one or more small eigenvalues. Hoerl and Kennard [3] proposed the ridge regression estimator to overcome this type of problem in linear regression analysis. In this paper, we generalize Hoerl and Kennard's [3] ridge estimator for the Tobit (or censored) regression.

2.1. Tobit ridge regression estimator

The presence of multi-collinearity has a number of potentially serious effects on the ML estimates of a Tobit model, as indicated in the previous section. Consequently, traditional methods proposed by Tobin [2] cannot be applied directly for estimating the parameter vector $\boldsymbol{\beta}$. To overcome this problem, we introduced a Tobit ridge regression estimator, which is obtained by modifying the ML estimator (see [11]).

Suppose that n_0 is the number of observations for $y_i = 0$, and n_1 is the number of observations for $y_i > 0$. For simplicity, let us first introduce some notations expressed in the following format [18]:

$\boldsymbol{y}_1 = (y_1, \ldots y_{n_1})'$ is a $(n_1 \times 1)$ vector of nonzero values for y_i

$\boldsymbol{X}_1 = (x_1, \ldots, x_{n_1})'$ is a $(n \times p)$ matrix entries of x_i for nonzero values on y_i

$$\boldsymbol{X}_0 = (x_{n_1+1}, \ldots, x_n)' \text{ is a } (n_0 \times p) \text{ matrix entreis of } x_i \text{ corresponding to } y_i = 0$$
$$\boldsymbol{\eta}_0 = (\eta_{n_1+1}, \ldots, \eta_n)' \text{ is a } (n_0 \times 1) \text{ vector values of } \eta_i \text{ corresponding to } y_i = 0$$

where elements of $\boldsymbol{\eta}_0$ can be obtained as $\eta_i = \frac{\phi(v_i)/\sigma}{1-\Phi(v_i)} = \left[\frac{1}{\sigma}\phi\left(\frac{x_i'\beta}{\sigma}\right)\right] / \left[1 - \Phi\left(\frac{x_i'\beta}{\sigma}\right)\right]$.

Using these notations, the ordinary likelihood function stated in (7) can be rewritten as

$$L = \prod_{i=1}^{n} \left[\frac{1}{\sigma}\phi\left(\frac{y_i - \boldsymbol{x}_i'\boldsymbol{\beta}}{\sigma}\right)\right]^{d_i} \times [1 - \Phi(v_i)]^{1-d_i}$$
$$= \prod_{1} \left[\frac{1}{\sigma}\phi\left(\frac{y_i - \boldsymbol{x}_i'\boldsymbol{\beta}}{\sigma}\right)\right] \times \prod_{0}[1 - \Phi(v_i)] \quad (10)$$

Here, it should be emphasized that the first terms of (10) are based on sample size n_1 for which $y_i > 0$, while the second terms are based on sample size n_0 for observations where $y_i = 0$. As previously indicated, the key idea is to estimate the parameters of the Tobit model by using a ridge regression based on ML. To achieve this, we added a penalty term to the likelihood function in (10), as in ordinary ridge regression. In light of these ideas for a given $k > 0$, the penalized ML criterion of (10) becomes

$$L^{pen} = L + \frac{k}{2}\boldsymbol{\beta}_2^2 \quad (11)$$

where $\frac{k}{2}\boldsymbol{\beta}_2^2$ is the penalty term for the ridge regularization and k is a ridge parameter. See Schaefer et al. [19] and Le Cessie and Van Houwelingen [20] for more detailed discussions about the ridge maximum likelihood. Penalized likelihood shrinks the ordinary likelihood estimates with penalty terms, and it solves the multi-collinearity problem with biased results.

The key idea is to obtain the Tobit ML ridge (MLR) estimator of the parameters in model (3), as previously denoted. For these purposes, we first obtain the natural logarithmic likelihood function of (11), given by

$$\ln L^{Pen} = \sum_{0} \ln(1 - \Phi(v_i)) + \sum_{1} \ln\left(\frac{1}{\sigma\sqrt{2\pi}}\right) - \sum_{1}\left(\frac{1}{2\sigma^2}(y_i - \boldsymbol{x}_i'\boldsymbol{\beta})^2\right) - \frac{k}{2}\boldsymbol{\beta}_2^2 \quad (12)$$

where $\sum_0(.)$ is over the values n_0 corresponding to $y_i = 0$, whereas $\sum_1(.)$ is over nonzero observations n_1 for y_i. One may note that (12) is obtained by adding a penalty term to (8). To maximize the likelihood in (12), we set its derivatives to zero. The first-order conditions for providing the Tobit MLR estimator of $\boldsymbol{\beta}$ are

$$\frac{\partial lnL^{pen}}{\partial \boldsymbol{\beta}} = -\sum_{0} \frac{\phi(v_i)\boldsymbol{x}_i}{1 - \Phi(v_i)} + \frac{1}{\sigma^2}\sum_{1}(y_i - \boldsymbol{x}_i'\hat{\boldsymbol{\beta}}_{MLR})\boldsymbol{x}_i - k\hat{\boldsymbol{\beta}}_{MLR} = 0 \quad (13)$$

Thus, after some algebraic manipulation, the Tobit MLR estimator is

$$\hat{\boldsymbol{\beta}}_{MLR} = \hat{\boldsymbol{\beta}}_R - \sigma(\boldsymbol{X}_1'\boldsymbol{X}_1 + k\boldsymbol{I_p})^{-1}\boldsymbol{X}_0'\eta_0 \quad (14)$$

where $\hat{\boldsymbol{\beta}}_R = (\boldsymbol{X}_1'\boldsymbol{X}_1 + k\boldsymbol{I_p})^{-1}\boldsymbol{X}_1\boldsymbol{y}_1$ is the ordinary ridge regression estimator using nonzero values of y_i. The implementation details of (14) are also provided in Appendix A1.

Note that the vector η_0 expressed in (14) depends on unknown parameters $\boldsymbol{\beta}$ and σ. Furthermore, (14) shows that the ordinary ridge regression estimates fail to capture the full effect of the covariates. Moreover, the estimator $\hat{\boldsymbol{\beta}}_{MLR}$ is also nonlinear in the parameters and therefore must be solved iteratively. In light of Fair [21], we introduced an algorithm based on iteration for obtaining the MLR estimator using (14). Details of the algorithm for fitting model (2) are given as follows.

2.1.1. Algorithm

Step 1: Compute initial guesses $\hat{\boldsymbol{\beta}}^{(0)} = \hat{\boldsymbol{\beta}}_R = (\boldsymbol{X}_1'\boldsymbol{X}_1 + k\boldsymbol{I_p})\boldsymbol{X}_1\boldsymbol{y}_1$ and $(\boldsymbol{X}_1'\boldsymbol{X}_1 + k\boldsymbol{I_p})\boldsymbol{X}_0'$.
Step 2: Choose a small positive number of σ, and denote this value by $\sigma^{(0)}$.
Step 3: Compute the vector $\eta_0^{(0)}$ using $\boldsymbol{\beta}^{(0)}$ and $\sigma^{(0)}$.
Step 4: Calculate the $\hat{\boldsymbol{\beta}}_{MLR}^{(0)}$ from (14) using $\eta_0^{(0)}$,$\boldsymbol{\beta}^{(0)}$, and $\sigma^{(0)}$.
Step 5: Determine the new estimate as the maximizer of (12), given by

$$\hat{\boldsymbol{\beta}}_{MLR}^{(1)} = \boldsymbol{\beta}^{(0)} + \lambda(\hat{\boldsymbol{\beta}}_{MLR}^{(0)} - \boldsymbol{\beta}^{(0)}), 0 < \lambda \leq 1. \tag{15}$$

where λ is a damping factor used in (15).
Step 6. Repeat Steps 2 and 5 until the iterations converge.

One should also note that the experiments in the simulation study showed that $\boldsymbol{\beta}^{(0)} = 0$ was a good starting value for the iterative procedure, although convergence was never guaranteed. However, if there are only a small number of censored values, $\boldsymbol{\beta}^{(0)} = \hat{\boldsymbol{\beta}}_R$ was a good initial value for vector $\boldsymbol{\beta}$. The parameter λ stated in step 5 is many times useful in such iterative processes to damp by taking λ to be less than one. Experience from simulation in this study shows that the magnitude of λ controls the jumps in each iteration. Therefore, the selection of the parameter λ is extremely important. In this context, some trials have been made for different values of λ such as $\lambda = 0.001, \lambda = 0.4$ and $\lambda = 1$. It seems that for $\lambda = 0.001$, the change in iterated estimates of β is really small, so the stability of the iteration process requires hundreds of repetitions. Since the amount of change is too large for a larger value of λ, the iteration for $\lambda = 1$ ends quickly, but does not give accurate predictions. As discussed in the study of Fair [21], Olsen [22] and Gajarado [11], it appears that the estimates provided with $\lambda = 0.4$ are numerically stabilized after 20 iterations. It is appeared that the estimates provided by $\lambda = 0.4$ numerically stabilized after 20 iterations, as discussed in the study of Fair [21], Olsen [22] and Gajarado [11].

3. Statistical properties of the MLR estimator

In this section, we summarize some properties of the MLR estimator $\hat{\boldsymbol{\beta}}_{\boldsymbol{MLR}}$ defined by (14). We know that the ridge estimator is a biased estimator, and this bias is proportional to the parameter$k$. Consequently, for a given $k > 0$, the Tobit MLR estimator expressed in (14) can be rewritten as

$$\hat{\boldsymbol{\beta}}_{\boldsymbol{MLR}} = (\boldsymbol{X}_1{}'\boldsymbol{X}_1 + k\boldsymbol{I_p})^{-1}[\boldsymbol{X}_1{}'\boldsymbol{X}_1\boldsymbol{\beta} - \sigma\boldsymbol{X}_0{}'\boldsymbol{\eta}_0] \tag{16}$$

and is abbreviated as

$$\boldsymbol{A}_k = (\boldsymbol{X}_1{}'\boldsymbol{X}_1 + k\boldsymbol{I_p})^{-1}. \tag{17}$$

The expected value, bias, and variance of the $\hat{\boldsymbol{\beta}}_{\boldsymbol{MLR}}$ estimator, are respectively given as follows:

$$E(\hat{\boldsymbol{\beta}}_{\boldsymbol{MLR}}) = \boldsymbol{A}_k(\boldsymbol{X}_1{}'\boldsymbol{X}_1\boldsymbol{\beta} - \sigma \boldsymbol{X}_0{}'\boldsymbol{\eta}_0) = \boldsymbol{\beta} - k\boldsymbol{A_k}\boldsymbol{\beta} - \boldsymbol{A}_k\sigma \boldsymbol{X}_0{}'\boldsymbol{\eta}_0 \tag{18}$$

$$[Bias(\hat{\boldsymbol{\beta}}_{MLR})] = \boldsymbol{A}_k\sigma \boldsymbol{X}_1{}'\boldsymbol{\eta}_0 + k\boldsymbol{A}_k\hat{\boldsymbol{\beta}}_{\boldsymbol{MLR}} \tag{19}$$

$$Var(\hat{\boldsymbol{\beta}}_{MLR}) = E\left(-\frac{\partial^2 logL^{pen}}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}'}\right) = (\boldsymbol{X}_1{}'\boldsymbol{R}\boldsymbol{X}_1 + k\boldsymbol{I}_p)^{-1} \tag{20}$$

where $\boldsymbol{R}$ is a $n \times n$ diagonal matrix, which is formed by its diagonal elements r_i values given by

$$r_i = -\frac{1}{\sigma^2}\left(v_i\phi(v_i) - \frac{\phi(v_i)^2}{1-\Phi(v_i)} - \Phi(v_i)\right), i = 1, \ldots, n$$

where $\phi(.)$ and $\Phi(.)$ denote the density and cumulative distribution function of standard normal distribution, respectively. The implementation details of Equations (19–20) can be found in Appendix A2.

The studies of Amemiya [18] and Van Wieringen [23] are helpful for understanding the variance of the parameters stated in (20). However, the expressions stated in Equations (18–20) are not directly usable since they depend on the unknown quantity σ^2. One, therefore, needs to determine an estimation for the variance σ^2. In a standard Tobit model, the estimate of this variance can be found by using residual sum of squares, as in OLS. Consequently, the estimate of the error variance is

$$\hat{\sigma}^2 = (\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}_{\boldsymbol{MLR}})'(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}_{\boldsymbol{MLR}})/n \tag{21}$$

It should be noted that the numerator of (21) represents the error terms arising from the measurements of ignored factors. The resulting estimator $\hat{\sigma}^2$ is also essentially biased. See Greene [24], [25] and Sun et al. [26] for the detailed asymptotic properties and bias of the estimator $\hat{\sigma}^2$.

3.1. Measuring the risk and efficiency

The bias stated in the previous section is only one criterion for evaluating the quality of an estimator. In general, the ill-effects of the deviation of $\hat{\boldsymbol{\beta}}_{\boldsymbol{MLR}}$ from $\boldsymbol{\beta}$ are referred to as the loss of information. Usually, the expected loss of an estimator $\hat{\boldsymbol{\beta}}_{\boldsymbol{MLR}}$ is measured by risk. This measurement is called the mean dispersion error (MDE). Our task is now to estimate the risk for the standard Tobit model. For convenience, we will work with the scalar-valued *MDE* matrix.

Definition 1: The risk is closely related to the matrix-valued *MDE* of an estimator $\hat{\boldsymbol{\beta}}_{\boldsymbol{MLR}}$ of the vector β. The scalar-valued version of the *MDE* matrix is defined as

$$SMDE(\hat{\boldsymbol{\beta}}_{MLR}, \boldsymbol{\beta}) = \sum_{j=1}^{p} E(\hat{\beta}_{MLRj} - \beta_j)^2 = tr\{MDE(\hat{\boldsymbol{\beta}}_{MLR}, \boldsymbol{\beta})\} \tag{22}$$

where $tr(\boldsymbol{A})$ denotes the trace of a matrix $\boldsymbol{A}$. It should be also noted that the first term on the right side of (22) is the squared error loss, described as $L = (\hat{\beta}_{MLRj} - \beta_j)^2$. In such a case, the risk of the estimator $\hat{\beta}_{MLRj}$, $E(\hat{\beta}_{MLRj} - \beta_j)^2$ is called *MDE*. It can also be given as follows:

$$\sum_{j=1}^{p} E(\hat{\beta}_{MLRj} - \beta_j)^2$$
$$= E(\hat{\boldsymbol{\beta}}_{MLR} - \boldsymbol{\beta}^2) = [Var(\hat{\boldsymbol{\beta}}_{MLR})] + [Bias(\hat{\boldsymbol{\beta}}_{MLR})]^T [Bias(\hat{\boldsymbol{\beta}}_{MLR})] = MDE. \quad (23)$$

This equation means that the *MDE* of an estimator is the sum of its variance and squared error. By applying (18–20), the *MDE* matrix stated in (23) can be rewritten as

$$MDE(\hat{\beta}_{MLRj}, \beta_j) = \sum_{j=1}^{p} E(\hat{\beta}_{MLR,j} - \beta_j)^2 = (\boldsymbol{X}_1{}'\boldsymbol{R}\boldsymbol{X}_1 + k\boldsymbol{I}_p)^{-1} + [\boldsymbol{A}_k \sigma \boldsymbol{X}_0{}' \boldsymbol{\eta}_0 + k\boldsymbol{A}_k \boldsymbol{\beta}]^2. \quad (24)$$

In other terms, the scalar-valued version of the *MDE* matrix (SMDE) in (22) can also be given by the following equation

$$SMDE(\hat{\boldsymbol{\beta}}_{MLR}, \boldsymbol{\beta}) = tr\{MDE(\hat{\boldsymbol{\beta}}_{MLR}, \boldsymbol{\beta})\}$$
$$= tr\{(\boldsymbol{X}_1{}'\boldsymbol{R}\boldsymbol{X}_1 + k\boldsymbol{I}_p)^{-1} + [\boldsymbol{A}_k \sigma \boldsymbol{X}_1{}' \boldsymbol{\eta}_0 + k\boldsymbol{A}_k \boldsymbol{\beta}]^2\}. \quad (25)$$

We can compare the quality of two estimators by looking at the ratio of their SMDE in (22) or (25). This ratio gives the following definition concerning the superiority of any two estimators.

Definition 3.2: The measure of the efficiency of an estimator $\hat{\beta}_{MLR1}$ relative to estimator $\hat{\beta}_{MLR2}$ is obtained by the ratio

$$RE(\hat{\beta}_{MLR1}, \hat{\beta}_{MLR2}) = \frac{R(\hat{\boldsymbol{\beta}}_{MLR2}, \boldsymbol{\beta})}{R(\hat{\boldsymbol{\beta}}_{MLR1}, \boldsymbol{\beta})} = \frac{SMDE(\hat{\beta}_{MLR2})}{SMDE(\hat{\beta}_{MLR1})} \quad (26)$$

where $R(.)$ denotes the scalar risk, which is also equivalent to (25). One should note that in comparing the efficiency of estimators, if $RE(\hat{\beta}_{MLR1}, \hat{\beta}_{MLR2}) > 1$, it can be said that $\hat{\beta}_{MLR1}$ is more efficient than $\hat{\beta}_{MLR2}$.

3.2. Asymptotic properties of MLR estimators

Equation (3) shows that the Tobit model uses only positive response values. Therefore, for positive values of y_i, the model can be written as follows

$$E[y_i | y_i \rangle 0] = \mathbf{x}_i{}'\boldsymbol{\beta} + E(\varepsilon_i | \varepsilon_i \rangle - \mathbf{x}_i{}'\boldsymbol{\beta}) = \mathbf{x}_i{}'\boldsymbol{\beta} + \sigma \frac{\phi(\nu_i)}{\Phi(\nu_i)}. \quad (27)$$

One can see that $E(\varepsilon_i | \varepsilon_i \rangle - \boldsymbol{x}_i'\boldsymbol{\beta})$ is obtained nonzero even if ε_i is not normally distributed. Consequently, one may say that when positive values of y_i are used, estimators will be biased in terms of Tobit model estimation. It should also be said, as indicated in (19), $\hat{\boldsymbol{\beta}}_{MLR}$ already has a bias term which is caused by a ridge penalty.

Goldberger [27] and Greene [24] evaluated the asymptotic bias for the OLS estimator of a Tobit model when data does not contain multicollinearity as

$$\lim_{n\to\infty} P\{|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}| > \delta\} = 0, \text{ which means } \hat{\boldsymbol{\theta}} \xrightarrow{p} \boldsymbol{\theta} \tag{28}$$

where $\boldsymbol{\theta} = (\boldsymbol{\beta}, \sigma^2)'$ and $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\beta}}, \hat{\sigma}^2)'$ for given $\delta > 0$. Note that convergence in (28) is valid when the below assumptions are ensured:

A1. *Elements of explanatory variables* $\boldsymbol{x}_i$ *normally distributed;*
A2. $\mathbf{x}_i$ *is independent from error terms* ε_i*.*

In this case, (3) can be written as follows

$$y_i = \begin{cases} \bar{\boldsymbol{x}}_i'\boldsymbol{\beta} + \varepsilon_i = z_i, \ if\ z_i > 0 \\ 0\ if\ z_i \le 0 \end{cases}$$

From assumptions A1 and A2, it follows that $\bar{\boldsymbol{x}}_i \sim N(0, \Sigma)$, distributed independently of ε_i. Accordingly, convergence can be written as follows

$$\hat{\boldsymbol{\beta}} \xrightarrow{p} \left[\frac{(1-\psi)}{(1-\omega^2\psi)}\right]\boldsymbol{\beta} \tag{29}$$

where

$$\psi = \frac{1}{\sigma_y} h\left(\frac{\beta_0}{\sigma_y}\right)\left(\beta_0 + \sigma_y h\left(\frac{\beta_0}{\sigma_y}\right)\right), \ h(\alpha) = \frac{\phi(\alpha)}{\Phi(\alpha)^2}$$

$$\omega^2 = \frac{1}{\sigma_y^2}\boldsymbol{\beta}'\Sigma\boldsymbol{\beta}, \sigma_y^2 = \sigma^2 + \boldsymbol{\beta}'\Sigma\boldsymbol{\beta}$$

Note that because of $0 < \psi < 1$ and $0 < \omega^2 < 1$, it can be shown that the OLS estimator of the Tobit model is biased. This was also proven by Goldberger [27] and Greene [24] as follows under the same assumptions of A1 and A2:

$$\hat{\boldsymbol{\beta}} \xrightarrow{p} \Phi(\beta_0/\sigma_y)\boldsymbol{\beta} \tag{30}$$

where $\hat{\boldsymbol{\beta}}$ is the vector of estimated regression coefficients by the OLS method for a Tobit regression of y_i on $\bar{x}_i$. Note that (30) is calculated using all observations of y_i not only positive observations. It can therefore be said that $\left(\frac{n_p}{n}\hat{\boldsymbol{\beta}}\right)$ is a consistent estimator of $\boldsymbol{\beta}$ where n_p is a number of positive response values. It should be noted that all of these inferences depend on the distribution of $\bar{x}_i$ values, which is assumed to be normal.

In addition, to show asymptotic properties of the ridge-based estimator $\hat{\boldsymbol{\beta}}_{MLR}$, some regularity conditions are given below:

C1. $\hat{\boldsymbol{\beta}}_{MLR}$ minimizes the penalized maximum likelihood function (12);

C2. The variance is a decreasing function when ridge parameter k is increasing (thus, when $k \to \infty$, the variance goes to zero);

C3. $[Bias(\hat{\boldsymbol{\beta}}_{MLR})]$ decreases together with ridge parameter k, which means for $k \to 0$, $[Bias(\hat{\boldsymbol{\beta}}_{MLR})]^T[Bias(\hat{\boldsymbol{\beta}}_{MLR})] \to 0$ where $[Bias(\hat{\boldsymbol{\beta}}_{MLR})]^T[Bias(\hat{\boldsymbol{\beta}}_{MLR})]$ is a continuous, monotonically increasing function of k (see [3]).

Under these conditions, asymptotic bias of $\hat{\beta}_{\mathrm{MLR}}$ can be expressed as

$$\lim_{n\to\infty}[Bias(\hat{\boldsymbol{\beta}}_{MLR})] = \lim_{n\to\infty}\{\boldsymbol{A}_k\sigma\boldsymbol{X}_1{}'\boldsymbol{\eta}_0 + k\boldsymbol{A}_k\hat{\boldsymbol{\beta}}_{\boldsymbol{MLR}}\} \tag{31}$$

Because of σ is dependent on sample size n which can be seen in Equation (21), when $n \to \infty$, (31) can be rearranged as

$$\lim_{n\to\infty}[Bias(\hat{\boldsymbol{\beta}}_{MLR})] = \lim_{n\to\infty}\{k\boldsymbol{A}_k\hat{\boldsymbol{\beta}}_{MLR}\} = k\boldsymbol{A}_k\hat{\boldsymbol{\beta}}_{MLR}$$

Thus, asymptotic bias of $\hat{\boldsymbol{\beta}}_{\boldsymbol{MLR}}$ is obtained as $k\boldsymbol{A}_k\hat{\boldsymbol{\beta}}_{MLR}$. In this case, one may say that this bias is dependent on ridge parameter k and from condition C3, it is heuristically said that when $k \to 0$, $[Bias(\hat{\boldsymbol{\beta}}_{MLR})] \to 0$. It is important to emphasize that because of the variance statement in C2, ridge parameter k has to be properly selected to provide balance between bias and variance.

4. Selection of the ridge parameter

There are various studies in the literature on choosing the ridge parameter k such as Hoerl et al. [28], Golub et al. [29], Pasha and Shah [30], and so on. However, although these researchers obtained reasonable results in choosing a ridge parameter, there are no absolute rules. In this paper, to find optimum ridge parameter k for estimating the Tobit ridge regression model, AIC_c, GCV, BIC,$RECP$, and Mollow's C_p criteria were used and performances were compared with both each other and $\hat{k}_{GM}$, as originally proposed by Kibria [10]. $\hat{k}_{GM}$ has also been used by Khalaf et al. [12], and it has given satisfying results under certain conditions. This study uses it as a benchmark method. Our real purpose is to see the effect of information criteria on selection of the ridge parameter. The six criteria that were used in this paper are explained as follows.

***AIC*$_c$ Criterion:** It was proposed by Hurvich et al. [31] to make classical Akaike information criterion robust for small sample sizes. Calculation of AIC_c is given by

$$AIC_c(k) = 1 + \log[\boldsymbol{y} - \hat{\boldsymbol{y}}^2/n] + [2\{p+1\}/n - p - 2] \tag{32}$$

where $\hat{\boldsymbol{y}} = \boldsymbol{X}\hat{\boldsymbol{\beta}}_{MLR}$, and p denotes the number of regression parameters in the Tobit model.

***BIC* Criterion:** Schwarz [32] proposed the BIC criterion by using Bayes estimators. The BIC criterion is

$$BIC(k) = \frac{1}{n}\boldsymbol{y} - \hat{\boldsymbol{y}}^2 + \left(\frac{\log(n)}{n}\right)p. \tag{33}$$

***GCV* Criterion:** It was developed by Craven and Wahba [33] and is calculated as follows

$$GCV(k) = n^{-1}\boldsymbol{y} - \hat{\boldsymbol{y}}^2/[n^{-1}p]^2 \tag{34}$$

***RECP* Criterion:** This is a risk estimation using classical pilots (RECP) is used pilot selection of ridge parameter k, k_p, computes $\hat{\boldsymbol{y}}_{k_p}$ and $\hat{\sigma}_p^2$, and then measures the risk between $\hat{\boldsymbol{y}}_{k_p}$ and $\boldsymbol{y}$. To choose pilot, k_p can be selected using one of the classical methods (see [34]). The RECP score is defined as

$$RECP(k) = 1/n\{\boldsymbol{y} - \hat{\boldsymbol{y}}_{k_p}{}^2 + \hat{\sigma}_{k_p}^2 p^2\} = 1/nE\boldsymbol{y} - \hat{\boldsymbol{y}}_{k_p}{}^2. \tag{35}$$

$\boldsymbol{C_p}$ Criterion : Mallows [35] proposed the $\boldsymbol{C_p}$ criterion for calculating the MDE in (23) scaled by $\hat{\sigma}^2$. The criterion can be obtained as follows

$$C_p(k) = 1/n\{\boldsymbol{y} - \hat{\boldsymbol{y}}^2 + 2\sigma^2 p - \sigma^2\} = 1/n\{\boldsymbol{y} - \hat{\boldsymbol{y}}^2 + 2\sigma^2 p - \sigma^2\}. \tag{36}$$

In practice, σ^2 is generally unknown. In this case, it has to be estimated with Equation (21). For details on the C_p, see Mallows [35] and Liang [36].

The $\hat{\boldsymbol{k}}_{\boldsymbol{GM}}$ method: The $\hat{\boldsymbol{k}}_{\boldsymbol{GM}}$ was proposed by Kibria [10] to select a ridge parameter, and Muniz and Kibria [37] made an extensive empirical study by using a number of ridge estimators including the $\hat{k}_{GM}$ given by

$$\hat{k}_{GM} = \hat{\sigma}^2 / \left\{ \left(\prod_{i=1}^{p} \hat{\beta}_i^2 \right)^{1/p} \right\} \tag{37}$$

where $\hat{\sigma}^2$ is calculated as in (21), and p is a number of the parameters.

5. Simulation study

This section reports the outcomes from a Monte Carlo simulation study, and it is designed to realize the main goal of this paper, comparing the performance of different ridge estimators. We therefore wish to find a good estimate of a ridge parameter and a suitable estimator obtained by this ridge parameter simultaneously. Since the degree of multi-collinearity among the covariates is of core importance, we first generated the correlated covariates using the following equation (see [10]):

$$x_{ij} = \sqrt{(1-\rho^2)}t_{ij} + \rho t_{ij},\ j = 1, \ldots, p \tag{38}$$

where p is the number of covariates, t_{ij} is the standard normal distribution, and $\rho = (0.80, 0.90, 0.99)$ denotes the three correlation levels between any two covariates. The observations of the latent response variable are constructed by

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \varepsilon_i,\ i = 1, 2, \ldots, n \tag{39}$$

where ε_i is a random variable from the normal distribution with a mean of zero and constant variance [i.e., $\varepsilon_i \sim N(0, \sigma^2)$], $\beta_0 = 1.5$, $\beta_1 = -2$, $\beta_2 = 0.7$, and $\beta_3 = 2.5$. Finally, the response variable in (39) is censored by using (3). Note also that the censoring rate (C.R.) is determined by a random Bernoulli distribution with probabilities at ratios specified in Table 1.

In addition, to what has been said above, a number of other factors can affect properties of ridge type estimators. The aforementioned factors in this simulation are things such

Table 1. Numerical values of some factors in the simulation setup.

Effective factors	Notation	Simulation design
The number of replications	$R.N$	1000
Sample size	n	35, 50, 150, 300
Censoring rates	$C.R$	5%, 40%
Correlation levels	ρ	0.80, 0.90, 0.99
Variance of the errors	σ^2	0.3, 1, 3

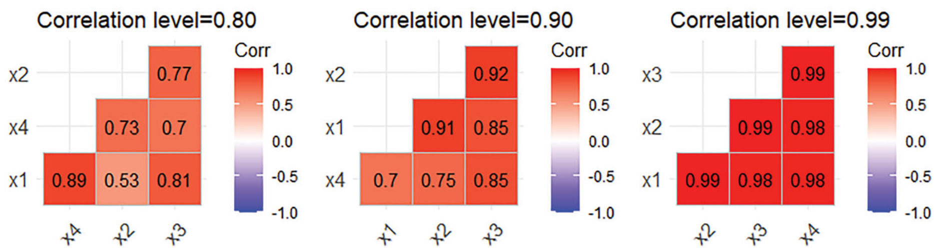

Figure 1. Correlation plots for different levels.

as sample size, the distribution of the error terms, correlation level, censoring rates, and the number of replications for each sample. For completeness, some specifications of the simulation setup are listed in Table 1.

We also examined three correlation levels, as shown in Table 1. For example, if $\rho = 0.80$, this allowed us to have about the same correlation level between all pairs of explanatory variables, as shown in Figure 1. Moreover, this case showed us that the eigenvalues of the $(\boldsymbol{X}'\boldsymbol{X})$ matrix end up being very large, which caused severe multi-collinearity in the explanatory variables, as displayed in the Figure 2.

5.1. Evaluation of the empirical results

The comparative outcomes of the Monte Carlo simulation experiments are summarized in the following figures and tables. It should be emphasized that in this simulation, many configurations were used to provide some intuition on the adequacy of the above ridge type estimators based on different selection criteria. Because 72 different simulation versions were examined, it is not possible to illustrate the details of each version. Therefore, a selection of the simulation results, performed under varying conditions, is given in the following sections.

Figures 3 and 4 display the box plots constructed by the biases of Tobit ridge regression estimates $\hat{\boldsymbol{\beta}}_{MLR}$ from model (39). As shown in these figures, when sample size n increases, the range for the biases of the estimates becomes narrower, as expected. The biases of the coefficient estimates from medium and large samples (i.e. $n = 50$ and $n = 200$) are also more stable than those from small samples. Figures 3 and 4 also compare the shrinkage parameter selection criteria on censored data. The biases of the estimates from the criteria on simulated data sets with censoring levels of 5% are given in Figure 3. Compared to Figure 4, the general trend shows that as the censoring level increases, the range of the biases increases. Hence, censoring rates are far more efficient on sample sizes.

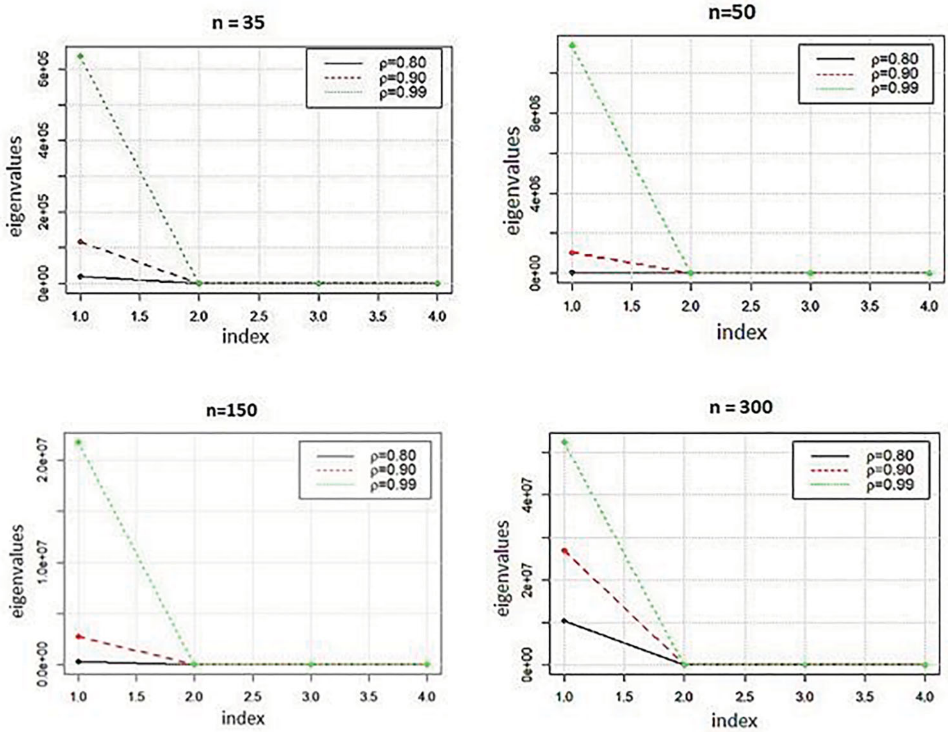

Figure 2. Scatter plot of eigenvalues from the $(\boldsymbol{X}'\boldsymbol{X})$ matrix.

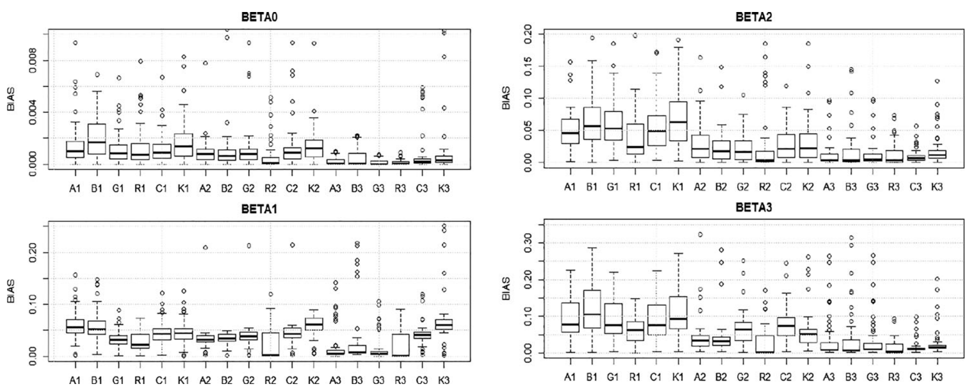

Figure 3. Boxplots of the biases from 1000 runs under simulation design, $\rho = 0.80, \sigma^2 = 0.3$, and *C.R.* $= 5\%$. Upper panel: A1, A2, and A3, the boxplots of the replications of the biases of $\beta_0 = 1.5$ from Tobit ridge estimates based on the *AICc* criterion are constructed using the algorithm defined in Section 2.1 for sample sizes of $n = 35, 50$, and 200, respectively. In a similar way, B1, B2, and B3 show the boxplots of the biases replications based on the *BIC*, G1, G2, and G3 denote *GCV*, R1, R2, and R3 define the *RECP*, C1, C2, and C3 represent the *Cp* and K1, K2, and K3 indicate the $\hat{k}_{GM}$ method. From top to bottom, the remaining panels are the same as the first panel except for $\beta_1 = -2$, $\beta_2 = 0.7$, and $\beta_3 = 2.5$, respectively.

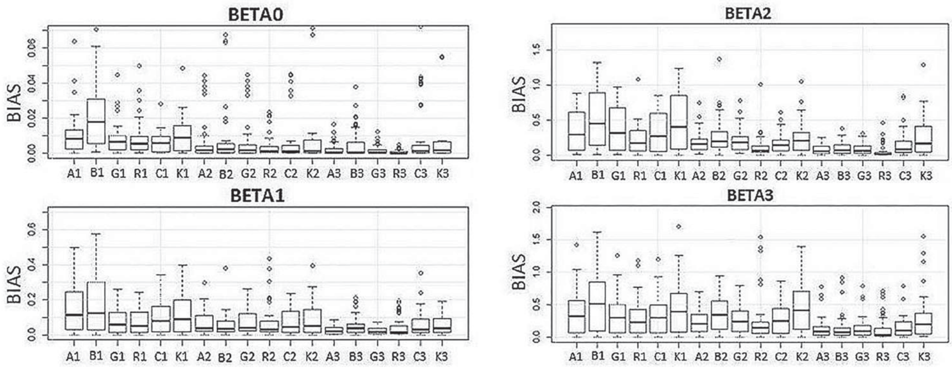

Figure 4. Similar to Figure 3 but for simulation design, $\rho = 0.99, \sigma^2 = 1$, and $C.R. = 40\%$.

Table 2. The outputs from the Tobit ridge estimators based on different criteria of parameters in model (39) with censored data for $\rho = 0.80, \sigma^2 = 1$, and $n = 50$.

		$C.R. = 5\%$				$C.R. = 40\%$			
Criteria	Summary Statistics	β_0	β_1	β_2	β_3	β_0	β_1	β_2	β_3
AICc	*Est*	1.68	−1.78	0.79	2.82	1.38	−2.07	0.59	2.79
	SD	0.06	0.2	0.24	0.2	0.27	0.77	0.74	0.78
	MDE	**0.04**	0.09	0.07	0.07	0.09	0.59	0.56	0.65
BIC	*Est*	1.87	−1.79	0.82	2.68	1.63	−1.58	1.09	3.3
	SD	0.09	0.3	0.36	0.25	0.31	1.07	1.04	1.09
	MDE	0.15	0.13	0.14	0.16	0.11	1.32	1.22	1.28
GCV	*Est*	1.78	−1.78	0.79	2.78	1.47	−2.02	0.64	2.84
	SD	0.05	0.17	0.2	0.16	0.26	0.67	0.65	0.67
	MDE	0.08	0.08	0.05	0.07	0.07	0.44	0.42	0.48
RECP	*Est*	1.69	−2.03	0.73	2.94	1.51	−1.89	0.72	2.78
	SD	**0.04**	**0.13**	**0.16**	**0.14**	**0.25**	**0.59**	**0.57**	0.7
	MDE	**0.04**	**0.02**	**0.03**	**0.02**	**0.06**	**0.36**	**0.33**	0.53
Cp	*Est*	1.68	−1.78	0.79	2.82	1.38	−2.07	0.59	2.79
	SD	0.06	0.2	0.24	0.2	0.27	0.77	0.74	0.78
	MDE	**0.04**	0.09	0.07	0.07	0.09	0.59	0.56	0.65
$\hat{k}_{GM}$	*Est*	1.87	−1.79	0.82	2.68	1.63	−1.58	1.09	3.3
	SD	0.09	0.3	0.36	0.25	0.31	1.07	1.04	1.09
	MDE	0.15	0.13	0.14	0.16	0.11	1.32	1.22	1.28

Note: Bold values denote the best scores.

The fits of the Tobit regression model (39) via MLR based on different selection criteria are summarized in Tables 1 and 2. These tables give the estimates of the parameters from the Tobit model, their averaged standard errors (*SDs*), and MDE values defined in (23) for each criterion. It should be noted that the rows labelled '*Est*' give the estimate vector $\hat{\boldsymbol{\beta}}_{MLR}$ defined in (14). The next rows marked '*SD*' denote the standard errors calculated based on $\hat{\sigma}^2$ in (21) and the root squares of diagonal elements of the matrix $Var(\hat{\boldsymbol{\beta}}_{MLR})$ given at (20). The rows marked '*MDE*' indicate the risk estimates related to the estimators. In these tables, Tobit ridge estimators are computed by optimum shrinkage parameter, which is selected with criteria considered here.

Table 3. Similar to Table 1 but for $\rho = 0.99, \sigma^2 = 3,$ and $n = 300$.

Criteria	Summary Statistics	C.R. = 5%				C.R. = 40%			
		β_0	β_1	β_2	β_3	β_0	β_1	β_2	β_3
AICc	*Est*	1.55	−2.3	1.28	3.35	1.68	−1.75	0.63	2.6
	SD	0.23	0.89	0.79	0.92	0.61	0.97	0.95	0.98
	MDE	0.06	0.87	0.95	0.96	0.4	1	0.91	1.12
BIC	*Est*	2.05	−2.15	0.99	3.42	1.28	−1.72	1.02	3.11
	SD	0.34	1.5	1.25	1.05	0.92	1.63	1.58	1.64
	MDE	0.42	2.26	1.65	1.29	0.89	2.74	2.6	2.7
GCV	*Est*	1.6	−2.24	1.14	3.16	1.43	−1.89	0.68	2.85
	SD	0.19	0.64	0.77	0.86	0.51	0.77	0.76	0.78
	MDE	**0.05**	0.46	0.79	0.77	0.27	0.61	0.58	0.63
RECP	*Est*	1.65	−2.12	1.13	3.08	1.43	−1.82	0.78	2.72
	SD	**0.16**	**0.57**	**0.67**	**0.76**	**0.43**	**0.63**	0.73	**0.57**
	MDE	**0.05**	**0.33**	**0.63**	**0.58**	**0.19**	**0.43**	0.54	**0.4**
Cp	*Est*	1.68	−2.28	1.01	3.18	1.41	−1.86	0.66	2.83
	SD	0.17	0.91	0.89	0.88	0.47	0.69	**0.68**	0.7
	MDE	0.06	0.91	0.89	0.81	0.23	0.5	**0.46**	0.52
$\hat{k}_{GM}$	*Est*	2	−2.38	1.05	3.36	1.63	−2.04	0.63	3.13
	SD	0.32	1.34	1.2	1.39	0.86	1.36	1.33	1.28
	MDE	0.35	1.94	1.56	2.05	0.76	1.85	1.77	1.66

Note: Bold values denote the best scores.

Notes in Tables 2 and 3 show that average SD and MDE values for $\hat{k}_{GM}$ and *BIC* selection methods are larger than those of the other selection criteria for almost all of the parameter estimates when the censoring level is sufficiently low. The *RECP* criterion has chosen a better estimate than the other four criteria and is a benchmark method for almost all experiments. In addition, the estimates obtained by C_p and *GCV* seem more reasonable. If results that obtained under heavy censorship are inspected, *RECP* still has the best performance, but the *BIC* method does not give satisfying results.

The scores in these Tables also prove that *BIC* and $\hat{k}_{GM}$ are highly sensitive to censorship. Therefore, it is not a suitable criterion for ridge parameter selection in a Tobit ridge regression. The following outcomes also support that argument. However, *GCV*, *RECP*, and C_p are relatively more resistant to censorship than the other criteria. C_p in particular gives better scores at a high censoring level. In terms of correlation levels, $\hat{k}_{GM}$ had the worst performance, and *BIC* was again second worst. Furthermore, one may note that for lower correlation levels, although AIC_c and *BIC* produce similar results, AIC_c always chooses a better ridge parameter than *BIC*. Consequently, inferences and comments based on *BIC* are also often valid for the AIC_c criterion.

To gain some further insight into the above ideas, the estimated SMDE values from estimators are also tabulated in Table 4 for only the highest value of variance $\sigma^2 = 3$. The other outcomes from different simulation configurations are similar, and they are not reported here. The SMDE values clearly provide evidence in support of the claims in the Table 3. Note also that when the variance of errors (*i.e.*, σ^2) increased, the SMDE values increased, as expected. In general, the Tobit ridge estimator based on the *RECP* criterion outperforms the others in terms of providing smaller SMDE values.

When dealing with the shrinkage parameter selection problem, a key problem is having a good perspective into bias and variance of the estimators since a balance between these two measurements forms the core of many parameter selection criteria. Therefore,

Table 4. Average *SMDE* values from the estimators based on different criteria for $\sigma^2 = 3$.

		C.R = 5%						C.R = 40%					
ρ	n	AIC_c	BIC	GCV	$RECP$	C_p	$\hat{k}_{GM}$	AIC_c	BIC	GCV	$RECP$	C_p	$\hat{k}_{GM}$
0.8	35	0.79	0.86	0.76	**0.673**	0.76	0.98	0.96	1.27	0.87	**0.834**	0.86	1.28
	50	0.7	0.74	0.68	**0.588**	0.7	0.88	0.82	0.99	0.76	**0.636**	0.74	1.06
	150	0.52	0.53	0.52	**0.489**	0.53	0.61	0.63	0.66	0.61	**0.507**	0.68	0.85
	300	0.53	0.53	0.53	**0.507**	0.53	0.59	0.557	0.56	0.55	**0.502**	0.61	0.74
0.9	35	1.06	1.3	0.99	**0.816**	0.92	1.12	1.033	1.47	0.91	1.228	**0.83**	1.33
	50	0.75	0.86	0.72	**0.663**	0.69	0.96	0.873	0.99	**0.81**	0.881	0.9	1.42
	150	0.57	0.58	0.56	**0.493**	0.57	0.75	0.728	0.84	0.69	**0.621**	0.76	1.26
	300	0.46	0.47	0.46	**0.42**	0.46	0.45	0.633	0.66	0.62	**0.59**	0.66	0.95
0.99	35	1.44	1.98	1.3	**1.065**	1.22	1.62	1.675	2.13	1.4	**1.094**	1.38	1.87
	50	1.43	1.95	1.29	**1.081**	1.22	1.6	1.575	2.02	1.33	1.628	**1.26**	1.7
	150	1.13	1.43	1.04	1.204	**1**	1.37	1.246	1.77	1.1	1.019	**1.1**	1.61
	300	1.01	1.2	0.95	1.035	**0.91**	1.21	1.062	1.36	1.08	1.109	**1.06**	1.42

Note: Bold values denote the best scores.

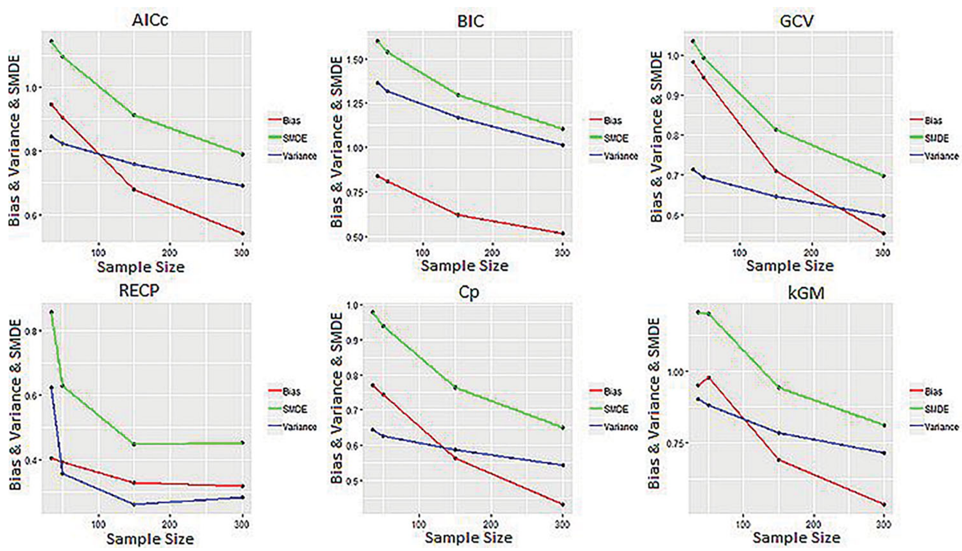

Figure 5. Bias-Variance decomposition plot for $\rho = 0.90, C.R. = 5\%, \sigma^2 = 3$.

Figures 5 and 6 represent bias-variance decomposition for six criteria in terms of SMDEs. Both figures were obtained for different designs. The figures clearly show that the Tobit ridge estimator is a biased estimator, which is most evident in Figure 5. The estimates in Figure 6 have more variance due to the high censoring levels. When the y-axis of the figures is inspected, it appears that the estimators obtained using information criteria have smaller SMDE, bias, and variance values than the $\hat{k}_{GM}$ benchmark. These figures also prove that *RECP* provides satisfying performance for this simulation study.

5.2. Comparing the efficiency

In order to illustrate and compare the efficiency of the selection methods based on correlated data, relative efficiency values are constructed from the SMDE values. Different

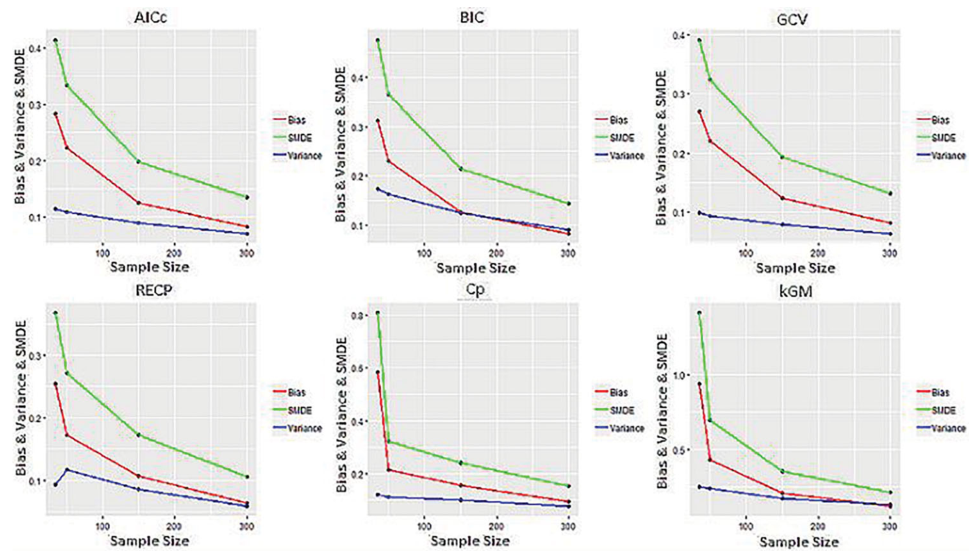

Figure 6. Similar to Figure 5 but for $\rho = 0.99$, $C.R. = 40\%$, $\sigma^2 = 1$.

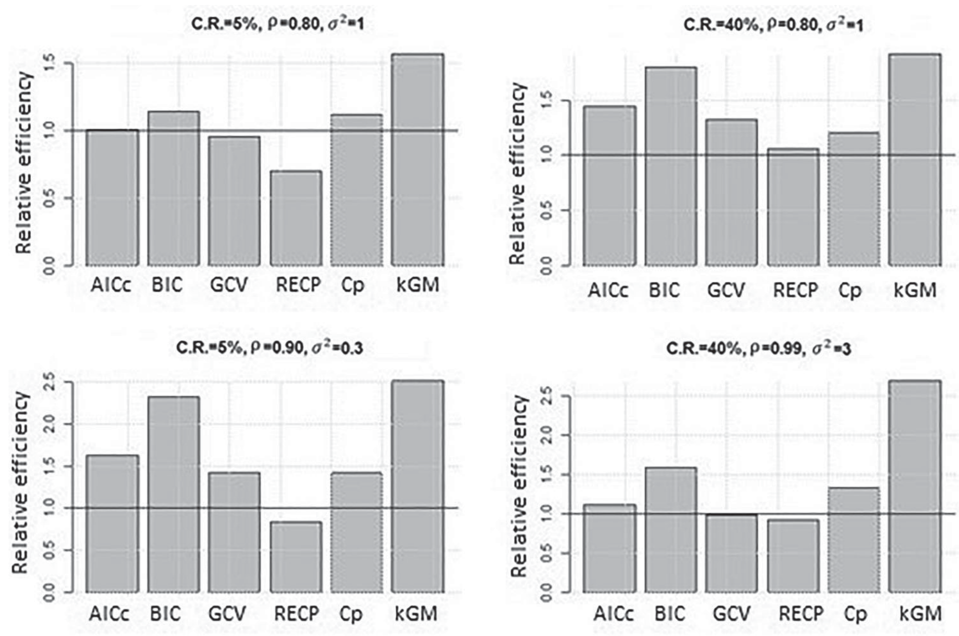

Figure 7. Bar-plots for relative efficiencies.

combinations are shown in Figure 7. As shown in Figure 7, relative efficiencies of *RECP* are better than others for all combinations, which can be crosschecked with Table 2. This case shows that *RECP* is more efficient than the other criteria, especially for highly correlated data.

When the bottom-right panel of Figure 7 is inspected, it reveals that the *RECP* criterion had the best efficiency rates for highly correlated data ($\rho = 0.99$) and a large variance of the error terms (*i.e.*, $\sigma^2 = 3$). This implies that the *RECP* criterion provides an optimal Tobit ridge estimator for fitting penalized ML criterion and the standard Tobit regression model discussed here. Moreover, *RECP* seems to work well in all simulation configurations. The bar plots displayed in Figure 7 indicate that AIC_c, GCV, and C_p had approximately the same performance due to the effect of replications. Note that their performances were better than those of the *BIC* and $\hat{k}_{GM}$ methods. Lastly, the $\hat{k}_{GM}$ method performed quite poorly in this study.

6. Real data example

In this section, a real data set was used to compare the performances of the Tobit ridge type estimators based on information criteria, which were used for selecting the ridge parameter. Gross domestic product per capita data obtained from Turkey was used and is accessible at https://data.worldbank.org. This data set contains eight variables, and each variable consists of 58 observations. The five most important variables affecting the GDP per capita (*gdppc*) are the percentage of import and exported goods (*impexp*), the population growth rate (*poprate*), the percentage of industrial production (*indstry*), the percentage of agricultural production (*agrclt*), and military spending (*miltry*). Hence, we used the regression model

$$gdppc_i = \beta_1(impexp_{i1}) + \beta_2(poprate_{i2}) + \beta_3(indstry_{i3}) + \beta_4(agrclt_{i4}) + \beta_5(miltry_{i5}) + \varepsilon_i,\ i = 1, \ldots, 58 \quad (40)$$

to determine GDP per capita data.

Collinearity was checked by simply calculating the correlations of the covariates stated in (40). Let $\boldsymbol{X}$ be a (58×5) matrix of the levels of the predictors in our real data example. The density of *gdppc* and the correlation plot of explanatory variables are displayed in Figure 8, which allows us to examine the relationship among the explanatory variables at the first stage.

The left panel of Figure 8 shows the density of variable *gdppc*. Note that density is important for this dataset because it allows us to visualize the censored observations. As seen from

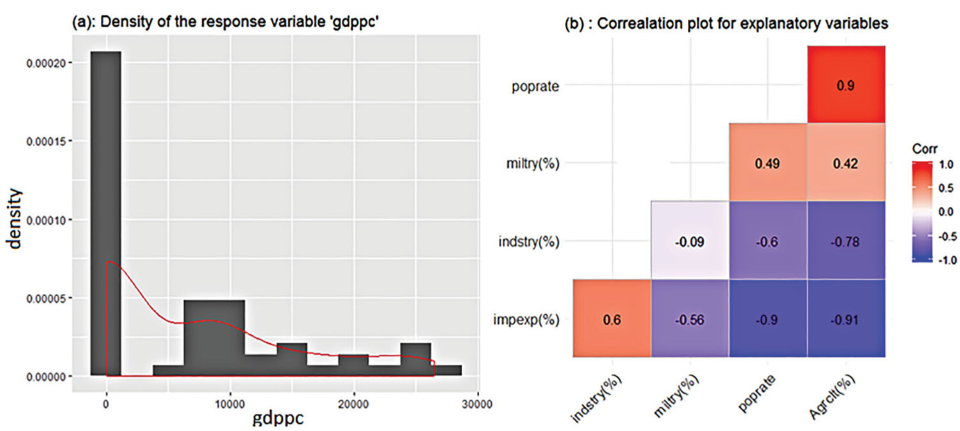

Figure 8. Descriptive plots for *GDP* per capita data modelled by Tobit-ridge regression.

the *gdppc*-axis, the dataset considered here is the left-censored, and these censored observations are indicated with zero. Note that *gdppc* values are not zero, because of they are incompletely observed, they take zero value which is a part of the Tobit methodology (see [2]). The right panel in this figure shows the correlations among the explanatory variables. Since some covariates are highly correlated, there is potential multi-collinearity in this real data set. Consequently, it is not possible to analyse this dataset with a classical regression model or a classical Tobit model.

A very simple measure of multi-collinearity can be provided by inspecting the characteristic roots or eigenvalues (say $\lambda_1, \ldots \lambda_k$) of $(\boldsymbol{X}'\boldsymbol{X})$. One or more small eigenvalues mean that there is collinearity among the columns of matrix $\boldsymbol{X}$. The measure most commonly used to detect multi-collinearity is the condition index (CI), which is computed as the ratio of minimum and maximum eigenvalues of $(\boldsymbol{X}'\boldsymbol{X})$, as given in (41). The eigenvalues of the $(\boldsymbol{X}'\boldsymbol{X})$ are $\lambda_1 = 106787, \lambda_2 = 21775.61, \lambda_3 = 1048.66, \lambda_4 = 20.43,$ *and* $\lambda_5 = 1.88$, respectively. Hence, for the GDP per capita data set, the CI is defined as $CI = \sqrt{[\lambda_{max}(\boldsymbol{X}'\boldsymbol{X})/\lambda_{min}(\boldsymbol{X}'\boldsymbol{X})]} = 238.33$ (41).

Since the value of CI exceeds 30, we must conclude that there is a strong collinearity problem in this data set [38]. To overcome the collinearity and the left- censored data simultaneously, a Tobit ridge estimator was used, which has been expressed in the previous section. To realize our purpose, the ridge parameter was chosen by $AIC_c, GCV, BIC, C_p, RECP$, and $\hat{k}_{GM}$ methods, respectively. The outcomes from the real data are summarized in the following table and figure.

Table 5 presents the Tobit ridge regression results using GDP per capita dataset for each criterion. In this table, the rows marked '*Est*' denote the estimated values of the regression coefficients. The next rows labelled '*SD*' indicate the standard deviations of the estimated coefficients. Note that the column marked '*SMDE*' provides the values of SMDE as defined in (22), whereas the column labelled '$Var(\varepsilon)$' shows the estimated variances of the error terms stated in (40). Important scores are indicated in bold. If one examines Table 5 carefully, one sees that *RECP* has smaller standard deviations and SMDE scores compared to other criteria. In addition, AIC_c, C_p, and *GCV* have provided the next-best results after *RECP*. Although the *BIC* method yields some small bias values, it has relatively large standard deviations for the regression coefficients and the largest SMDE value in comparison

Table 5. Outcomes from the Tobit ridge estimators based on different criteria.

Criteria	Summary Statistics	β_1	β_2	β_3	β_4	β_5	*SMDE*	$Var(\varepsilon)$
AICc	*Est*	0.236	10.959	0.306	−0.489	−4.868	5.703	2.876
	SD	0.089	3.263	0.033	0.302	0.842		
BIC	*Est*	0.224	6.175	0.377	−1.581	−3.691	9.058	3.84
	SD	0.057	4.257	0.245	0.149	0.071		
GCV	*Est*	0.236	10.937	0.307	−0.289	−4.867	4.155	2.377
	SD	0.053	2.059	0.023	0.197	0.642		
RECP	*Est*	0.255	8.789	0.942	−0.537	−4.69	**3.94**	**2.061**
	SD	**0.006**	**1.202**	**0.021**	**0.014**	**0.034**		
Cp	*Est*	0.25	7.55	0.451	−0.478	−4.692	5.426	3.261
	SD	0.007	1.205	0.028	0.015	0.04		
$\hat{k}_{GM}$	*Est*	0.207	13.205	0.254	−0.634	−4.718	6.425	3.893
	SD	0.107	3.7	0.229	0.41	1.316		

Note: Bold values denote the best scores.

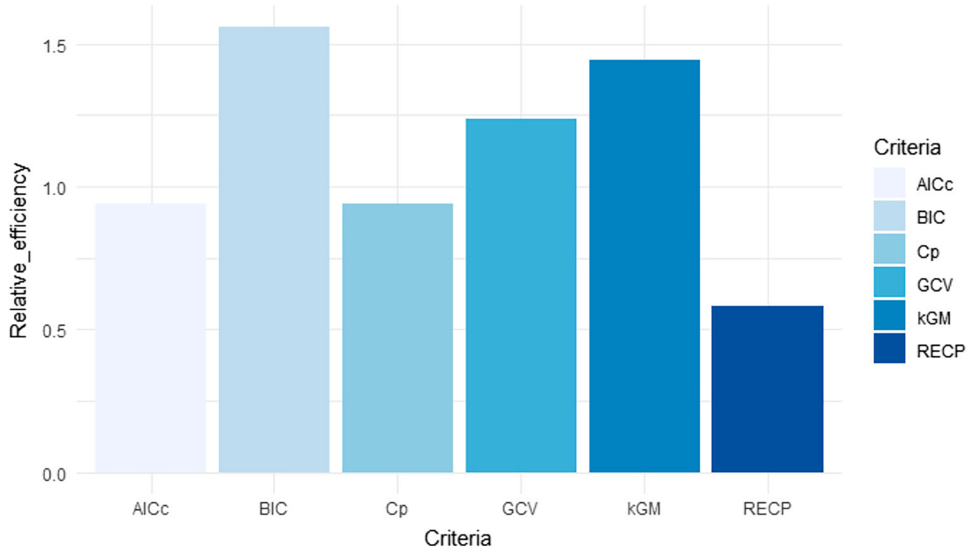

Figure 9. Bar plot for the relative efficiencies obtained from ridge estimators based on criteria.

with the others. As for $\hat{k}_{GM}$, it performs similarly to *BIC*. From these real data results, we can say that the results of the simulations and the real data study are in accordance with one another.

Figure 9 represents the relative efficiency values from the criteria for the GDP per capita dataset. It clearly shows that *RECP* is the most efficient method in selecting the ridge parameter. Interpretations given for Table 5 are likewise acceptable for Figure 9. Note finally that AIC_c, C_p, and *GCV* follow *RECP* in terms of efficiency, and *BIC* and $\hat{k}_{GM}$ do not perform well, as in the simulation experiments.

7. Conclusions and recommendations

In this paper, we introduced the Tobit ridge estimators to estimate the parameters of a Tobit model with collinear data. To efficiently calculate these estimators we needed an optimum shrinkage parameter. The optimum parameter was determined using information criteria, such as $AIC_c, BIC, GCV, RECP$, and Cp. The outcomes obtained from these criteria were compared to those found with $\hat{k}_{GM}$, which has been used as a benchmark method in this paper. Thus, six different Tobit ridge estimators based on ML (i.e. $\hat{\boldsymbol{\beta}}_{MLR}$ defined in (14)) were provided for the parameters of a Tobit regression model with left-censored and collinear data.

To compare the performance of the estimators, the values of SMDE, biases, and variances of regression coefficients, variances of the models and relative efficiencies were used as evaluation measurements. Outcomes from the simulation and with real data show that *RECP* performed better than the others, whereas *BIC* did not perform well. It should be noted that $\hat{k}_{GM}$ was used by Khalaf et al. [12] to select the shrinkage parameter for the Tobit ridge estimator, and it had given satisfying results in their study. However, in our study, the performance of $\hat{k}_{GM}$, similar to that of *BIC*, did not perform well.

The following conclusions are expressed to summarize the outcomes from the Monte Carlo simulation experiments and real data study:

- The Monte Carlo simulation results, performed under varying conditions, show that the quality of the parameter estimates is poorly affected by the correlation and censoring levels. Concerning this, $\hat{k}_{GM}$ and *BIC* have performed poorly in terms of providing a ridge estimator for the Tobit model with left-censored data, whereas *RECP*, *Cp*, and *GCV* have performed relatively better.
- From the boxplots given in Figures 3 and 4, it is shown that the quantities of the biases obtained from the estimators under the smaller samples are much larger than those obtained from the larger samples. This result implies that the censorship and correlation levels in the data are highly effected by the sample size. In this sense, *RECP* also provides the low-biased estimates.
- Although *RECP* generally had the best *SD* and *MDE* scores (see Tables 2 and 3), when simulation results are inspected in detail, we see that in some of the high censoring levels C_p and *GCV* have the same scores. Furthermore, the scores in Tables 2–4 denote that $\hat{k}_{GM}$ and *BIC* performed the worst, especially at high censoring levels.
- As shown in Figures 5 and 6, which show the bias-variance decomposition, it is more appropriate to use the Tobit ridge estimators based on shrinkage parameters selected by information criteria (i.e. *AICc*, *GCV*, *Cp*, *RECP*, and *BIC*) than classical methods, such as the $\hat{k}_{GM}$ method.
- According to the results from the GDP data, all methods perform satisfactorily. There is in fact little difference between them. Unsurprisingly, *RECP* had the minimum SMDE value in this real data example, as well as the simulations since it produces the estimators with minimum variances.
- In this study, as can be seen in explanations given above, although $\hat{k}_{GM}$ is a commonly used and successful estimator for the ridge parameter, it has an unsatisfying performance in terms of Tobit ridge estimator. If simulation study is inspected, it can be realized that performance of $\hat{k}_{GM}$ is getting worse for high censoring level which is the major cause of the unsatisfying results.

As a result of this study, we conclude that the *RECP* criterion is most appropriate in estimating the parameters of a Tobit regression model because the estimator with the ridge parameter selected by *RECP* has produced the estimates with the best numerical performance for all simulation configurations and the real dataset. Additionally, the *AICc*, *GCV*, and *Cp* criteria are also efficient for several simulation configurations. By contrast, $\hat{k}_{GM}$ and *BIC* perform the worst. Ultimately, *RECP* can be recommended as the best selection criterion.

Acknowledgements

We would like to thank the editor, the associate editor, and the anonymous referees for beneficial comments and suggestions.

Disclosure statement

No potential conflict of interest was reported by the author(s).

References

[1] Amemiya T. Advanced econometrics. Cambridge: Harvard University Press; 1985.
[2] Tobin J. Estimation of relationships for limited dependent variables. Econometrica. 1958;26:24–36.
[3] Hoerl AE, Kennard RW. Ridge regression: biased estimation for nonorthogonal problems. Technometrics. 1970;12:55–67.
[4] Scott J. Regression models of categorical and limited dependent variables. Thousand Oaks (CA): Sage; 1997.
[5] Roozbeh M. Shrinkage ridge estimators in semiparametric regression models. J Multivar Anal. 2015;136:56–74.
[6] Amini M, Roozbeh M. Optimal partial ridge estimation in restricted semiparametric regression models. J Multivar Anal. 2015;136:26–40.
[7] Roozbeh M. Optimal QR-based estimation in partially linear regression models with correlated errors using GCV criterion. Comput Stat Data Anal. 2018;117:45–61.
[8] Akdeniz F, Roozbeh M. Generalized difference-based weighted mixed almost unbiased ridge estimator in partially linear models. Stat Papers. 2019;60:1717–1739.
[9] Roozbeh M, Hesamian G, Akbari MG. Ridge estimation in semi-parametric regression models under stochastic restriction and correlated elliptically contoured errors. J Comput Appl Math. 2020;378. DOI:10.1016/j.cam.2020.112940
[10] Kibria BMG. Performance of some new ridge regression estimators. Commun Stat Simul Comput. 2003;32:419–435.
[11] Gajarado KAM. An extension of the normal censored regression model. Estimation and applications [PhD dissertation]. Santiago: Pontifica Universidad Catolica de Chile; 2009.
[12] Khalaf G, Mansson K, Sjolander P, et al. A Tobit regression estimator. Commun Stat Theory Methods. 2014;43(1):131–140.
[13] Mansson K, Shukur G. On ridge parameters in logistic regression. Commun Stat Theory Methods. 2011;40(18):3366–3381.
[14] Haq MS, Kibria BMG. A shrinkage estimator for there stricted linear regression model: ridge regression approach. J Appl Stat Sci. 1996;3(4):301–316.
[15] Fang Y. Asymptotic equivalence between cross-validations and Akaike information criteria in mixed effect models. J Data Sci. 2011;9(1):15–21.
[16] Aydın D, Yüzbaşı B, Ahmed SE. Modified ridge type estimator in partially linear regression models and numerical comparisons. J Comput Theor Nanosci. 2016;13(10):7040–7053.
[17] Yılmaz E, Yüzbaşı B, Aydın D. Choice of smoothing parameter for kernel type ridge estimators in semiparametric regression models. Revstat Stat J. 2018. REVSTAT-113-2017-R2.
[18] Amemiya T. Multivariate regression and simultaneous equation models when the dependent variables are truncated normal. Econometrica. 1974;42:999–1012.
[19] Schaefer RL, Roi D, Wolfe RA. A ridge logistic estimator. Commun StatTheory Methods. 1984;13(1):99–113.
[20] Le Cessie S, van Houwelingen JC. Ridge estimators in logistic regression. Appl Stat. 1992;41(1):191–201.
[21] Fair R. A note on computation of the Tobit estimator. Econometrica. 1977;45:1723–1727.
[22] Olsen RJ. Note on the uniqueness of the maximum likelihood estimator for the Tobit model. Econometrica. 1978;46(5):1211–1215.
[23] Van Wieringen WN. Lecture notes on ridge regression, arXiv preprint, arXiv:1509.09169; 2015.
[24] Greene WH. On the asymptotic bias of the ordinary least squares estimator of the Tobit model. Econometrica. 1981;49(2):505–513.
[25] Greene WH. The behavior of the maximum likelihood estimator of limited dependent variable models in the presence of fixed effects. Econ J. 2004;7(1):98–119.
[26] Sun Z, Guo Y, Xie T, et al. Model diagnostics of parametric Tobit model based on cumulative residuals. J Korean Stat Soc. 2020. DOI:10.1007/s42952-020-00069-2
[27] Goldberger AS. Linear regression after selection. J Econ. 1981;15(3):357–366.
[28] Hoerl AE, Kennard RW, Baldwin KF. Ridge regression: some simulations. Commun Stat. 1975;4:105–123.

[29] Golub GH, Heath M, Wahba G. Generalized cross-validation as a method for choosing a good ridge parameter. Technometrics. 1979;21(2):215–223.
[30] Pasha GR, Shah MA. Application of ridge regression to multicollinear data. J Res Sci. 2004;15(1):97–106.
[31] Hurvich C, Simonoff M, and Tasi JS, et al. Smoothing parameter selection in nonparametric regression using an improved Akaike information criterion. J R Statist Soc B. 1998;60:271–293.
[32] Schwarz G. Estimating the dimension of a model. Ann Stat. 1978;6:461–464.
[33] Craven P, Wahba G. Smoothing noisy data with spline functions. Num Math. 1979;31:377–403.
[34] Lee TCM. Smoothing parameter selection for smoothing splines: a simulation study. Comput Stat Data Anal. 2003;42:139–148.
[35] Mallows C. Some comments on C_p. Technometrics. 1973;15:661–675.
[36] Liang H. Estimation partially linear models and numerical comparison. Comput Stat Data Anal. 2006;50:675–687.
[37] Muniz G, Kibria BMG. On some ridge regression estimators: an empirical comparisons. Commun Stat Simul Comput. 2009;38(3):621–630.
[38] Belsley D, Kuh E, Welch R. Regression diagnostics: identifying influential data and sources of collinearity. New York (NY): Wiley; 1980.

Appendices

Appendix 1. Derivation of Equation (14)

Matrix and vector form of the Equation (13) can be obtained similar to Fair [21] but for ridge solution as follows to provide simplicity of the solution:

$$-\boldsymbol{X}_0^T\boldsymbol{\eta}_0 + \frac{1}{\sigma}\boldsymbol{X}_1^T(\boldsymbol{Y}_1 - \boldsymbol{X}_1\boldsymbol{\beta}) - k\boldsymbol{\beta} = 0 \tag{A1}$$

From that, derivation of the estimator of $\boldsymbol{\beta}$ is given by

$$\begin{aligned}
&\frac{1}{\sigma}\boldsymbol{X}_1^T\boldsymbol{Y}_1 - \boldsymbol{X}_1^T\boldsymbol{X}_1\boldsymbol{\beta} - k\boldsymbol{\beta} - \boldsymbol{X}_0^T\boldsymbol{\eta}_0 = 0\\
&\boldsymbol{X}_1^T\boldsymbol{X}_1\boldsymbol{\beta} + k\boldsymbol{\beta} = \frac{1}{\sigma}\boldsymbol{X}_1^T\boldsymbol{Y}_1 - \boldsymbol{X}_0^T\boldsymbol{\eta}_0\\
&(\boldsymbol{X}_1^T\boldsymbol{X}_1 + k\boldsymbol{I})\boldsymbol{\beta} = \frac{1}{\sigma}\boldsymbol{X}_1^T\boldsymbol{Y}_1 - \boldsymbol{X}_0^T\boldsymbol{\eta}_0\\
&\boldsymbol{\beta} = \frac{1}{\sigma}(\boldsymbol{X}_1^T\boldsymbol{X}_1 + k\boldsymbol{I})^{-1}\boldsymbol{X}_1^T\boldsymbol{Y}_1 - (\boldsymbol{X}_1^T\boldsymbol{X}_1 + k\boldsymbol{I})^{-1}\boldsymbol{X}_0^T\boldsymbol{\eta}_0
\end{aligned}$$

Thus, $\boldsymbol{\beta}$ solution in Equation (14) can be obtained for Tobit ridge model.

Appendix 2. Details of Equations (19–20)

Derivation of Equation (19) is obtained as follows

$$\begin{aligned}
E(\hat{\boldsymbol{\beta}}_{MLR}) &= E[(\boldsymbol{X}_1'\boldsymbol{X}_1 + k\boldsymbol{I})^{-1}[\boldsymbol{X}_1'\boldsymbol{Y}_1 - \sigma\boldsymbol{X}_0\boldsymbol{\eta}_0]]\\
&= E[(\boldsymbol{X}_1'\boldsymbol{X}_1 + k\boldsymbol{I})^{-1}[\boldsymbol{X}_1'\boldsymbol{X}_1\boldsymbol{\beta} - \sigma\boldsymbol{X}_0\boldsymbol{\eta}_0]]\\
&= E[(\boldsymbol{X}_1'\boldsymbol{X}_1 + k\boldsymbol{I})^{-1}\boldsymbol{X}_1'\boldsymbol{X}_1\boldsymbol{\beta} - (\boldsymbol{X}_1'\boldsymbol{X}_1 + k\boldsymbol{I})^{-1}\sigma\boldsymbol{X}_0\boldsymbol{\eta}_0]\\
&= (\boldsymbol{X}_1'\boldsymbol{X}_1 + k\boldsymbol{I})^{-1}(\boldsymbol{X}_1'\boldsymbol{X}_1 + k\boldsymbol{I} - k\boldsymbol{I})\beta - (\boldsymbol{X}_1'\boldsymbol{X}_1 + k\boldsymbol{I})^{-1}\sigma\boldsymbol{X}_0\boldsymbol{\eta}_0\\
&= (\boldsymbol{X}_1'\boldsymbol{X}_1 + k\boldsymbol{I})^{-1}[(\boldsymbol{X}_1'\boldsymbol{X}_1 + k\boldsymbol{I})\boldsymbol{\beta} - k\boldsymbol{I}\boldsymbol{\beta}] - (\boldsymbol{X}_1'\boldsymbol{X}_1 + k\boldsymbol{I})^{-1}\sigma\boldsymbol{X}_0\boldsymbol{\eta}_0\\
&= [\boldsymbol{I} - k(\boldsymbol{X}_1'\boldsymbol{X}_1 + k\boldsymbol{I})^{-1}]\boldsymbol{\beta} - (\boldsymbol{X}_1'\boldsymbol{X}_1 + k\boldsymbol{I})^{-1}\sigma\boldsymbol{X}_0\boldsymbol{\eta}_0\\
&= \boldsymbol{\beta} - k(\boldsymbol{X}_1'\boldsymbol{X}_1 + k\boldsymbol{I})^{-1}\boldsymbol{\beta} - (\boldsymbol{X}_1'\boldsymbol{X}_1 + k\boldsymbol{I})^{-1}\sigma\boldsymbol{X}_0\boldsymbol{\eta}_0
\end{aligned}$$

If $\boldsymbol{A}_k = (\boldsymbol{X}_1{}'\boldsymbol{X}_1 + k\boldsymbol{I})^{-1}$ then

$$E(\hat{\boldsymbol{\beta}}_{MLR}) = \boldsymbol{\beta} - k\boldsymbol{A}_k\boldsymbol{\beta} - \boldsymbol{A}_k\sigma\boldsymbol{X}_0\boldsymbol{\eta}_0$$

and from that bias of the $\boldsymbol{\beta}$ could be written as

$$BIAS(\hat{\boldsymbol{\beta}}_{MLR}) = k\boldsymbol{A}_k\boldsymbol{\beta} + \boldsymbol{A}_k\sigma\boldsymbol{X}_0\boldsymbol{\eta}_0$$

For Equation (20) it should be written first,

$$E\left(\frac{\partial^2 logL^{pen}}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}'}\right) = -\sum_0 \frac{\phi(v_i)}{(1-\Phi(v_i))}\left[\phi(v_i) - \frac{1}{\sigma^2}(1-\Phi(v_i)\boldsymbol{x}_i^T\boldsymbol{\beta}\right]\boldsymbol{x}_i\boldsymbol{x}_i^T - \frac{1}{\sigma^2}\sum_1 \boldsymbol{x}_i\boldsymbol{x}_i^T - k$$

In order to make this expression simple, using with the probability limits of this second derivative, scoring method is applied. Thus, second derivative of the penalized log likelihood function according to $\boldsymbol{\beta}$ can be written more easily

$$E\left(-\frac{\partial^2 logL^{pen}}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}'}\right) = (\boldsymbol{X}_1'\boldsymbol{R}\boldsymbol{X}_1 + k\boldsymbol{I}_p)^{-1}$$

Thus it can be said that Equation (20) is ensured (see [11,21] for details).