STATISTICS, OPTIMIZATION AND INFORMATION COMPUTING
Stat., Optim. Inf. Comput., Vol. **x**, Month 202x, pp 0–11.
Published online in International Academic Press (www.IAPress.org)

Modified cross-validation procedure in selection shrinkage parameter of Poisson ridge regression model

Mazin M. Alanaz [1], Zakariya Yahya Algamal [2,*]

[1]*Department of Operation Research and Intelligence Techniques, University of Mosul, Iraq*
[2]*Department of Statistics and Informatics, University of Mosul, Mosul, Iraq*

Abstract Poisson regression is a fundamental framework for modeling count data; however, the maximum likelihood estimator becomes unreliable under multicollinearity, yielding unstable estimates with large variance and weak inferential performance. Ridge regression offers a remedy by shrinking regression coefficients, but its effectiveness depends crucially on selecting an appropriate shrinkage parameter. This paper proposes a modified cross-validation (MCV) procedure for Poisson ridge regression that explicitly targets this instability. The method repeatedly reassigns observations to folds, re-computes the optimal shrinkage parameter over many replications, and then uses an appropriately chosen quantile of the resulting distribution of optimal values as the final tuning parameter; the quantile is itself selected via cross-validated prediction error. Extensive Monte Carlo simulations, conducted over varying sample sizes, numbers of predictors, intercept values, and correlation levels, show that the MCV-based Poisson ridge estimator consistently achieves lower mean squared error than maximum likelihood, standard cross-validation, and generalized cross-validation, with particularly notable gains under strong multicollinearity. A real data analysis further demonstrates that the proposed procedure improves prediction performance in practical count data applications.

Keywords Multicollinearity, ridge regression, cross-validation, shrinkage, Monte Carlo simulation

DOI: 10.19139/soic-2310-5070-3321

1. Introduction

In regression modeling, data in the form of counts are usually common. “Count data regression modeling has received much attention in medicine, behavioral sciences, psychology, and econometrics [1, 2, 3]. The Poisson regression model is the most basic models under count data regression models [4, 34, 35, **?**].

In dealing with the Poisson regression model, it is assumed that there is no correlation among the explanatory variables[36, 37, 38, 39, 40]. In practice, however, this assumption often not holds, which leads to the problem of multicollinearity. In the presence of multicollinearity, when estimating the regression coefficients for Poisson regression model using the maximum likelihood (ML) method, the estimated coefficients are usually become unstable with a high variance, and therefore low statistical significance with incorrect signs [5, 6, 7, 44].

Numerous remedial methods have been proposed to overcome the problem of multicollinearity. The ridge regression method [8, 46, 45, 47] has been consistently demonstrated to be an attractive and alternative to the ML estimation method.

Ridge regression is a shrinkage method that shrinks all regression coefficients toward zero to reduce the large variance [9]. The ridge regression performance greatly relies on the choice of shrinkage parameter. Consequently, choosing a suitable value of the shrinkage parameter is an important part of ridge regression model fitting [10].

*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)

Several methods, which they are based on the original ridge regression of [8], are available for estimating the ridge shrinkage parameter in the literature [11, 12, 13, 14, 15, 16, 17]. Cross-validation method (CV), a data-driven approach, on the other hand, is a practically useful approach for handling the shrinkage selection problem in ridge regression [18, 48, 49]. This is due to the attractive property of the CV, which does not assume any underlying distribution about the data. Furthermore, CV can consider a natural choice when the target of model fitting is prediction [19].

The idea behind the CV is to randomly split the data into k mutually exclusive folds of approximately equal size. Among the k folds, one fold is retained as validation data set for testing the model fitting, and the remaining $k-1$ folds are used as training data set to fit the model with a specific value of the shrinkage parameter [50, 51, 52, 53]. Then, the prediction performance over these splits is averaged to represent the predictability of the fitted model. After that, the best value of the shrinkage parameter is the corresponding value to the small prediction error. It is clear that CV method is greatly dependent on the fold assignment process which leads to large variability in selecting the shrinkage parameter value and, consequently, will negatively affect the prediction performance of the ridge model.

In this paper, a modification of CV is proposed to address the variability of shrinkage parameter selection. This modification is based on repeated fold assignment. And then, a proper quantile value of the best shrinkage parameter values, which are obtained over the repeated fold assignment, is utilized. Due to this proposed modification, the shrinkage parameter selection is shown to have better performance in terms of model prediction.

2. Poisson ridge regression model

Count data often arise in epidemiology, social, and economic studies. This type of data consists of positive integer values. Poisson distribution is a well-known distribution that fit to such type of data. Poisson regression model is used to model the relationship between the counts as response variable and potentially explanatory variables [16, 20, 21, 22, 23, 24, 25, 26].

Let y_i be the response variable and follows a Poisson distribution with mean ϕ_i, then the probability density function is defined as

$$f\left(y_i\right)=\frac{e^{-\phi_i}{\phi_i}^{y_i}}{y_i!}, \quad y_i=0,1,\ldots;\ i=1,2,\ldots,n. \tag{1}$$

In a Poisson regression model, $\ln(\phi_i)=\mathbf{x}_i^T\boldsymbol{\beta}$ is expressed as a linear combination of explanatory variables $\mathbf{x}_i=(x_{i1},...,x_{ip})^T$. The $\ln(\phi_i)$ is called as canonical link function which making the relationship between explanatory variables and response variable linear. The most common method of estimating the coefficients of Poisson regression model is to use the maximum likelihood method. Given the assumption that the observations are independent, the log-likelihood function is defined as

$$\ell(\boldsymbol{\beta})=\sum_{i=1}^{n}\left\{y_i\mathbf{x}_i^T\boldsymbol{\beta}-\exp(\mathbf{x}_i^T\boldsymbol{\beta})-\ln y_i!\right\}. \tag{2}$$

The ML estimator is then obtained by computing the first derivative of the Eq. (3) and setting it equal to zero, as

$$\frac{\partial\ell(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}}=\sum_{i=1}^{n}\left[y_i-\exp(\mathbf{x}_i^T\boldsymbol{\beta})\right]\mathbf{x}_i=0. \tag{3}$$

Because Eq. (4) is nonlinear in $\boldsymbol{\beta}$, the iteratively weighted least squares (IWLS) algorithm can be used to obtain the ML estimators of the Poisson regression parameters (PR) as

$$\hat{\boldsymbol{\beta}}_{ML}=(\mathbf{X}^T\hat{\mathbf{W}}\mathbf{X})^{-1}\mathbf{X}^T\hat{\mathbf{W}}\hat{\mathbf{v}}, \tag{4}$$

where $\hat{\mathbf{W}}=\mathrm{diag}(\hat{\phi}_i)$ and $\hat{\mathbf{v}}$ is a vector where ith element equals to $\hat{v}_i=\ln(\hat{\phi}_i)+((y_i-\hat{\phi}_i)/\hat{\phi}_i)$. The ML estimator is asymptotically normally distributed with a covariance matrix that corresponds to the inverse of the Hessian matrix

$$cov(\hat{\boldsymbol{\beta}}_{ML})=\left[-E\left(\frac{\partial^2\ell(\boldsymbol{\beta})}{\partial\beta_i\,\partial\beta_k}\right)\right]^{-1}=(\mathbf{X}^T\hat{\mathbf{W}}\mathbf{X})^{-1}. \tag{5}$$

The mean squared error (MSE) of Eq. (5) can be obtained as

$$\begin{aligned} \text{MSE}\,(\hat{\boldsymbol{\beta}}_{ML}) &= E(\hat{\boldsymbol{\beta}}_{ML} - \hat{\boldsymbol{\beta}})^T(\hat{\boldsymbol{\beta}}_{ML} - \hat{\boldsymbol{\beta}}) \\ &= tr[(\mathbf{X}^T\hat{\mathbf{W}}\mathbf{X})^{-1}] \\ &= \textstyle\sum_{j=1}^{p} \frac{1}{\lambda_j}, \end{aligned} \tag{6}$$

where λ_j is the eigenvalue of the $\mathbf{X}^T\hat{\mathbf{W}}\mathbf{X}$ matrix.

In the presence of multicollinearity, the matrix $\mathbf{X}^T\hat{\mathbf{W}}\mathbf{X}$ becomes ill-conditioned leading to high variance and instability of the ML estimator of the Poisson regression parameters. As a remedy, Månsson and Shukur [27] proposed the Poisson ridge estimator (PRE) as

$$\begin{aligned} \hat{\boldsymbol{\beta}}_{PRE} &= (\mathbf{X}^T\hat{\mathbf{W}}\mathbf{X} + h\mathbf{I})^{-1}\mathbf{X}^T\hat{\mathbf{W}}\mathbf{X}\hat{\boldsymbol{\beta}}_{ML} \\ &= (\mathbf{X}^T\hat{\mathbf{W}}\mathbf{X} + h\mathbf{I})^{-1}\mathbf{X}^T\hat{\mathbf{W}}\hat{\mathbf{v}}, \end{aligned} \tag{7}$$

where $h \geq 0$. The ML estimator can be considered as a special estimator from Eq. (7) with $h = 0$. Regardless of $h$value, the MSE of the $\hat{\boldsymbol{\beta}}_{PRE}$is smaller than that of $\hat{\boldsymbol{\beta}}_{ML}$because the MSE of $\hat{\boldsymbol{\beta}}_{PRE}$is equal to [28]

$$\text{MSE}(\hat{\boldsymbol{\beta}}_{PRE}) = \sum_{j=1}^{p} \frac{\lambda_j}{(\lambda_j + h)^2} + h^2 \sum_{j=1}^{p} \frac{\alpha_j}{(\lambda_j + h)^2}, \tag{8}$$

where α_j is defined as the jth element of $\gamma\hat{\boldsymbol{\beta}}_{ML}$and γ is the eigenvector of the $\mathbf{X}^T\hat{\mathbf{W}}\mathbf{X}$ matrix. Comparing with the MSE of Eq. (6), $\text{MSE}(\hat{\boldsymbol{\beta}}_{PRE})$ is always small for $h > 0$.

3. The proposed method

The efficiency of ridge regression model strongly depends on appropriately choosing the shrinkage parameter. A choice of shrinkage parameter that is too small leads to overfitting the PRE, while shrinkage parameter that is too large shrinks $\boldsymbol{\beta}$ by too much, making a bias-variance tradeoff [29, 54, 55]. Numerous methods, which they are based on the original ridge regression of Hoerl and Kennard [8], are available for estimating the ridge shrinkage parameter in the literature [11, 12, 13, 14, 15, 16, 17, 30]. Furthermore, information criteria, such as the Akaike information criterion (AIC) and the Bayesian information criterion (BIC) can be also used to select the ridge shrinkage parameter [29, 56, 57].

Cross-validation is one of the most widely used methods to select the shrinkage parameter. In practice, the k-fold CV (k-CV) is widely used where the data is partitioned into k nearly equal size folds. In these k folds, one fold will leave as the test set and use the remaining k $-$ 1 folds as the training set. Start with a grid containing the initial values of the shrinkage parameter, h, the k-CV is calculated for each h. Then, the optimal value of h is the one that has the minimum CV prediction error. The k-CV method was extended to its generalized form, generalized cross-validation (GCV), by Golub, Heath [30].

The k-CV is defined as:

$$\text{k}-\ \text{CV}_{(h)} = \frac{1}{k}\sum_{i=1}^{k}(y_i - \hat{y}_i{}_{(h)}^{-k(i)})^2, \tag{9}$$

where $\hat{y}_i{}_{(h)}^{-k(i)} = \mathbf{x}_i^{-k(i)}\hat{\boldsymbol{\beta}}_{PRE(h)}^{-k(i)}$ is the fitted ridge regression model. Then the optimal value of λ can be obtained by minimizing Eq. (9) as

$$h_{\ optimal} = \underset{r=1,2,\ldots,R}{\arg\min}\ \text{k}-\ \text{CV}_{(h_r)}. \tag{10}$$

It is worth mentioning that CV method is greatly dependent on the fold assignment process which leads to large variability in selecting the shrinkage parameter value and, consequently, will negatively affect the prediction performance of the ridge model.

To enhance k-CV method and to decrease its variability in selecting the shrinkage parameter value, a modification is proposed. Specifically, the steps involved in our modification of k-CV are summarized in the following algorithm:
Algorithm: Complete implementation of the proposed modification
Step 1: Assign observations to folds for CV randomly.
Step 2: Fit the ridge regression using this fold assignment.
Step 3: Let $\hat{h}_r$ represents the optimal shrinkage parameter value obtained from the ridge regression model.
Step 4: Repeat steps 1 – 3 for R times
Step 5: Let $S(R) = (\hat{h}_1, \hat{h}_2, ..., \hat{h}_R)$ represents the R values of $\hat{h}_r$.
Step 6: Let Π represents a sequence of π values, where $0 < \pi < 1$ is the quantile of $S(R)$.
Step 6: Determine $\hat{h}(\pi)$, the quantile of $S(R)$.
Step 7: Find the ridge regression estimator by fitting the ridge regression with the optimum shrinkage parameter, $\hat{h}(\pi)$.
Step 8: Calculate the CV prediction error of the Step 7.
Step 9: Repeat steps 6 – 8 for $\pi \in \Pi$.
Step 10: Select the best quantile, $\hat{\pi}$, with the smallest CV prediction error.
Step 11: Find the optimum ridge regression estimator by fitting the ridge regression with optimum shrinkage parameter, $\hat{h}(\hat{\pi})$.

Obviously, the quantile value will alleviate the CV variability in shrinkage selection because of the re-estimation of the ridge regression.

4. Simulation results

In this section, a Monte Carlo simulation experiment is used to examine the performance of the new estimator with different degrees of multicollinearity. The response variable of n observations is generated from Poisson regression model by:

$$\phi_i = \exp(\mathbf{x}_i^T \boldsymbol{\beta}), \tag{11}$$

where $\boldsymbol{\beta} = (\beta_0, \beta_1, ..., \beta_p)$ with $\sum_{j=1}^{p} \beta_j^2 = 1$ and $\beta_1 = \beta_2 = ... = \beta_p$ [14, 27]. In addition, because the value of intercept, β_0, has an effect on ϕ_i, three values are chosen $\beta_0 \in \{1, 0, -1\}$, where decreasing the value of β_0 leads to lower average value of ϕ_i, which leads to less variation [58, 59, 60].

The explanatory variables $\mathbf{x}_i^T = (x_{i1}, x_{i2}, ..., x_{in})$ have been generated from the following formula

$$x_{ij} = (1 - \rho^2)^{1l2} w_{ij} + \rho w_{ip}, \quad i = 1, 2, ..., n, \quad j = 1, 2, ..., p, \tag{12}$$

where ρ represents the correlation between the explanatory variables and w_{ij}'s are independent standard normal pseudo-random numbers. Because the sample size has direct impact on the prediction accuracy, three representative values of the sample size are considered: 30, 50 and 100. In addition, the number of the explanatory variables is considered as $p = 4$ and $p = 8$ because increasing the number of explanatory variables can lead to increase the MSE. Further, because we are interested in the effect of multicollinearity, in which the degrees of correlation considered more important, three values of the pairwise correlation are considered with $\rho = \{0.90, 0.95, 0.99\}$. For a combination of these different values of n, p, β_0, and ρ the generated data is repeated 1000 times and the averaged mean squared errors (MSE) is calculated as:

$$\text{MSE}(\hat{\boldsymbol{\beta}}) = \frac{1}{1000} \sum_{i=1}^{1000} (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})^T (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}), \tag{13}$$

where $\hat{\boldsymbol{\beta}}$ is the estimated coefficients for the used estimator.

The estimated MSE of Eq. (13) for CV, GCV, and our proposed method (MCV), for the combination of n, p, β_0, and ρ, are respectively summarized in tables 1, 2, and 3. Several observations can be made.

First, in terms of ρ values, there is increasing in the MSE values when the correlation degree increases regardless the value of n, p, β_0. However, MCV performs better than CV, GCV, and MLE. For instance, in Table 1, when

$p = 4$, $n = 100$, and $\rho = 0.95$, the MSE of MCV was about 56.74%, 39.11%, and 20.82% lower than that of MLE, CV, and GCV respectively.

Second, regarding the number of explanatory variables, it is easily seen that there is increasing in the MSE values when the p increasing from four variables to eight variables. Although this increasing can affected the quality of an estimator, MCV is achieved the lowest MSE comparing with MLE, CV, and GCV, for different n, ρ, β_0.

Third, with respect to the value of n, The MSE values decreases when n increases, regardless the value of ρ, p, β_0. However, MCV still consistently outperforms CV and GCV by providing the lowest MSE.

Fourth, in terms of the value of the intercept and for a given values of ρ, p, n, MCV is always show smaller MSE comparing with the other methods.

To summary, all the considered values of n, ρ, p, β_0, MCV is superior to CV, clearly indicating that the new proposed estimator is more efficient.

Table 1. MSE values when $\beta_0 = -1$

			ML	CV	GCV	MCV
		ρ				
$p = 4$	$n = 30$	0.90	6.249	6.002	5.735	5.135
		0.95	6.877	6.63	6.363	5.763
		0.99	7.275	7.028	6.761	6.161
	$n = 50$	0.90	4.62	4.373	4.106	3.506
		0.95	5.695	5.448	5.181	4.581
		0.99	5.887	5.64	5.373	4.773
	$n = 100$	0.90	4.463	4.216	3.949	3.349
		0.95	4.673	4.426	4.159	3.559
		0.99	5.428	5.181	4.914	4.314
$p = 8$	$n = 30$	0.90	6.354	6.107	5.84	5.24
		0.95	6.973	6.726	6.459	5.859
		0.99	7.388	7.141	6.874	6.274
	$n = 50$	0.90	4.889	4.642	4.375	3.775
		0.95	6.032	5.785	5.518	4.918
		0.99	6.357	6.11	5.843	5.243
	$n = 100$	0.90	4.799	4.552	4.285	3.685
		0.95	5.074	4.827	4.56	3.96
		0.99	5.632	5.385	5.118	4.518

Table 2. MSE values when $\beta_0 = 0$

			ML	CV	GCV	MCV
		ρ				
$p = 4$	$n = 30$	0.90	6.2731	6.0261	5.7591	5.1591
		0.95	6.9011	6.6541	6.3871	5.7871
		0.99	7.2991	7.0521	6.7851	6.1851
	$n = 50$	0.90	4.6441	4.3971	4.1301	3.5301
		0.95	5.7191	5.4721	5.2051	4.6051
		0.99	5.9111	5.6641	5.3971	4.7971
	$n = 100$	0.90	4.4871	4.2401	3.9731	3.3731
		0.95	4.6971	4.4501	4.1831	3.5831
		0.99	5.4521	5.2051	4.9381	4.3381
$p = 8$	$n = 30$	0.90	6.3781	6.1311	5.8641	5.2641
		0.95	6.9971	6.7501	6.4831	5.8831
		0.99	7.4121	7.1651	6.8981	6.2981
	$n = 50$	0.90	4.9131	4.6661	4.3991	3.7991
		0.95	6.0561	5.8091	5.5421	4.9421
		0.99	6.3811	6.1341	5.8671	5.2671
	$n = 100$	0.90	4.8231	4.5761	4.3091	3.7091
		0.95	5.0981	4.8511	4.5841	3.9841
		0.99	5.6561	5.4091	5.1421	4.5421

Moreover, Figures 1-6 show the performance of the used method over several correlation values. It is clear from these Figures that the proposed method, MCV, exhibits very good performance compared to ML, CV, and GCV.

Table 3. MSE values when $\beta_0 = 1$

			ML	CV	GCV	MCV
		ρ				
$p = 4$	$n = 30$	0.90	6.471	6.224	5.957	5.357
		0.95	7.099	6.852	6.585	5.985
		0.99	7.497	7.25	6.983	6.383
	$n = 50$	0.90	4.842	4.595	4.328	3.728
		0.95	5.917	5.67	5.403	4.803
		0.99	6.109	5.862	5.595	4.995
	$n = 100$	0.90	4.685	4.438	4.171	3.571
		0.95	4.895	4.648	4.381	3.781
		0.99	5.65	5.403	5.136	4.536
$p = 8$	$n = 30$	0.90	6.576	6.329	6.062	5.462
		0.95	7.195	6.948	6.681	6.081
		0.99	7.61	7.363	7.096	6.496
	$n = 50$	0.90	5.111	4.864	4.597	3.997
		0.95	6.254	6.007	5.74	5.14
		0.99	6.579	6.332	6.065	5.465
	$n = 100$	0.90	5.021	4.774	4.507	3.907
		0.95	5.296	5.049	4.782	4.182
		0.99	5.854	5.607	5.34	4.74

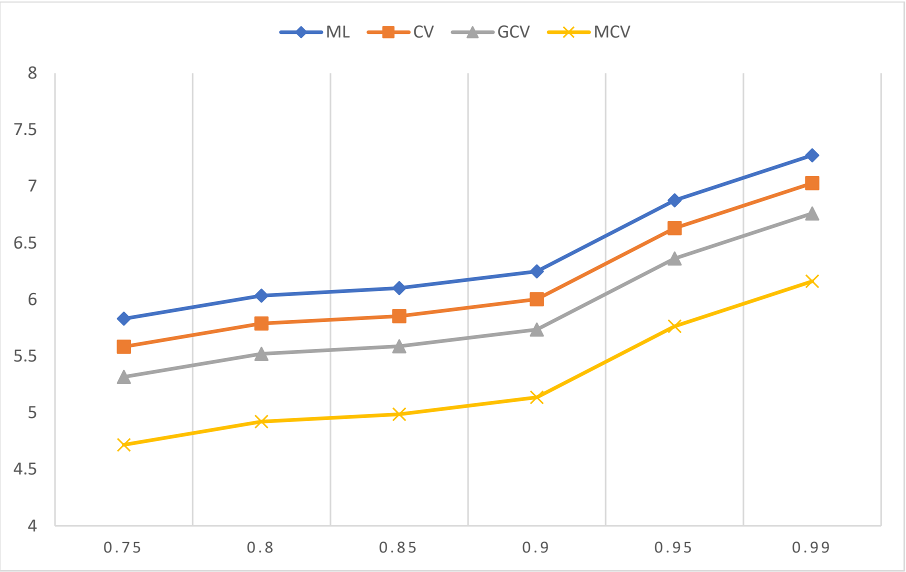

Figure 1. The performance of used methods when $p = 4$, $n = 30$ $\beta_0 = -1$

5. Real application

To investigate the usefulness of reviewed biased estimators, an application related to the football English league, season 2016-2017 is employed. This data contains twenty teams, where the response variable represents the number of won matches. The six considerable predictors included the number of yellow cards (x_1), the number of red cards (x_2), the total number of substitutions (x_3), the number of matches with 2.5 goals on average (x_4), the number of matches that ended with goals (x_5), and the ratio of the goal scores to the number of matches (x_6).

First, the deviance test [33] is used to check whether the Poisson regression model is fit well to this data or not. The result of the residual deviance test is equal to 8.373 with 14 degrees of freedom and the p-value is 0.869. It is indicated form this result that the Poisson regression model fits very well to this data.

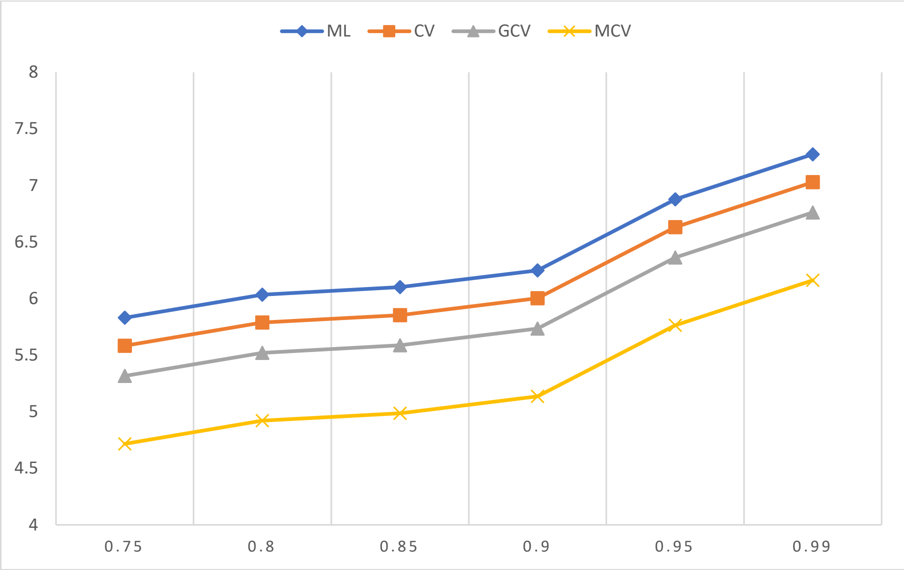

Figure 2. The performance of used methods when $p = 4$, $n = 50$ $\beta_0 = -1$

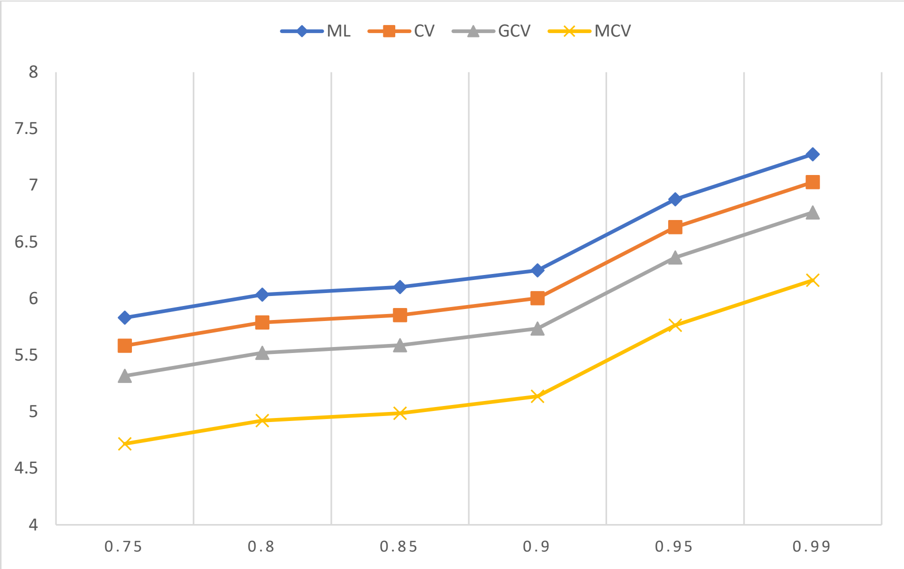

Figure 3. The performance of used methods when $p = 4$, $n = 100$ $\beta_0 = -1$

Second, to check whether there are relationships between the explanatory variables or not, Figure 7 displays the correlation matrix among the six explanatory variables. It is obviously seen that there are correlations greater than 0.86 between x_1 and x_6, x_1 and x_4, x_2 and x_4, and, x_4 and x_6.

Third, to test the existence of multicollinearity, the eigenvalues of the matrix $\mathbf{X}^T\hat{\mathbf{W}}\mathbf{X}$ are obtained as 993.758, 142.907, 75.560, 38.999, 21.424, and 1.016. The determined condition number $\text{CN} = \sqrt{\lambda_{\max}/\lambda_{\min}}$ of the data is 31.274 indicating that the multicollinearity issue is exist.

The estimated Poisson regression coefficients, standard errors which are computed by using bootstrap with 500 replications, and MSE values for the ML, CV, GCV, and MCV estimators are listed in Table 4. According to

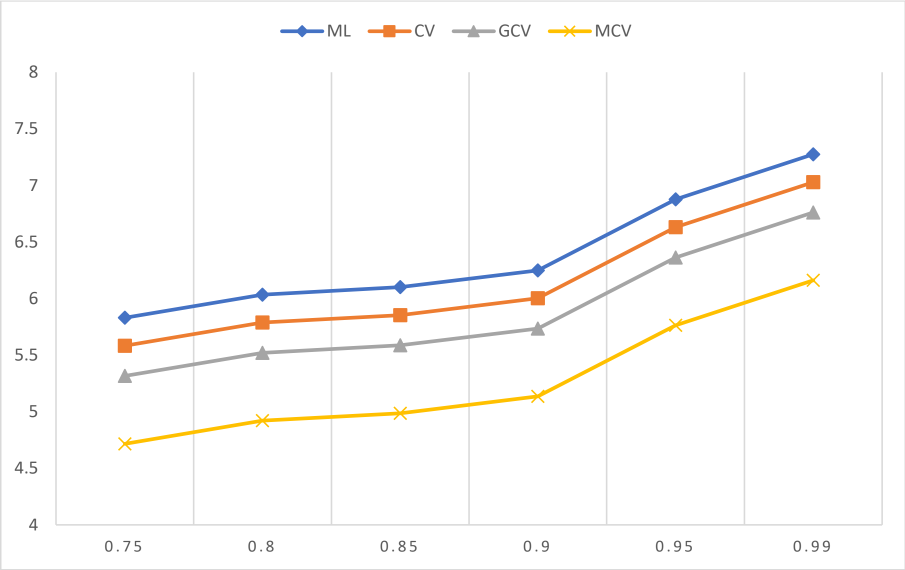

Figure 4. The performance of used methods when $p = 8$, $n = 30$ $\beta_0 = -1$

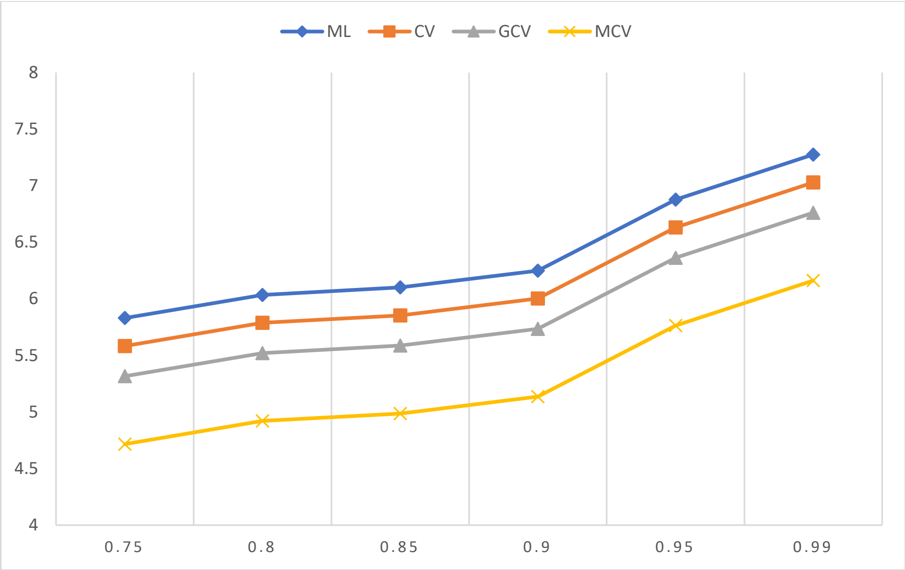

Figure 5. The performance of used methods when $p = 8$, $n = 50$ $\beta_0 = -1$

Table 4, it is clearly seen that the MCV estimator shrinkages the value of the estimated coefficients efficiently. Additionally, in terms of the calculated standard errors, the MCV and GCV show substantial decreasing comparing with MLE.

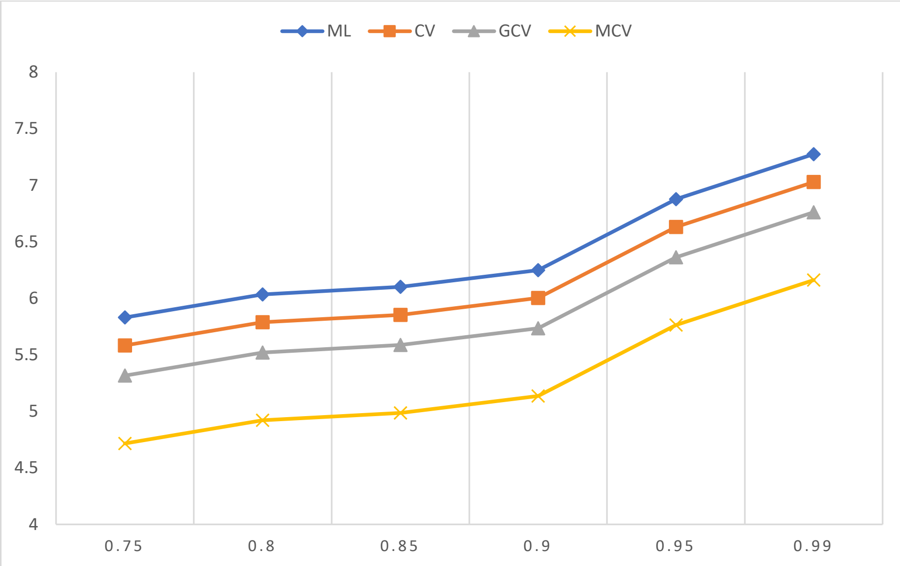

Figure 6. The performance of used methods when $p = 8$, $n = 100$ $\beta_0 = -1$

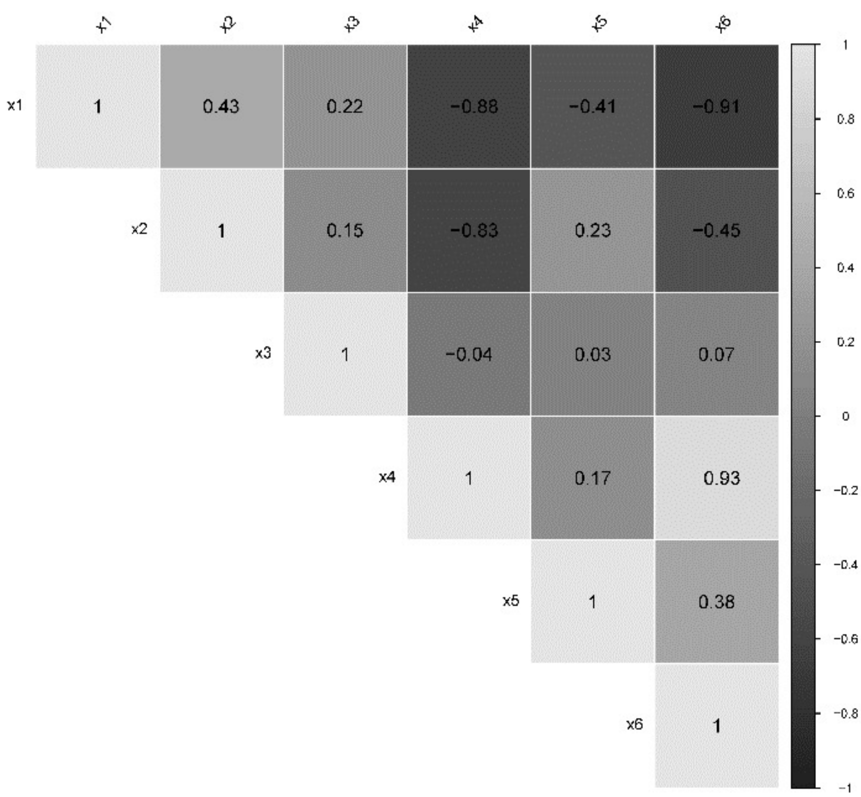

Figure 7. The correlation matrix among the six explanatory variables

Table 4. The estimated coefficients and MSE values for the ML, CV, GCV, and MCV estimators. The number in parenthesis is the standard error".

	MLE	CV	GCV	MCV
$\hat{\beta}_1$	-1.219	-1.016	-0.915	-0.887
	(0.151)	(0.007)	(0.011)	(0.127)
$\hat{\beta}_2$	0.441	0.510	4.004	0.024
	(0.151)	(0.001)	(0.002)	(0.139)
$\hat{\beta}_3$	0.575	0.416	0.412	0.276
	(0.175)	(0.008)	(0.008)	(0.108)
$\hat{\beta}_4$	-3.476	-2.034	-2.014	-0.147
	(0.313)	(0.008)	(0.007)	(0.231)
$\hat{\beta}_5$	-2.432	-2.12	-2.017	-1.011
	(0.160)	(0.004)	(0.003)	(0.133)
$\hat{\beta}_6$	5.121	0.004	0.003	0.184
	(0.287)	(0.003)	(0.001)	(0.230)
MSE	3.681	2.108	2.041	1.221

6. Conclusion

The study establishes that instability in cross-validation–based tuning is a major source of inefficiency for Poisson ridge regression when multicollinearity is present, and that explicitly targeting this instability can substantially improve estimation and prediction. By repeatedly reassigning folds, aggregating the resulting optimal shrinkage parameters, and selecting a suitable quantile in a data-driven manner, the proposed modified cross-validation (MCV) procedure produces a more stable shrinkage choice and, consequently, a more reliable Poisson ridge estimator. Across an extensive Monte Carlo design that varies sample size, number of predictors, correlation strength, and intercept values, the MCV-based estimator consistently yields the smallest mean squared error compared with maximum likelihood, standard cross-validation, and generalized cross-validation, with the advantage becoming more pronounced as multicollinearity strengthens or the number of predictors increases. These gains persist regardless of the intercept specification, indicating that the proposed method is robust to different mean structures in the Poisson model. The real data application confirms that these improvements translate into better predictive performance in practice, reinforcing the practical value of MCV as a tuning strategy for count data regression in applied fields such as epidemiology, social sciences, and economics. Overall, the results suggest that incorporating resampling-based stabilization and quantile selection into cross-validation offers a simple yet powerful extension of existing ridge tuning methods for Poisson regression.

REFERENCES

1. Z.Y. Algamal, *Diagnostic in Poisson regression models*, Electronic Journal of Applied Statistical Analysis, vol. 5, no. 2, pp. 178–186, 2012.
2. Y. Asar, and A. Genç, *A New Two-Parameter Estimator for the Poisson Regression Model*, Iranian Journal of Science and Technology, Transactions A: Science, vol. 42, no. 2, pp. 793–803, 2017.
3. S. Coxe, S.G. West, and L.S. Aiken, *The analysis of count data: a gentle introduction to poisson regression and its alternatives*, J Pers Assess, vol. 91, no. 2, pp. 121–136, 2009.
4. Z. Wang, et al., *Penalized count data regression with application to hospital stay after pediatric cardiac surgery*, Stat. Meth. Med. Res., In press, 2014.
5. Y.-J. Jou, C.-C.L. Huang, and H.-J. Cho, *A VIF-based optimization model to alleviate collinearity problems in multiple linear regression*, Computational Statistics, vol. 29, no. 6, pp. 1515–1541, 2014.
6. M.I. ALheety, and B.M.G. Kibria, *A generalized stochastic restricted ridge regression estimator*, Communications in Statistics—Theory and Methods, vol. 43, pp. 4415–4427, 2014.
7. F.S.M. Batah, M.R. Özkale, and S.D. Gore, *Combining unbiased ridge and principal component regression estimators*, Communications in Statistics - Theory and Methods, vol. 38, no. 13, pp. 2201–2209, 2009.
8. A.E. Hoerl, and R.W. Kennard, *Ridge regression: Biased estimation for nonorthogonal problems*, Technometrics, vol. 12, no. 1, pp. 55–67, 1970.

9. Y. Asar, and A. Genç, *New shrinkage parameters for the Liu-type logistic estimators*, Communications in Statistics - Simulation and Computation, vol. 45, no. 3, pp. 1094–1103, 2015.
10. T. Sökūt Açar, and M.R. Özkale, *Cross validation of ridge regression estimator in autocorrelated linear regression models*, Journal of Statistical Computation and Simulation, vol. 86, no. 12, pp. 2429–2440, 2015.
11. G. Khalaf, and G. Shukur, *Choosing ridge parameter for regression problems*, Communications in Statistics - Theory and Methods, vol. 34, no. 5, pp. 1177–1182, 2005.
12. Y. Asar, A. Karaibrahimoğlu, and A. Genç, *Modified ridge regression parameters: A comparative Monte Carlo study*, Hacettepe Journal of Mathematics and Statistics, vol. 43, no. 5, pp. 827–841, 2014.
13. G. Muniz, and B.M.G. Kibria, *On some ridge regression estimators: An empirical comparisons*, Communications in Statistics - Simulation and Computation, vol. 38, no. 3, pp. 621–630, 2009.
14. B.M.G. Kibria, *Performance of some new ridge regression estimators*, Communications in Statistics - Simulation and Computation, vol. 32, no. 2, pp. 419–435, 2003.
15. R. Hamed, A.E.L. Hefnawy, and A. Farag, *Selection of the ridge parameter using mathematical programming*, Communications in Statistics - Simulation and Computation, vol. 42, no. 6, pp. 1409–1432, 2013.
16. M. Alkhamisi, G. Khalaf, and G. Shukur, *Some modifications for choosing ridge parameters*, Communications in Statistics - Theory and Methods, vol. 35, no. 11, pp. 2005–2020, 2006.
17. A.E. Hefnawy, and A. Farag, *A combined nonlinear programming model and Kibria method for choosing ridge parameter regression*, Communications in Statistics - Simulation and Computation, vol. 43, no. 6, pp. 1442–1470, 2014.
18. M.R. Özkale, *Predictive performance of linear regression models*, Statistical Papers, vol. 56, no. 2, pp. 531–567, 2015.
19. J.A. Sabourin, W. Valdar, and A.B. Nobel, *A permutation approach for selecting the penalty parameter in penalized model selection*, Biometrics, vol. 71, no. 4, pp. 1185–1194, 2015.
20. Z.Y. Algamal, and M.H. Lee, *Adjusted Adaptive LASSO in High-dimensional Poisson Regression Model*, Modern Applied Science, vol. 9, no. 4, pp. 170–177, 2015.
21. Y. Asar, *Some new methods to solve multicollinearity in logistic regression*, Communications in Statistics - Simulation and Computation, DOI: 10.1080/03610918.2015.1053925, 2015.
22. Y. Asar, *Some new methods to solve multicollinearity in logistic regression*, Communications in Statistics - Simulation and Computation, vol. 46, no. 4, pp. 2576–2586, 2015.
23. W.J. Hemmerle, and M. Boulanger Carey, *Some properties of generalized ridge estimators*, Communications in Statistics - Simulation and Computation, vol. 12, no. 3, pp. 239–253, 2007.
24. B.M.G. Kibria, *Some Liu and ridge-type estimators and their properties under the ill-conditioned Gaussian linear regression model*, Journal of Statistical Computation and Simulation, vol. 82, no. 1, pp. 1–17, 2012.
25. B.M.G. Kibria, K. Månsson, and G. Shukur, *Some ridge regression estimators for the zero-inflated Poisson model*, Journal of Applied Statistics, vol. 40, no. 4, pp. 721–735, 2013.
26. S. KaÇiranlar, and I. Dawoud, *On the performance of the Poisson and the negative binomial ridge predictors*, Communications in Statistics - Simulation and Computation, pp. 0–0, 2017.
27. K. Månsson, and G. Shukur, *A Poisson ridge regression estimator*, Economic Modelling, vol. 28, no. 4, pp. 1475–1481, 2011.
28. B.M.G. Kibria, K. Månsson, and G. Shukur, *A Simulation Study of Some Biasing Parameters for the Ridge Type Estimation of Poisson Regression*, Communications in Statistics - Simulation and Computation, vol. 44, no. 4, pp. 943–957, 2015.
29. P.S. Boonstra, B. Mukherjee, and J.M. Taylor, *A small-sample choice of the tuning parameter in ridge regression*, Statistica Sinica, vol. 25, no. 3, pp. 1185–1206, 2015.
30. S. Bhat, and R. Vidya, *A class of generalised ridge estimator*, Communications in Statistics - Simulation and Computation, pp. 00–00, 2016.
31. G.H. Golub, M. Heath, and G. Wahba, *Generalized cross-validation as a method for choosing a good ridge parameter*, Technometrics, vol. 21, no. 2, pp. 215–223, 1979.
32. Y. Asar, and A. Genç, *A New Two-Parameter Estimator for the Poisson Regression Model*, Iranian Journal of Science and Technology, Transactions A: Science, vol. 42, no. 2, pp. 793–803, 2017.
33. D.C. Montgomery, E.A. Peck, and G.G. Vining, *Introduction to linear regression analysis*, John Wiley & Sons, New York, 2015.
34. Algamal, Z. Y., & Asar, Y. *Liu-type estimator for the gamma regression model* , Communications in Statistics-Simulation and Computation, vol. 49, no . 8, p. 2035-2048, 2020.
35. 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.
36. 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.
37. Algamal, Z. Y., Lee, M. H., & Al-Fakih, A. M. *High-dimensional quantitative structure–activity relationship modeling of influenza neuraminidase a/PR/8/34 (H1N1) inhibitors based on a two-stage adaptive penalized rank regression* , Journal of Chemometrics, vol. 30 , no.2 m p. 50-57 , 2016.
38. Algamal, Z. Y., Lee, M. H., Al-Fakih, A. M., & Aziz, M. *High-dimensional QSAR classification model for anti-hepatitis C virus activity of thiourea derivatives based on the sparse logistic regression model with a bridge penalty* , Journal of Chemometrics, vol. 31 , p.6, p. e2889 , 2017.
39. Algamal, Z. Y., Qasim, M. K., Lee, M. H., & Ali, H. T. M. *High-dimensional QSAR/QSPR classification modeling based on improving pigeon optimization algorithm* , Chemometrics and Intelligent Laboratory Systems, vol. 206, p. 104170 , 2020.
40. Abonazel, M. R., Algamal, Z. Y., Awwad, F. A., & Taha, I. M. *A new two-parameter estimator for beta regression model: method, simulation, and application* , Frontiers in Applied Mathematics and Statistics, vol. 7,p. 780322 ,2022.
41. Algamal, Z. Y., & Abonazel, M. R. *Developing a Liu-type estimator in beta regression model* , Concurrency and Computation: Practice and Experience, vol.34 ,no. 5 , p. e6685.

42. 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.
43. Salih, A. M., Algamal, Z., & Khaleel, M. A. *A new ridge-type estimator for the gamma regression model*, Iraqi Journal for Computer Science and Mathematics, vol . 5 , no .1 , p. 85-98, 2023.
44. Alharthi, A. M., Kadir, D. H., Al-Fakih, A. M., Algamal, Z. Y., Al-Thanoon, N. A., & Qasim, M. K. *Quantitative structure-property relationship modelling for predicting retention indices of essential oils based on an improved horse herd optimization algorithm*, SAR and QSAR in Environmental Research, vol . 34 , no .10 , p. 831-846, 2023.
45. 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
46. 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
47. 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
48. 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
49. 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
50. 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 Trans. Math, vol. 21, p.403-414, 2022
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. Alharthi, A. M., Kadir, D. H., Al-Fakih, A. M., Algamal, Z. Y., Al-Thanoon, N. A., Qasim, M. K. *Improving golden jackel optimization algorithm: An application of chemical data classification*, Chemometrics and Intelligent Laboratory Systems, vol.250, 105149, 2024.
54. Qasim, O. S., Algamal, Z. Y. *A gray wolf algorithm for feature and parameter selection of support vector classification* , International Journal of Computing Science and Mathematics, vol. 13, 93-102, 2021.
55. Al-Fakih, A. M., Qasim, M. K., Algamal, Z. Y., Alharthi, A. M., Zainal-Abidin, M. H. *QSAR classification model for diverse series of antifungal agents based on binary coyote optimization algorithm* SAR and QSAR in Environmental Research, vol.34, no.4, p. 285-298, 2023.
56. Al-Fakih, A. M., Algamal, Z. Y., Qasim, M. K. *An improved opposition-based crow search algorithm for biodegradable material classification* SAR and QSAR in Environmental Research, vol.33, no.5, p. 403-415, 2022.
57. Hawa, N. S., Mustafa, M. Y., Kibria, B. G., & Algamal, Z. Y. *Bootstrap Liu-type estimator for Conway-Maxwell-Poisson regression model* Communications in Statistics-Simulation and Computation, p. 1-12, 2025.
58. Naziyah, A. A., & Algamal, Z. Y. *Jackknifed Liu-type estimator in Poisson regression model* Journal of the Iranian Statistical Society, vol.19, no.1, p. 21-37, 2020.
59. Al-Taweel, Y., & Algamal, Z. Y. *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.
60. Al-Fakih, A. M., Qasim, M. K., Algamal, Z. Y., Alharthi, A. M., & Zainal-Abidin, M. H. *QSAR classification model for diverse series of antifungal agents based on binary coyote optimization algorithm* Statistics, SAR and QSAR in Environmental Research, vol.34, no.4, p. 285-298, 2023.