STATISTICS, OPTIMIZATION AND INFORMATION COMPUTING
Stat., Optim. Inf. Comput., Vol. **x**, Month 202x, pp 0–12.
Published online in International Academic Press (www.IAPress.org)

An efficient shrinkage estimator for Cox proportional hazards models

Abeer Adil Hamed Mohammed [1], Ammar Ahmed Ali [2], Zakariya Yahya Algamal [3,*]

[1]*Department of Mathematics, College of Education for Pure Sciences, Al-Hamdaniya University, Iraq*
[2]*Department of Educational and Psychological Sciences, College of Education for Humanities, University of Mosul, Iraq*
[3]*Department of Statistics and Informatics, University of Mosul, Mosul, Iraq*

Abstract Multicollinearity between prognostic covariates between medical survival studies is a frequent occurrence that may cause severe instability of Cox proportional hazards regression model resulting in inflated variances, wide confidence intervals, and imprecise estimates of the hazard ratio. To counter this issue, we suggest a variation of the KibriaLukman estimator (CPHRM-LKLE) of the Cox model which is a combination of the shrinkage properties of the ridge and KibriaLukman estimators by an optimally chosen biasing parameter based on minimizing the mean squared error. We obtain the covariance structure and the mean squared error of the proposed estimator and then provide the theoretical conditions when CPHRM-LKLE outperforms the conventional maximum likelihood estimator and various other existing shrinkage estimators, such as ridge, Liu, Liu-type, and Kibria-Lukman estimators. Monte Carlo simulation study is carried out in a wide range of designs, sample size, number of covariates, level of pairwise correlation and censoring proportion. In all cases, CPHRM-LKLE would always achieve the lowest average mean squared error, the performance improvement is more vivid in cases of extreme multicollinearity, increased censoring, and more complicated models. The empirical usefulness of the method is further demonstrated by two empirical survival data of reported multicollinearity, the National Wilm tumor study and the NCCTG lung cancer cohort. CPHRM-LKLE provides the best overall mean squared error with the direction to covariates and clinical interpretability of covariate effects, and provides more stable relapse and mortality risk estimates than the other methods. The findings indicate that CPHRM-LKLE is a practical and strong alternative of Cox regression is there is the presence of correlated prognostic factors.

Keywords NCCTG lung cancer, Cox regression model, National Wilm's tumor studym, multicollinearity, biasing parameter, Kibria–Lukman estimator

DOI: 10.19139/soic-2310-5070-3725

1. Introduction

Survival analysis is the branch of statistics devoted to studying the time until an event occurs, explicitly accommodating incomplete observation of that time and often allowing for time-varying risk. It underpins much of modern clinical, epidemiologic, and reliability research, where understanding not only whether but also when events occur is essential [1, 2]. In survival analysis, the outcome is a non-negative random variable representing time from a well-defined origin, such as diagnosis, treatment initiation, hospital admission, to a specified event, such as death, relapse, device failure, hospital discharge. Unlike standard continuous outcomes, many observations are censored, for some individuals.

Regression modeling in survival analysis provides a framework for relating time-to-event outcomes to explanatory variables while correctly accounting for censoring and the peculiar distributional features of survival times. From a modeling perspective, survival regression links a function of the event process, typically the hazard rate or the survival time, to a linear predictor constructed from covariates [3, 45, 46, 47, 48, 49].

*Correspondence to: Zakariya Yahya Algamal (Email: zakariya.algamal@uomosul.edu.iq). Department of Statistics and Informatics, University of Mosul, Mosul, Iraq.

ISSN 2310-5070 (online) ISSN 2311-004X (print)
Copyright © 202x International Academic Press

The Cox proportional hazards regression model (CPHRM) is the dominant regression framework for analyzing time-to-event data when censoring is present and the goal is to assess how covariates affect the hazard of an event over time. It combines flexible, nonparametric modeling of the underlying event rate with a parametric structure for covariate effects, making it both practical and powerful for biomedical and epidemiologic research [4, 5, 6, 7, 8, 50, 51].

Multicollinearity is a key concern in the CPHRM because it arises when two or more covariates are highly correlated, creating near-linear dependencies in the design matrix and undermining the stability of coefficient estimates. In this setting, the maximum likelihood estimator (MLE) becomes flat in directions corresponding to these dependencies, leading to inflated standard errors, wide confidence intervals, and sensitivity of the estimated log-hazard ratios to small changes in the dataset or model specification [52, 53, 54]. As a result, individual coefficients for correlated predictors may appear statistically non-significant, may have counterintuitive signs, or may take on implausibly large magnitudes, even though the joint predictive contribution of the group of variables is substantial [9, 10, 11, 55]. The issue of multicollinearity in CPHM has been examined by numerous scholars, who have also suggested novel biased estimating methods as alternatives to the MLE [12, 13, 14, 15, 16, 17]. Although these estimators improve the performance of CPHM in the presence of multicollinearity, there remains scope to construct estimators that achieve further reductions in mean squared error by exploiting additional structure in the biasing mechanism [56, 57, 58, 59].

Motivated by Aladeitan, Adebimpe [18], the main contribution of this paper is the development of a new biased estimator for the CPHRM that is specifically tailored to handle multicollinearity more effectively than existing methods. Building on the Kibria–Lukman estimator and ridge estimation, the study proposes a modified Kibria–Lukman estimator that combines the strengths of both approaches within the CPHRM framework. By exploiting additional structure in the biasing mechanism, the proposed estimator aims to further reduce the inflated mean squared error of the maximum likelihood estimator under multicollinearity, thereby improving estimation accuracy and stability relative to currently available biased estimators for the CPHRM.

2. CPHRM

The statistical area of survival analysis focuses on time-to-event data, or more specifically, the amount of time passing between a clearly defined beginning event and a specific endpoint. One of the most well-liked and practical models in survival analysis is the CPHRM [19].

Consider the observed time be representing as $\{(t_i, \tau_i, c_i) : i = 1, ..., n\}$ where t_i is the survival time if $\tau_i = 1$, censored time if $\tau_i = 0$ and $c_i = (c_{i1}, c_{i2}, ..., c_{ip})$ is a p-dimensional explanatory variables. Then, the CPHRM can be defined as

$$h\left(t_i|z_i\right) = h_0\left(t_i\right) \exp\left(\beta^T c_i\right), \tag{1}$$

where $h_0\left(t_i\right)$ is the baseline function and $\beta = (\beta_1, ..., \beta_p)^T$ is an unknown regression coefficient of $p \times 1$ vector. The following partial likelihood represents the combined probability of all realized events, assuming that the subjects are statistically independent of one another

$$L\left(\beta\right) = \prod_{i=1}^{n} \frac{\exp\left(\beta^T c_i\right)}{\sum_{j \in R_i} \exp\left(\beta^T c_j\right)}, \tag{2}$$

where R_i is the set of subjects that are at risk just before time t_i.

Typically, the partial log likelihood function (Eq. (2)) is minimized in order to estimate the regression parameters of Eq. (1) by MLE method

$$\hat{\beta}_{\mathrm{CPHRM-MLE}} = \arg\min_{\beta} \left(\log L(\beta)\right) = -\sum_{i=1}^{n} \left[\beta^T c_i - \log \sum_{j \in R_i} \exp(\beta^T c_j)\right]. \tag{3}$$

Then,

$$\hat{\beta}_{\mathrm{CPHRM-MLE}} = (\hat{C}^T \hat{C})^{-1} \hat{C}^T \hat{L}, \tag{4}$$

where $\hat{C}$ is a $\sum(q_i+1)\times p$ ($q_i+1$is the size of ith risk set) matrix and $\hat{L}$ is a $p\times 1$ vector. The covariance (Cov) matrix of $\hat{\beta}_{\mathrm{CPHRM-MLE}}$is defined as

$$Cov(\hat{\beta}_{\mathrm{CPHRM-MLE}}) = \left[-\frac{\partial^2 L(\beta)}{\partial\beta\,\partial\beta'}\bigg|_{\beta=\hat{\beta}_{\mathrm{CPHM-MLE}}}\right]^{-1} = (\hat{C}^T\hat{C})^{-1} \tag{5}$$

The mean squared error (MSE) of Eq. (4) can be obtained as

$$\mathrm{MSE}(\hat{\beta}_{\mathrm{CPHRM-MLE}}) = \sum_{j=1}^{p}\frac{1}{\delta_j}, \tag{6}$$

where δ_j is the eigenvalue of the $\hat{C}^T\hat{C}$ matrix.

3. Collinearity issue in PCHRM

In the presence of multicollinearity, the matrix $\hat{C}^T\hat{C}$ becomes ill-conditioned leading to high variance and instability of the MLE of the CPHRM parameters. Therefore, in the context of multicollinearity issue, the goal is to improve MLE performance. The accuracy of parameter estimate can be increased by using shrinkage estimators [20, 60, 61].

To solve this issue in CPHRM, the ridge and Liu-type estimators of Hoerl and Kennard [21] and Liu [22] were expanded [13, 14]. Ahmad, Aslam and Ahmad [12], in terms of Liu [23], proposed the Liu estimator for the CPHRM. Further, Ahmad, Aslam [24] proposed Aslam-Ahmad estimator for the CPHRM. Additionally, Algamal [15] proposed a linearized ridge estimator in CPHRM. Furthermore, Seifollahi, Algamal and Arashi [16] extended the Kibria-Lukman estimator in the CPHRM. Moreover, Seifollahi and Arashi [25] improved the Kibria-Lukman estimator by developing linear shrinkage, Stein and its positive counterpart, pretest, and shrinkage pretest estimators in the CPHRM.

The ridge estimator in CPHRM, CPHRM-RE can be defined as:

$$\hat{\beta}_{\mathrm{CPHRM-RE}} = (\hat{C}^T\hat{C}+kI_p)^{-1}\,\hat{C}^T\hat{C}\,\hat{\beta}_{\mathrm{CPHRM-MLE}} \tag{7}$$

where $k>0$is a biasing factor. The covariance and MSE of $\hat{\beta}_{\mathrm{CPHRM-RE}}$can be written as

$$Cov(\hat{\beta}_{\mathrm{CPHRM-RE}}) = (\hat{C}^T\hat{C}+kI)^{-1}\,\hat{C}^T\hat{C}(\hat{C}^T\hat{C}+kI)^{-1} \tag{8}$$

$$\mathrm{MSE}(\hat{\beta}_{\mathrm{CPHRM-RE}}) = \sum_{j=1}^{p}\frac{\delta_i}{(\delta_j+k)^2} + k^2\sum_{j=1}^{p}\frac{\vartheta_j^2}{(\delta_j+k)^2} \tag{9}$$

where ϑ_j is defined as the jth element of $\gamma\,\hat{\beta}_{\mathrm{CPHRM-MLE}}$and γ is the eigenvector of the$\hat{C}^T\hat{C}$.

On the other hand, the Liu estimator in the CPHRM, CPHRM-LE, can be defined as

$$\hat{\beta}_{\mathrm{CPHRM-LE}} = (\hat{C}^T\hat{C}+I_p)^{-1}\,(\hat{C}^T\hat{C}+dI_p)\hat{\beta}_{\mathrm{CPHRM-MLE}}, \tag{10}$$

where $0<d<1$ is a biasing factor. The covariance and MSE of $\hat{\beta}_{\mathrm{CPHRM-LE}}$can be written as

$$Cov(\hat{\beta}_{\mathrm{CPHRM-LE}}) = (\hat{C}^T\hat{C}+I_p)^{-1}\,(\hat{C}^T\hat{C}+dI_p)\,\hat{C}^T\hat{C}\,(\hat{C}^T\hat{C}+dI_p)(\hat{C}^T\hat{C}+I_p)^{-1}, \tag{11}$$

$$\mathrm{MSE}(\hat{\beta}_{\mathrm{CPHRM-LE}}) = \sum_{j=1}^{p}\frac{(\delta_j+d)^2}{\delta_j(\delta_j+1)^2} + (d-1)^2\sum_{j=1}^{p}\frac{\vartheta_j^2}{(\delta_j+1)^2} \tag{12}$$

A Kibria-Lukman estimator (KLE) is a recent ridge-type biased estimator which was proposed to stabilize the inference in linear and generalized linear regression models when multicollinearity causes a breakdown in the

efficiency of more traditional estimators like ordinary least squares (OLS) and MLE. Kibria and Lukman [26] suggested the KLE as a one-parameter ridge-type estimator that is derived by mixing the structures of the classical ridge and Liu estimators in the standard linear regression model. Statistical properties of the KLE have been examined in the context of the MSE which is a combination of both bias and variance. A number of papers apply KLE to GLMs, especially Poisson or logistic regression, in which multicollinearity worsens the performance of MLE [27, 28, 29, 30, 31, 32, 44].

Seifollahi, Algamal and Arashi [16] extended the KLE to the CPHRM. This estimator, CPHRM-KLE can be defined as:

$$\hat{\beta}_{\text{CPHRM}-\text{KLE}} = (\hat{C}^T\hat{C} + kI_p)^{-1}(\hat{C}^T\hat{C} - kI_p)\hat{\beta}_{\text{CPHRM}-\text{MLE}} \tag{13}$$

where $k > 0$. The covariance and MSE of $\hat{\beta}_{\text{CPHRM}-\text{KLE}}$can be written as

$$Cov(\hat{\beta}_{\text{CPHRM}-\text{KLE}}) = (\hat{C}^T\hat{C} + I_p)^{-1}(\hat{C}^T\hat{C} - kI_p)\,\hat{C}^T\hat{C}(\hat{C}^T\hat{C} - kI_p)(\hat{C}^T\hat{C} + I_p)^{-1}, \tag{14}$$

$$\text{MSE}(\hat{\beta}_{\text{CPHRM}-\text{KLE}}) = \sum_{j=1}^{p}\frac{(\delta_i - k)^2}{\delta_i(\delta_i + k)^2} + 4k^2\sum_{j=1}^{p}\frac{\vartheta_j^2}{(\delta_j + k)^2}. \tag{15}$$

4. The proposed estimator

Several modifications of the KLE have been proposed in the literature [29, 33, 34, 35, 36, 37, 38, 39, 40]. Aladeitan, Adebimpe [18] suggested the modified KLE, which was particularly used to estimate the Poisson regression model under collinearity. The estimator has inherited the privilege of the Kibria-Lukman with the ridge estimator. Although most of the estimators mentioned above are capable of doing so be either single-parameter or two-parameter estimators.

Following Oladapo, Idowu [41], we proposed an extension of the modified KLE in the CPHRM, named as CPHRM-LKLE, by replacing the$\hat{\beta}_{\text{CPHRM}-\text{MLE}}$ in Eq. (13) with $\hat{\beta}_{\text{CPHRM}-\text{NLE}} = (C^TC + I_p)^{-1}(C^TC - dI_p)\hat{\beta}_{\text{CPHRM}-\text{MLE}}$.

The CPHRM-LKLE will provide better stability and accuracy of the estimator when the traditional estimators can be ineffective because there is a high level of collinearity between predictor variables. In CPHRM-LKLE

The CPHRM-LKLE is defined as

$$\begin{aligned}\hat{\beta}_{\text{CPHRM}-\text{LKLE}} &= (C^TC + kI_p)^{-1}(C^TC - kI_p)\hat{\beta}_{\text{CPHRM}-\text{NLE}} \\ &= (C^TC + kI_p)^{-1}(C^TC - kI_p)(C^TC + I_p)^{-1}(C^TC - dI_p)\hat{\beta}_{\text{CPHRM}-\text{MLE}},\end{aligned} \tag{16}$$

where $k > 0$ and $0 < d < 1$. The MSE of Eq. (16) can be written as

$$\text{MSE}(\hat{\beta}_{\text{CPHRM}-\text{MKLE}}) = \sum_{j=1}^{p}\frac{\delta_i(\delta_i - k)^2(\delta_i - d)^2}{\delta_i(\delta_i + k)^2(\delta_i + 1)^2} + \sum_{j=1}^{p}\frac{(2k\delta_i + d\delta_i - kd + \delta_i + k)^2\vartheta_j^2}{\delta_i(\delta_i + k)^2(\delta_i + 1)^2}. \tag{17}$$

5. Monte Carlo Simulation study

The effectiveness of our suggested estimator, CPHRM-LKLE, is examined in this section. In accordance with McDonald and Galarneau [42], predictors with varying levels of multicollinearity are produced by

$$z_{ij} = (1 - \rho^2)^{1l2}b_{ij} + \rho b_{ip}, \quad i = 1, 2, ..., n, \quad j = 1, 2, ..., p, \tag{18}$$

where b_{ij}'s are produced from a standard normal distribution and ρ^2 denotes the correlation between the explanatory variables. Since multicollinearity can potentially affect the estimates of small samples, the small sample sizes in terms of the number of explanatory factors were used. Here, $n \in \{30, 75, 150, 500\}$. Two values of the number of explanatory variables were set to be $p = 4$ and $p = 8$. Choosing several explanatory variables

is essential for capturing the complexity of relationships between predictors and the time-to-event outcome. The pairwise correlation are considered with $\rho \in \{0.85, 0.90, 0.95, 0.99\}$. A Bernoulli distribution with censoring (cen) probability produces the censoring time by $cen = 0.2$ and $cen = 0.4$. The selection of multiple values of censoring in a CPHRM is significant due to a number of reasons, most of which are associated with dealing with the complexities of survival data and enhancing the effectiveness of the model estimates [62, 63, 64]. The baseline survival time was assumed to follow an exponential distribution with parameter equal to 1.5. The parameters, β, were set to be (0.50, 0.32, -0.61, -0.28) for $p = 4$ and (0.39, 0.30, -0.55, -0.30, -0.36, 0.45, 0.65, -0.80) for $p = 8$. The generated data is repeated 5000 times for a combination of all these factors, and the averaged mean squared errors (AMSE) are computed as

$$\text{AMSE}(\hat{\beta}) = \frac{1}{5000}\sum_{i=1}^{5000}(\hat{\beta} - \beta)^T(\hat{\beta} - \beta), \tag{19}$$

where $\hat{\beta}$ is the obtained shrinkage estimator. The CPHRM-LKLE estimator with $k = 1/\max\left(\varphi_j^2\right)$ was compared with: Eq. (1) CPHRM-MLE in Eq. (4), Eq. (2) CPHRM-RE in Eq. (7) with $k = 1/\max\left(\varphi_j^2\right)$, Eq. (3) CPHRM-LE in Eq. (10) with $d$as in İnan and Tez [13], and Eq. (4) CPHRM-KLE estimator with $k = 1/\max\left(\varphi_j^2\right)$. Tables 1–4 summarize the AMSE over 5000 replications for all competing estimators of the CPHRM under varying degrees of multicollinearity, censoring, sample size, and model dimension. In each configuration we compare the CPHRM-MLE, the CPHRM-RE, the CPHRM-LE, the CPHRM-LTE, the CPHRM-KLE, and the CPHRM-LKLE , where the smallest AMSE in each scenario is highlighted in bold.

In Table 1 we consider the low-dimensional setting with four covariates and moderate censoring, and pairwise correlations. Across all combinations of n and of ρ, the proposed CPHRM-LKLE achieves the lowest AMSE, outperforming CPHRM-MLE, CPHRM-RE, CPHRM-LE, CPHRM-LTE, CPHRM-KLE. As expected, the AMSE increases as ρ approaches 0.99, indicating the detrimental effect of severe multicollinearity on all estimators; however, the increase for CPHRM-LKLE is consistently smaller in magnitude. For example, at $n = 30$ and $\rho = 0.99$, the AMSE decreases from 7.047 for CPHRM-MLE to 3.529 for CPHRM-LKLE , representing a substantial reduction in overall estimation error. Increasing the sample size leads to lower AMSE for all methods, yet CPHRM-LKLE remains uniformly superior, indicating that its advantage is not restricted to small samples.

Table 2 reports results for the higher-dimensional scenario with eight covariates under the same grid of sample sizes, correlations, and a moderate censored fraction. The AMSE values are generally larger than those in Table 1, reflecting the increased difficulty of estimating more parameters in the presence of strong multicollinearity. Despite this, the proposed estimator, CPHRM-LKLE , continues to dominate its competitors in every configuration, again yielding the smallest AMSE. The benefit of CPHRM-LKLE is most pronounced under high collinearity and small samples; for instance, when $n = 30$ and $\rho = 0.99$, CPHRM-MLE has AMSE 7.209, while CPHRM-LKLE reduces it to 3.691, nearly halving the average error. This pattern suggests that the added biasing structure in CPHRM-LKLE is particularly effective in stabilizing coefficient estimation when the design matrix is highly ill-conditioned.

In Table 3 we return to the four-covariate model but increase the censoring level, thereby reducing the effective information available for estimation. As anticipated, the AMSE values are slightly larger than those observed in Table 1 for comparable $n$and ρ, confirming the adverse impact of censoring on precision. Nevertheless, CPHRM-LKLE continues to provide the smallest AMSE in all design points, again outperforming both traditional and existing shrinkage estimators. For example, at $n = 150$ and $\rho = 0.95$, the AMSEs for CPHRM-MLE, CPHRM-RE, CPHRM-LE, CPHRM-LTE, and CPHRM-KLE are 6.925, 4.161, 3.523, 3.379, and 3.355, respectively, whereas CPHRM-LKLE attains an AMSE of 3.317. These findings indicate that the proposed estimator maintains its efficiency gains even when information is lost due to censoring, a setting that closely mirrors practical medical survival studies.

Table 4 presents the most challenging scenario, combining a higher-dimensional model with eight covariates, strong multicollinearity, and a higher censoring rate. This configuration yields the largest AMSE values across all estimators, particularly for CPHRM-MLE, highlighting the severe instability induced by multicollinearity in conjunction with heavy censoring. Even in this demanding context, CPHRM-LKLE consistently achieves the lowest AMSE for all sample sizes and correlation levels. For example, at $n = 500$and $\rho = 0.99$, the AMSE reduces

from 7.104 for CPHRM-MLE to 3.586 for CPHRM-LKLE , while all other shrinkage estimators lie between these two extremes. The persistence of this dominance across increasing p, ρ, and censoring supports the theoretical MSE comparisons developed earlier and confirms that the modified Kibria–Lukman biasing mechanism translates into substantial finite-sample efficiency gains in CPHRM under multicollinearity.

Related to the performance comparison between CPHRM-LKLE and CPHRM-KLE, across all simulation scenarios in Tables 1-4, the proposed CPHRM-LKLE systematically improves on the original CPHRM-KLE in terms of AMSE, regardless of sample size, correlation level, censoring rate, or model dimension. In the four-covariate settings, Tables 1 and 3, CPHRM-LKLE always yields the smallest AMSE, with CPHRM-KLE being the second best, indicating that the modification effectively refines an already competitive estimator. The gain becomes more pronounced as multicollinearity intensifies, for instance, at $n = 30$and $\rho = 0.99$ in Table 1, CPHRM-KLE achieves an AMSE of 3.578, whereas CPHRM-LKLE further reduces it to 3.529. Similar improvements are observed under higher censoring in Table 3; for example, at $n = 150$ and $\rho = 0.95$, CPHRM-KLE and CPHRM-LKLE yield AMSEs of 3.355 and 3.317, respectively. Further, in the eight-covariate scenarios, Tables 2 and 4, where censoring and dimensionality exert greater impact, the relative advantage of CPHRM-LKLE persists. For instance, in Table 2 with $n = 30$ and $\rho = 0.99$, CPHRM-KLE reports an AMSE of 3.742, while CPHRM-LKLE reduces this to 3.691. These consistent, though numerically modest, reductions across all designs show that embedding the ridge-type component into the Kibria–Lukman structure yields a uniformly more efficient shrinkage estimator for CPHRM in the presence of multicollinearity.

In conclusion, across all four simulation tables, both the original, CPHRM-KLE, and the modified version, CPHRM-LKLE , clearly outperform the classical CPHRM-MLE when the predictors are highly correlated, but the modified version is consistently the most accurate. In practical terms, this means that CPHRM-LKLE provides more stable estimates of covariate effects, which should translate into more reliable estimates of relative risks in clinical survival studies.

Table 1. The AMSE values when $p = 4$ and $cen = 0.2$

n	ρ	CPHRM-MLE	CPHRM-RE	CPHRM-LE	CPHRM-KLE	CPHRM-MKLE
30	0.85	6.5	4.089	3.452	3.283	3.242
	0.90	6.756	4.126	3.485	3.323	3.287
	0.95	6.993	4.229	3.591	3.423	3.385
	0.99	7.115	4.444	3.803	3.646	3.597
75	0.85	6.413	4.002	3.365	3.196	3.155
	0.90	6.669	4.039	3.402	3.233	3.202
	0.95	6.906	4.142	3.504	3.336	3.303
	0.99	7.032	4.357	3.716	3.559	3.511
150	0.85	6.392	3.981	3.344	3.175	3.134
	0.90	6.653	4.022	3.377	3.212	3.179
	0.95	6.885	4.121	3.483	3.315	3.277
	0.99	7.007	4.336	3.695	3.541	3.489
500	0.85	6.284	3.873	3.236	3.067	3.026
	0.90	6.544	3.912	3.269	3.104	3.071
	0.95	6.777	4.013	3.375	3.207	3.169
	0.99	6.899	4.232	3.587	3.432	3.381

6. Real survival data Application

In order to assess the quality of the suggested estimator in practice, two real data sets are used. The first data set from the National Wilm's Tumor Study, which is described in Breslow and Chatterjee [43] and used by Ahmad, Aslam and Ahmad [12], Algamal [15], and Seifollahi, Algamal and Arashi [31]. This dataset has 4028 observations on the following variables are included in the dataset. Time to relapse (time variable), Eq. (1) histology from local institution (z1), Eq. (2) Patients whose tumors contain one of the uncommon cell types known as "unfavorable histology" are more likely to relapse and die than are patients with tumors of "favorable histology" (z2). Eq. (3)

Table 2. The AMSE values when $p = 8$ and $cen = 0.2$

n	ρ	CPHRM-MLE	CPHRM-RE	CPHRM-LE	CPHRM-KLE	CPHRM-MKLE
30	0.85	6.662	4.251	3.614	3.445	3.404
	0.90	6.919	4.291	3.647	3.482	3.449
	0.95	7.155	4.391	3.753	3.585	3.547
	0.99	7.277	4.606	3.965	3.81	3.759
75	0.85	6.575	4.164	3.527	3.358	3.317
	0.90	6.831	4.201	3.564	3.395	3.362
	0.95	7.072	4.304	3.666	3.499	3.465
	0.99	7.192	4.519	3.88	3.721	3.672
150	0.85	6.554	4.143	3.506	3.337	3.296
	0.90	6.812	4.184	3.539	3.374	3.341
	0.95	7.047	4.283	3.645	3.477	3.439
	0.99	7.169	4.5	3.857	3.703	3.651
500	0.85	6.446	4.035	3.4	3.229	3.19
	0.90	6.702	4.072	3.431	3.266	3.233
	0.95	6.939	4.175	3.537	3.369	3.331
	0.99	7.061	4.394	3.749	3.592	3.543

Table 3. The AMSE values when$p = 4$ and $cen = 0.4$

n	ρ	CPHRM-MLE	CPHRM-RE	CPHRM-LE	CPHRM-KLE	CPHRM-MKLE
30	0.85	6.61	4.197	3.56	3.391	3.35
	0.90	6.864	4.234	3.593	3.43	3.395
	0.95	7.101	4.337	3.699	3.531	3.493
	0.99	7.223	4.552	3.911	3.754	3.705
75	0.85	6.521	4.11	3.473	3.304	3.263
	0.90	6.777	4.147	3.51	3.341	3.312
	0.95	7.014	4.25	3.612	3.444	3.411
	0.99	7.139	4.465	3.824	3.667	3.621
150	0.85	6.5	4.089	3.452	3.283	3.242
	0.90	6.759	4.13	3.485	3.32	3.287
	0.95	6.993	4.229	3.591	3.423	3.385
	0.99	7.115	4.444	3.803	3.649	3.597
500	0.85	6.392	3.981	3.344	3.175	3.134
	0.90	6.65	4.02	3.377	3.212	3.179
	0.95	6.885	4.121	3.483	3.315	3.277
	0.99	7.007	4.34	3.695	3.541	3.489

Table 4. The AMSE values when $p = 8$ and $cen = 0.4$

n	ρ	CPHRM-MLE	CPHRM-RE	CPHRM-LE	CPHRM-KLE	CPHRM-MKLE
30	0.85	6.773	4.362	3.725	3.556	3.515
	0.90	7.03	4.402	3.76	3.593	3.56
	0.95	7.266	4.502	3.864	3.696	3.661
	0.99	7.39	4.717	4.076	3.921	3.87
75	0.85	6.686	4.275	3.641	3.469	3.431
	0.90	6.942	4.312	3.675	3.506	3.473
	0.95	7.183	4.415	3.777	3.61	3.576
	0.99	7.303	4.63	3.991	3.832	3.783
150	0.85	6.665	4.254	3.617	3.45	3.407
	0.90	6.923	4.295	3.65	3.485	3.452
	0.95	7.159	4.394	3.756	3.591	3.55
	0.99	7.28	4.611	3.97	3.814	3.762
500	0.85	6.557	4.146	3.511	3.34	3.301
	0.90	6.813	4.183	3.542	3.377	3.344
	0.95	7.05	4.286	3.651	3.48	3.442
	0.99	7.172	4.505	3.86	3.703	3.654

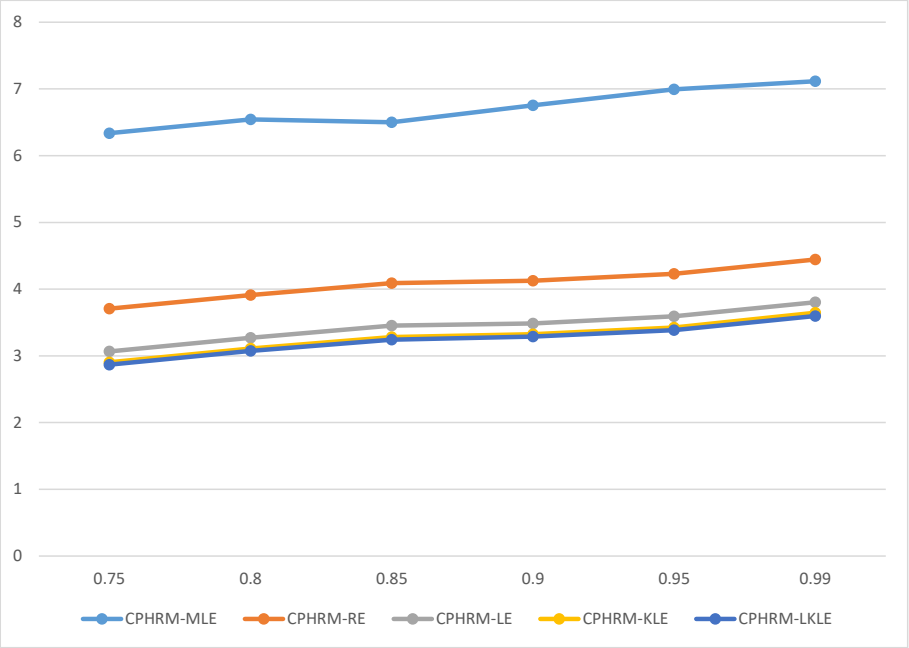


Figure 1. The performance of used methods when $p = 4$, $n = 30$ $cen = 0.2$

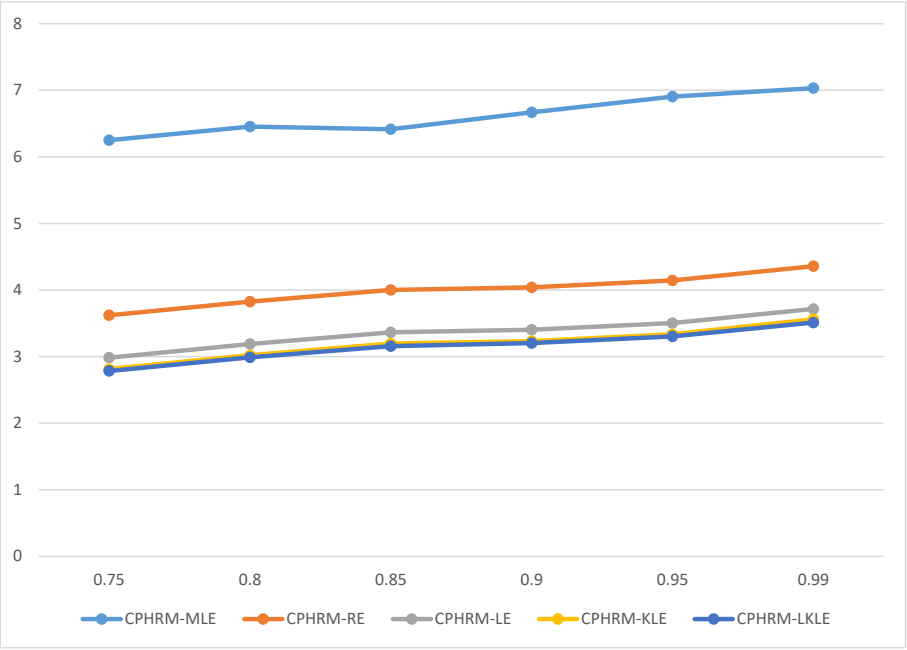


Figure 2. The performance of used methods when $p = 4$, $n = 75$ $cen = 0.2$

Disease stage (z3); divided into four categories, Eq. (4) study (z4); classified into two categories, and Eq. (5) age in months (z5). The correlation between the histology of the local institution and the central lab is 0.74, which is very

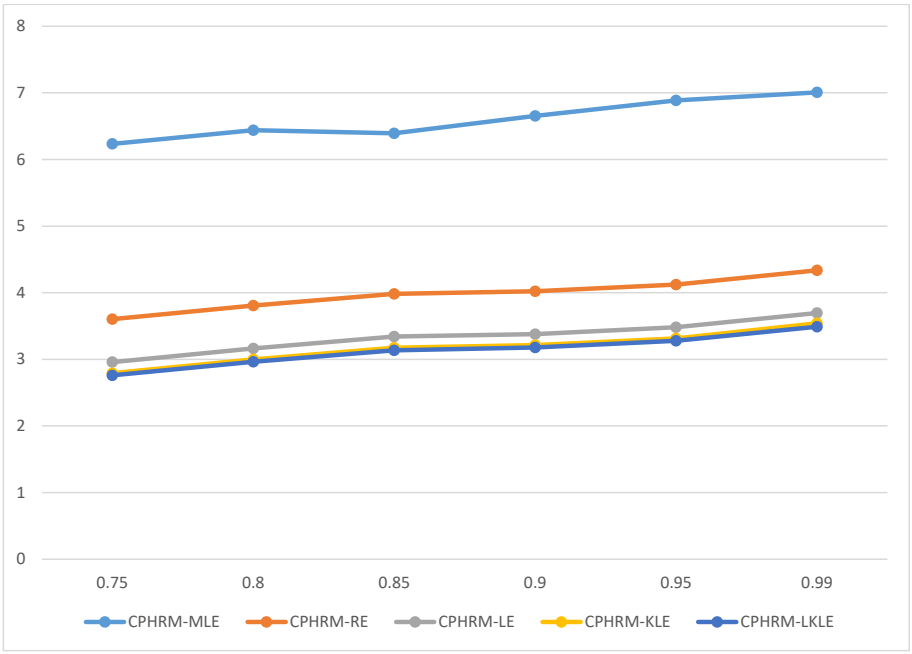


Figure 3. The performance of used methods when $p = 4$, $n = 150$ $cen = 0.2$

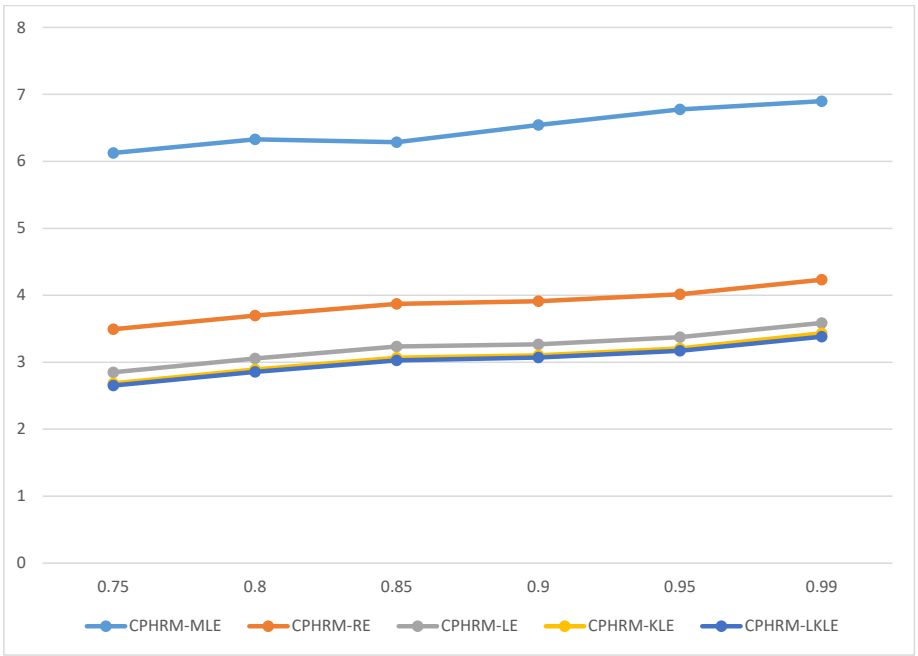


Figure 4. The performance of used methods when $p = 4$, $n = 500$ $cen = 0.2$

positive. The conditional index value that was obtained was 153.30. These results demonstrate that multicollinearity exists to a large extent.

The second data set is related to NCCTG lung cancer, which was obtained from https:\\www.kaggle.com\ukveteran\ncctg-lung-cancer-data. This data is comprised of survival data of advanced lung cancer patients in North Central cancer treatment group of 228 patients. The measures used in this data are time (Survival time in days) as response variable, y, age (zz1), sex (Male=1 and Female=2) (zz2), ECOG performance score with five scores (zz3), Karnofsky performance scorerated by physician (zz4), and weight loss in last six months (zz5). According to Seifollahi, Algamal and Arashi [31] and Seifollahi and Arashi [25], the condition number was quite large around 37.18 which means that there were a problem of multicollinearity.

Tables 5 and 6 contain the MSE values and estimated CPHRM coefficients for the CPHRM-MLE, CPHRM-RE, CPHRM-LE, CPHRM-LTE, CPHRM-KLE, and CPHRM-LKLE estimators for the first and the second data sets, respectively.

In Table 5, across all estimators, the shrinkage estimators CPHRM-RE, CPHRM-LE, CPHRM-LTE, CPHRM-KLE, and CPHRM-LKLE slightly pull the coefficients toward zero compared with CPHRM-MLE, reflecting variance reduction under multicollinearity. Among them, CPHRM-LKLE yields coefficients very close to CPHRM-MKL, with modest additional shrinkage, preserving the same clinical interpretation of risk factors. Crucially, CPHRM-LKLE attains the smallest overall MSE with 0.04173, improving on CPHRM-MLE with 0.04774 and also on the best existing shrinkage competitor, CPHRM-KLE with 0.04568. This indicates that CPHRM-LKLE offers more stable estimates of relapse risk across patients, which is particularly valuable when using histology and stage jointly for prognostic modeling and decision-making.

From Table 6, the shrinkage estimators again pull coefficients slightly toward zero relative to CPHRM-MLE, with CPHRM-LKLE typically providing the most moderated estimates while preserving the direction and relative magnitudes of the effects. For example, the hazard contrast between sexes softens from -0.5643 under CPHRM-MLE to -0.4917 under CPHRM-MKLE, and the ECOG effect decreases from 0.0202 to 0.0157, reflecting reduced instability in these strongly correlated predictors. Consistent with the simulation results, CPHRM-LKLE achieves the lowest overall MSE with 0.07486, improving upon CPHRM-MLE with 0.0867 and the CPHRM-KLE with 0.0778, which indicates more precise estimation of the underlying risk structure in this clinically complex dataset.

Table 5. National Wilm's tumor study results

variables	CPHRM-MLE	CPHRM-RE	CPHRM-LE	CPHRM-KLE	CPHRM-LKLE
z1	0.1857	0.1939	0.2091	0.2021	**0.2013**
z2	1.4685	1.4574	1.4367	1.4463	**1.4288**
z3	0.3517	0.3512	0.3502	0.3507	**0.3479**
z4	-0.0999	-0.0990	-0.0973	-0.0981	**-0.0971**
z5	0.0060	0.0060	0.0060	0.0060	**0.0059**
MSE	0.04774	0.04679	0.04610	0.04568	**0.04173**

Table 6. NCCTG lung cancer results

variables	CPHRM-MLE	CPHRM-RE	CPHRM-LE	CPHRM-KLE	CPHRM-LKLE
zz1	0.0116	0.0118	0.0118	0.0119	**0.1117**
zz2	-0.5643	-0.5286	-0.5254	-0.4930	**-0.4922**
zz3	0.0202	0.0182	0.0181	0.0163	**0.0155**
zz4	0.8340	0.7781	0.7730	0.7718	**0.7709**
zz5	-0.0117	-0.0110	-0.0110	-0.0104	**-0.0101**
MSE	0.0867	0.0812	0.0811	0.0778	**0.07504**

7. Conclusion

This paper proposed a variant of the Kibria-Lukman estimator of shrinkage of the CPHRM to overcome the negative effects of multicollinearity on the survival regression. The proposed CPHRM-LKLE reduces the bias variance trade-off by incorporating an extra ridge-type structure to the KL estimator with the selection of the biasing parameter based on an explicit mean squared error minimization criterion. Theory comparisons When mild regularity assumptions are satisfied, theoretical comparisons demonstrate that CPHRM-LKLE outperforms the classical maximum likelihood estimator and a variety of other known biased estimators with respect to mean squared error. Monte Carlo experiments have established that these benefits in the asymptotic sense of the word are robustly reflected in substantial finite-sample gains in a rich variety of practical contexts, including small or moderate sample sizes, strong pairwise correlations among covariates, variable amounts of predictors, and different degrees of censoring. CPHRM-LKLE will give the lowest mean squared error as the average in all of the simulated cases and the relative decrease in the error will be stronger as multicollinearity and censoring rise. The two empirical uses, which include the tumor relapse of Wilm and survival of advanced lung cancer, is another indication that CPHRM-LKLE is more accurate in the estimations of the coefficients, and the total mean squared error, but still provides clinically relevant direction of effects and magnitudes. Combined, this evidence suggests that CPHRM-LKLE is a useful and stable instrument to examine the medical time-to-event data when the prognostic variables are correlated, providing less volatile risk estimates and retaining the same interpretability. It is possible to further develop future work, which may be recommended to focus on high-dimensional penalized Cox models, stratified or time-varying coefficient Cox models, and other types of survival models that are often applied in contemporary biomedical studies.

REFERENCES

1. M. Cockeran, S.G. Meintanis, and J.S. Allison, *Goodness-of-fit tests in the Cox proportional hazards model*, Communications in Statistics-Simulation Computation, vol. 50, no. 12, pp. 4132–4143, 2021.
2. T. Emura, Y.-H. Chen, and H.-Y. Chen, *Survival prediction based on compound covariate under Cox proportional hazard models*, PloS One, vol. 7, no. 10, pp. e47627, 2012.
3. J.S. da Conceição et al., *Modified Weibull regression model for recurrent events: properties and application*, Communications in Statistics-Simulation and Computation, pp. 1–15, 2026.
4. U. Karabey and N.A. Tutkun, *Model selection criterion in survival analysis*, AIP Conference Proceedings, vol. 1863, pp. 120003, 2017.
5. C. Leng and H.H. Zhang, *Model selection in nonparametric hazard regression*, Journal of Nonparametric Statistics, vol. 18, no. 7-8, pp. 417–429, 2006.
6. A.A. Ewees et al., *A Cox Proportional-Hazards Model Based on an Improved Aquila Optimizer with Whale Optimization Algorithm Operators*, Mathematics, vol. 10, no. 8, 2022.
7. G.Y.I. Abdallh and Z.Y. Algamal, *An Investigational Modeling Approach for Improving Gene Selection using Regularized Cox Regression Model*, Mathematical Biology and Bioinformatics, vol. 18, no. 2, pp. 282–293, 2023.
8. Ewees, A. A., Algamal, Z. Y., Abualigah, L., Al-Qaness, M. A., Yousri, D., Ghoniem, R. M., Abd Elaziz, M. *Boosting Arithmetic Optimization Algorithm with Genetic Algorithm Operators for Feature Selection: Case Study on Cox Proportional Hazards Model*, Mathematics, vol. 9, no. 18, 2021.
9. Z. Algamal et al., *Modified Jackknifed Ridge Estimator in Bell Regression Model: Theory, Simulation and Applications*, Iraqi Journal for Computer Science and Mathematics, vol. 4, no. 1, pp. 146–154, 2023.
10. Z.Y. Algamal et al., *Performance of the Ridge and Liu Estimators in the zero-inflated Bell Regression Model*, Journal of Mathematics, vol. 2022, pp. 1–15, 2022.
11. Z.Y. Algamal, *Shrinkage parameter selection via modified cross-validation approach for ridge regression model*, Communications in Statistics - Simulation and Computation, vol. 49, no. 7, pp. 1922–1930, 2018.
12. S. Ahmad, M. Aslam, and S. Ahmad, *Extending the Liu estimator for the Cox proportional hazards regression model with multicollinearity*, Communications in Statistics - Simulation and Computation, pp. 1–14, 2023.
13. D. İnan and M. Tez, *A New Estimator for Cox Proportional Hazard Regression Model in Presence of Collinearity*, Communications in Statistics - Theory and Methods, vol. 41, no. 13-14, pp. 2437–2444, 2012.
14. X. Xue, M.Y. Kim, and R.E. Shore, *Cox regression analysis in presence of collinearity: an application to assessment of health risks associated with occupational radiation exposure*, Lifetime Data Analysis, vol. 13, no. 3, pp. 333–350, 2007.
15. Z.Y. Algamal, *A modified ridge estimator in Cox regression model*, Communications in Statistics-Simulation and Computation, pp. 1–8, 2025.
16. S. Seifollahi, Z.Y. Algamal, and M. Arashi, *New insights into multicollinearity in the Cox proportional hazard models: the Kibria-Lukman estimator and its application*, Journal of Applied Statistics, pp. 1–14, 2025.

17. S. Seifollahi and M. Arashi, *Improved survival analysis with shrinkage Kibria–Lukman estimators in the Cox model: Application to lung cancer data*, Statistical Methods in Medical Research, 2026.
18. B.B. Aladeitan et al., *Modified Kibria-Lukman (MKL) estimator for the Poisson Regression Model: application and simulation*, F1000Research, vol. 10, pp. 548, 2021.
19. D.R. Cox, *Regression models and life-tables*, Journal of the Royal Statistical Society: Series B, vol. 34, no. 2, pp. 187–202, 1972.
20. M.I. Ullah, M. Aslam, and S. Altaf, *lmridge: A Comprehensive R Package for Ridge Regression*, The R Journal, vol. 10, no. 2, pp. 326–346, 2018.
21. A.E. Hoerl and R.W. Kennard, *Ridge regression: Biased estimation for nonorthogonal problems*, Technometrics, vol. 12, no. 1, pp. 55–67, 1970.
22. K. Liu, *Using Liu-type estimator to combat collinearity*, Communications in Statistics - Theory and Methods, vol. 32, no. 5, pp. 1009–1020, 2003.
23. K. Liu, *A new class of biased estimate in linear regression*, Communications in Statistics - Theory and Methods, vol. 22, no. 2, pp. 393–402, 1993.
24. S. Ahmad et al., *Development of the Aslam-Ahmad estimator for the Cox proportional hazards regression model with multicollinearity*, Journal of Statistical Computation and Simulation, pp. 1–24, 2024.
25. S. Seifollahi and M. Arashi, *Improved survival analysis with shrinkage Kibria-Lukman estimators in the Cox model: Application to lung cancer data*, Statistical Methods in Medical Research, 2026.
26. B.G. Kibria and A.F. Lukman, *A new ridge-type estimator for the linear regression model: simulations and applications*, Scientifica, vol. 2020, no. 1, pp. 9758378, 2020.
27. M.R. Abonazel, A.A. Saber, and F.A. Awwad, *Kibria–Lukman estimator for the Conway–Maxwell Poisson regression model: Simulation and applications*, Scientific African, vol. 19, pp. e01553, 2023.
28. I. Dawoud, M.R. Abonazel, and F.A. Awwad, *Generalized Kibria-Lukman estimator: method, simulation, and application*, Frontiers in Applied Mathematics and Statistics, vol. 8, pp. 880086, 2022.
29. M.N. Akram et al., *Kibria–Lukman estimator for the zero inflated negative binomial regression model: theory, simulation and applications*, Communications in Statistics-Simulation and Computation, vol. 54, no. 5, pp. 1464–1480, 2025.
30. G.A. Shewa and F.I. Ugwuowo, *Kibria-Lukman type estimator for gamma regression model*, Concurrency and Computation: Practice and Experience, vol. 35, no. 1, pp. e7441, 2023.
31. S. Seifollahi, Z.Y. Algamal, and M. Arashi, *New insights into multicollinearity in the Cox proportional hazard models: the Kibria-Lukman estimator and its application*, Journal of Applied Statistics, vol. 52, no. 14, pp. 2672–2685, 2025.
32. A.F. Lukman et al., *A new ridge-type estimator for the gamma regression model*, Scientifica, vol. 2021, no. 1, pp. 5545356, 2021.
33. M.N. Akram et al., *A new modified ridge-type estimator for the beta regression model: simulation and application*, AIMS Mathematics, vol. 7, no. 1, pp. 1035–1057, 2021.
34. F.A. Almulhim et al., *New two parameter hybrid estimator for zero inflated negative binomial regression models*, Scientific Reports, vol. 15, no. 1, pp. 21239, 2025.
35. N.A. Alreshidi et al., *Modified Kibria–Lukman Estimator for the Conway–Maxwell–Poisson Regression Model: Simulation and Application*, Mathematics, vol. 13, no. 5, 2025.
36. H. Alrweili, *Kibria-Lukman Hybrid Estimator for the Conway–Maxwell–Poisson Regression Model*, Electronic Journal of Applied Statistical Analysis, vol. 17, no. 2, 2024.
37. A.T. Hammad et al., *A new effective biased estimator for Poisson regression model under multicollinearity*, Statistics, vol. 60, no. 2, pp. 543–581, 2025.
38. G.A. Shewa and F.I. Ugwuowo, *Kibria-Lukman type estimator for gamma regression model*, Concurrency and Computation: Practice and Experience, vol. 35, no. 1, 2022.
39. G.A. Shewa and F.I. Ugwuowo, *A new hybrid estimator for linear regression model analysis: Computations and simulations*, Scientific African, vol. 19, 2023.
40. Algboory, Ahmed Mutlag and Alkhateeb, Ahmed Naziyah and Algamal, Zakariya Yahya, *A Modified Jackknife Liu-Type Estimator for the Gamma Regression Models Data*, Statistics, Optimization & Information Computing, vol. 14, No. 5,pp. 2142–2154, 2025.
41. O.J. Oladapo et al., *A New Biased Two-Parameter Estimator in Linear Regression Model*, Equations, vol. 3, pp. 73–92, 2023.
42. G.C. McDonald and D.I. Galarneau, *A Monte Carlo evaluation of some ridge-type estimators*, Journal of the American Statistical Association, vol. 70, no. 350, pp. 407–416, 1975.
43. N.E. Breslow and N. Chatterjee, *Design and analysis of two-phase studies with binary outcome applied to Wilms tumour prognosis*, Journal of the Royal Statistical Society: Series C, vol. 48, no. 4, pp. 457–468, 1999.
44. Alkhateeb, Ahmed Naziyah and Al-Qazaz, Qutaiba N Nayef, *Variable Selection in Weibull Accelerated Survival Model Based on Chaotic Sand Cat Swarm Algorithm*, Statistics, Optimization & Information Computing, vol. 13, no. 5, pp. 2105–2118, 2025.
45. Algamal, Z. Y. *Efficiency of a hybrid Kibria–Lukman estimator in the Cox regression model with survival data application* , Hacettepe Journal of Mathematics and Statistics,(Advanced Online Publication), p. 1-16, 2026.
46. ALbazzaz, Z. M., Younus, F. A., & Algamal, Z. Y. *Dawoud–Kibria estimator for the Cox regression model*, Computational Journal of Mathematical and Statistical Sciences, 6(2), p. 79-97, 2026
47. Al-Taweel, Y., & Algamal, Z. *Some almost unbiased ridge regression estimators for the zero-inflated negative binomial regression model*, Periodicals of Engineering and Natural Sciences, vol. 8, no.(1), p.248-255, 2020
48. Algamal, Z. Y., & Lee, M. H. *Applying penalized binary logistic regression with correlation based elastic net for variables selection*, Journal of Modern Applied Statistical Methods, vol. 14, no.(1), p.15, 2015
49. Algamal, Z. Y., Lee, M. H., Al-Fakih, A. M., & Aziz, M. *High-dimensional QSAR modelling using penalized linear regression model with L 1/2-norm*, SAR and QSAR in Environmental Research, vol. 27, no.(9), p.703-719, 2016
50. Shamany, R., Alobaidi, N. N., & Algamal, Z. Y. *A new two-parameter estimator for the inverse Gaussian regression model with application in chemometrics*, Electronic Journal of Applied Statistical Analysis, vol. 12, no.(2), p.453-464, 2019
51. Algamal, Z. Y., Qasim, M. K., Lee, M. H., & Ali, H. T. M. *QSAR model for predicting neuraminidase inhibitors of influenza A viruses (H1N1) based on adaptive grasshopper optimization algorithm*, SAR and QSAR in Environmental Research, vol.31, no.11,

p.803-814, 2020.
52. Al-Fakih, A. M., Algamal, Z. Y., Lee, M. H., Aziz, M., Ali, H. T. M. *QSAR classification model for diverse series of antifungal agents based on improved binary differential search algorithm*, SAR and QSAR in Environmental Research, vol.30, no.2, p.131-143, 2019.
53. Mustafa, M. Y., & Algamal, Z. Y. *Neutrosophic inverse power Lindley distribution: A modeling and application for bladder cancer patients*, International Journal of Neutrosophic Science, 21(2), p. 216-223, 2023
54. Algamal, Z. Y., Abonazel, M. R., Awwad, F. A., & Eldin, E. T. *Modified jackknife ridge estimator for the Conway-Maxwell-Poisson model*, Scientific African, 19, e01543, 2023
55. Alobaidi, N. N., Shamany, R. E., & Algamal, Z. Y. *A new ridge estimator for the negative binomial regression model*, Thailand Statistician, 19(1), 115-124, 2021
56. Bibani, A. A., & Algamal, Z. Y. *Survival function estimation for fuzzy Gompertz distribution with neutrosophic data*, International Journal of Neutrosophic Science, 21(3), p.137-42, 2023
57. Lukman, A. F., Dawoud, I., Kibria, B. G., Algamal, Z. Y., & Aladeitan, B. *A new ridge-type estimator for the gamma regression model* Scientifica, vol. 1, p.5545356, 2021.
58. Algamal, Z. Y., Alhamzawi, R.,& Ali, H. T. M. *Gene selection for microarray gene expression classification using Bayesian Lasso quantile regression*, Computers in biology and medicine, vol. 97, p. 145-152, 2018
59. Algamal, Z., & Ali, H. M. *An efficient gene selection method for high-dimensional microarray data based on sparse logistic regression* , Electronic Journal of Applied Statistical Analysis, vol . 10 , no . 1 , 242-256 , 2017.
60. Algamal, Z. Y., & Lee, M. H. *A new adaptive L1-norm for optimal descriptor selection of high-dimensional QSAR classification model for anti-hepatitis C virus activity of thiourea derivatives* , SAR and QSAR in Environmental Research, vol. 28, no. 1 , p. 75-90, 2017.
61. Algamal, Z. Y., & Lee, M. H. *A novel molecular descriptor selection method in QSAR classification model based on weighted penalized logistic regression*, Journal of Chemometrics, vol. 31, no.(10), e2915, 2017
62. Qasim, M. K., Algamal, Z. Y., & Ali, H. M. *A binary QSAR model for classifying neuraminidase inhibitors of influenza A viruses (H1N1) using the combined minimum redundancy maximum relevancy criterion with the sparse support vector machine*, SAR and QSAR in Environmental Research, vol. 29, no.(7), p.517-527, 2018
63. Awwad, F. A., Odeniyi, K. A., Dawoud, I., Algamal, Z. Y., Abonazel, M. R., Kibria, B. G., & Eldin, E. T. *New two-parameter estimators for the logistic regression model with multicollinearity*, WSEAS Transactions on Mathematics, vol. 21, p.403-414, 2022
64. Algamal, Z. Y., Qasim, M. K., & Ali, H. T. M. *A QSAR classification model for neuraminidase inhibitors of influenza A viruses (H1N1) based on weighted penalized support vector machine* , SAR and QSAR in Environmental Research, vol. 28 , no. 5, p. 415-426 , 2017.