STATISTICS, OPTIMIZATION AND INFORMATION COMPUTING
Stat., Optim. Inf. Comput., Vol. **15**, April 2026, pp 3129–3137.
Published online in International Academic Press (www.IAPress.org)

Bootstrap Liu Estimator for Almon Distributed Lag Model

Husam Waleed Yaseen [1], Salwa Salahuddin Qasim [2], Zakariya Yahya Algamal [1,*]

[1] *Department of Statistics and Informatics, University of Mosul, Mosul, Iraq*
[2] *College of Physical Education and Sports Sciences, University of Mosul, Mosul, Iraq*

Abstract The Almon distributed lag model is used to study how an explanatory variable affects a dependent variable spread out over a number of time periods, as opposed to an influence that happens instantly. Most of the time, Almon technique is used to estimate the parameters in the distributed lag model (DLM). Still, this estimator becomes very unstable if the explanatory variables and their delays are highly correlated. A new bootstrapped Liu shrinkage estimator is suggested in this research to deal with multicollinear challenges in the DLM. It was achieved by gradually narrowing the selection of the biasing parameters. According to the findings of the Monte Carlo study, the new methods lead to lesser MSE in all the cases compared to the standard methods. The use of the tested methods in real-world situations supports the assumption that they are better than the other ones.

Keywords Liu estimator, multicollinearity, Almon estimator, bootstra, distributed lag model

DOI: 10.19139/soic-2310-5070-3182

1. Introduction

Regression modeling forms the cornerstone of statistical analysis for predicting continuous outcomes based on one or more predictors. It quantifies relationships between variables while accounting for uncertainty through probabilistic frameworks.

In health sciences, regression models forecast disease incidence from risk factors; in environmental modeling, they predict water quality from pollutants. Economic forecasting employs time-series regression like ARIMA extensions. For your work in statistical modeling, integrating metaheuristics enhances feature selection in high-dimensional data.

Linear models suit normally distributed data with linear trends, while generalized linear models (GLMs) handle non-normal responses like counts (Poisson) or binaries (logistic). Nonlinear regression fits curved relationships, and regularized variants like Ridge or Lasso prevent overfitting by penalizing large coefficients.

Key assumptions include linearity, independence, homoscedasticity, and normality of residuals. Violations require transformations or robust methods like generalized least squares. Diagnostics involve plotting residuals, checking Durbin-Watson for autocorrelation, and using R-squared for fit, where higher values indicate better explanation of variance.

Distributed lag model (DLM) is a model that is commonly adopted to predict the current values of a response variable based on both the current values of a regressor (explanatory variable) and its lagged values. This model is often employed in economics and many other fields to look at actions that cause results to appear gradually and build a detailed picture of time-related relationships. DLM is particularly useful in fields such as environmental

*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)

epidemiology and economics, where understanding the impact of past events on current outcomes is crucial [1, 2, 3, 23, 24, 25, 26, 27].

The Almon distributed lag model imposes a polynomial structure on lag coefficients to address multicollinearity in finite distributed lag regressions. It reduces the parameter space by assuming lags follow a low-degree polynomial, enabling estimation in time series data where effects persist over time.

Consider the finite distributed lag model as

$$yt = \beta 0xt + \beta 1xt-1 + \cdots + \beta pxt-p + ut \tag{1}$$

Supposing certain conditions are met, OLS is the simplest approach to determine the distributed lags' parameters. Among the assumptions is that the greatest lag value is fixed, and both error and lagged explanators should have no restrictions. Nevertheless, when explanatory variables are correlated with their one-time shifts, it leads to big variations in the coefficients. If multicollinearity happens, the OLS can no longer work properly. Also, this method does not work properly when the number of observations is not well above the number of lags [28, 29, 30, 31, 32, 33, 34].

Imposing a shape on the lag distribution will reduce the effects of collinearity. Let us assume that the lag weights follow a smooth pattern that can be represented by a low degree polynomial. Almon [4] introduced this idea, and the resulting finite lag model is often called the Almon distributed lag, or a polynomial distributed lag. The Almon estimator is the best linear unbiased estimator (BLUE) for the DLM. Therefore, this estimator has a widespread usage in applied econometrics due to its ease of estimation. However, in order to avoid the problem of multicollinearity, Almon [4] proposed a desirable technique for estimation of the DLM, which has become a solution to the problems arising from multicollinearity. Almon [4] assumed that the coefficients i can be well approximated by a polynomial of degree r which is less than p. This polynomial is named the Almon distributed lag or polynomial distributed lag [35, 36, 37, 38, 39, 40, 41, 42]. The matrix notation of this prior information Eq.(2) is given as follow:

$$Y = X\beta + U \qquad \ldots(2)$$

in matrix notation where:

$$y = \begin{bmatrix} y_{p+1} \\ y_{p+2} \\ . \\ . \\ . \\ y_T \end{bmatrix}, \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ . \\ . \\ . \\ \beta_p \end{bmatrix}, X = \begin{bmatrix} x_{p+1} & x_p & \cdots & x_1 \\ x_{p+2} & x_{p+1} & \cdots & x_2 \\ \vdots & \vdots & \ddots & \vdots \\ x_T & x_{T-1} & \cdots & x_{T-p} \end{bmatrix}, u = \begin{bmatrix} u_{p+1} \\ u_{p+2} \\ . \\ . \\ . \\ u_T \end{bmatrix}$$

It is obvious that y is a (T-p) x 1 vector of observations on response variable, β is a (p+1) x 1 vector of unknown lag coefficient or lag weights, X is a (T-p) x (p+1) matrix containing values of explanatory variable and its p period lag values, u is (T-p) x 1 vector of normal distributed random errors with mean zero and a constant variance. The OLS of the parameter vector β in Eq.(2) be $\widehat{\beta}ols = \left(\acute{X}X\right)^{-1}\acute{X}y$. Application of the OLS method directly to the DLM, given in Eq.(1) for estimation purpose may have two serious problems: First, if the number of lags is large enough but the sample size is small, it may not possible to estimate parameters because of inadequate degrees of freedom to carry out the traditional test of significance. And, second, the successive (lag) values tend to be highly correlated causing the problem of multicollinearity which lead to imprecise estimation of parameters.

To tackle the issue of multicollinearity, Almon [4] proposed a technique for estimation of the DLM which is being widely used by practitioners. It is assumed that the coefficients β_i can be represented by a polynomial of degree r in i which is less than (p) (the lag length)

$$\beta_i = \theta_0 + \theta_1 i + \theta_2 i^2 + \cdots + \theta_r i^r ; i = 0, 1, 2, \ldots, p \; and \; p \geq r \geq 0 \; \ldots(3)$$

$$\beta = R\theta \qquad \ldots(4)$$

where

$$R = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ 1 & 1 & 1 & \dots & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & p & p^2 & \dots & p^r \end{pmatrix} \quad , \quad \theta = \begin{pmatrix} \theta_0 \\ \theta_1 \\ \vdots \\ \theta_r \end{pmatrix}$$

are (p+1)x(r+1) matrix and (r+1)x1 vector, respectively by substituting Eq.(4) in Eq.(2), we can get

$$y = XR\theta + u \qquad \dots(5)$$

$$= Z\theta + u \ , \ where \ Z = XR$$

Then, the OLS of θ is

$$\widehat{\theta} = \left(\acute{Z}Z\right)^{-1}\acute{Z}y \qquad \dots(6)$$

$$\widehat{\beta}_{OLS} = R\widehat{\theta} \qquad \dots(7)$$

which is the Almon estimator (AE) of β, it's the best linear unbiased estimator of β under the assumption that β=R θ .

2. Bootstrap Liu estimator in DLM

Several shrinkage estimators were studied to deal with multicollinearity issue in DLM [5, 6, 7, 8, 9, 64, 65, 66]. Accounting for the multicollinearity issue, Liu [10] suggested an estimator called the Liu estimator to combine the Stein estimator with the ridge estimator. Unlike ridge estimator, Liu's estimator is a simple linear function, so choosing the shrinkage parameter is much easier than choosing the ridge parameter.

Consequently, the Liu estimator in DLM is defined as

$$\hat{\theta}^{Liu}_{Almon} = (Z'Z + I)^{-1} + (Z'Z + dI)\hat{\theta}^{OLS}_{Almon} \tag{2}$$

$$\hat{\beta}^{Liu}_{Almon} = R\hat{\theta}^{Liu}_{Almon} \tag{3}$$

where d is a shrinkage parameter, $0 < d < 1$.

The matrix mean squared error (MMSE) of $\hat{\beta}^{Liu}_{Almon}$ is defined as follows:

$$\begin{aligned} \text{MMSE}\ (\hat{\beta}^{Liu}_{Almon}) &= E(\hat{\beta}^{Liu}_{Almon} - \boldsymbol{\beta})(\hat{\beta}^{Liu}_{Almon} - \boldsymbol{\beta})^T \\ &= Var(\hat{\beta}^{Liu}_{Almon}) + \left[bias(\hat{\beta}^{Liu}_{Almon})\right]\left[bias(\hat{\beta}^{Liu}_{Almon})\right]^T , \end{aligned} \tag{4}$$

where

$$Var(\hat{\beta}^{Liu}_{Almon}) = \mathbf{V}_d\mathbf{B}^{-1}\mathbf{V}_d^T, \tag{5}$$

and

$$bias(\hat{\beta}^{Liu}_{Almon}) = [\mathbf{V}_d - \mathbf{I}]\ \boldsymbol{\beta}, \tag{6}$$

where $\mathbf{B} = Z'Z$ and $\mathbf{V}_d = (\mathbf{B} + \mathbf{I})^{-1}(\mathbf{B} + d\,\mathbf{I})$. The MSE of the estimator $\hat{\beta}^{Liu}_{Almon}$ can be defined as

$$\begin{aligned} \text{MSE}\ (\hat{\beta}^{Liu}_{Almon}) &= tr\left[\mathbf{V}_d\mathbf{B}^{-1}\mathbf{V}_d^T\right] + [\mathbf{V}_d - \mathbf{I}]\ \boldsymbol{\beta}\boldsymbol{\beta}^T\left[\mathbf{V}_d - \mathbf{I}\right]^T \\ &= \sum_{j=1}^{p}\frac{(\lambda_j+d)^2}{\lambda_j(\lambda_j+1)^2} + (d-1)^2\sum_{j=1}^{p}\frac{\alpha_j^2}{(\lambda_j+1)^2} , \end{aligned} \tag{7}$$

where α_j is defined as the jth element of $\gamma\hat{\boldsymbol{\beta}}_{OLS}$and γ is the eigenvector of the $Z'Z$. To find the optimal value of d, the first derivative of Eq. (7) with respect to d and setting the resulting to zero and solving for d, the optimal value is obtained as [11]

$$d_{optimal} = \frac{(\alpha^2 - 1)}{(1/\lambda) + \alpha^2}. \tag{8}$$

According to Eq. (14), when $\alpha_j^2 < 1$ the $d_{optimal}$ becomes negative and becomes positive when $\alpha_j^2 > 1$. To guarantee that $d_{optimal}$be between 0 and 1, the following methods have been proposed to estimate the $d_{optimal}$ [43, 44, 45, 46, 47, 48, 49, 50, 51, 52]:

$d_1 = \max\left(0, \frac{(\hat{\alpha}_{\max}^2 - 1)}{(1/\hat{\lambda}_{\max}) + \hat{\alpha}_{\max}^2}\right)$, $d_2 = \max\left(0, median\left(\frac{(\hat{\alpha}_j^2 - 1)}{(1/\hat{\lambda}_j) + \hat{\alpha}_j^2}\right)\right)$, and $d_3 = \max\left(0, \frac{1}{p}\sum_{j=1}^{p}\frac{(\hat{\alpha}_j^2 - 1)}{(1/\hat{\lambda}_j) + \hat{\alpha}_j^2}\right)$.

Bootstrap is one more resampling method used for studying data. A bootstrap analysis means finding the sampling distribution of a statistic by re-sampling the data with replacement [53, 54, 55, 56]. Even so, by adopting the assumption that observations are all independent and identically distributed, bootstrap may be off the mark if the data contains dependencies [14, 57, 58, 59, 60, 61].

By sampling pairs of observations instead of independent observations, bootstrap deals with the issue [15, 62, 63]. To find the parameters of the pairs, two observations from the total data set are arbitrarily chosen and a sample is taken from those two [16]. After that, the statistic is calculated for all pairs and these results are used to create a big sample to find the distribution of the statistic. Pair bootstrap ensures that data plays a key role and gives accurate estimates of sample distribution in the event of correlation or related observations [17]. Time series is a typical example where the observations have values connected with each other. The authors Chaubey, Khurana and Chandra [18] derived the bootstrapped method for confidence interval of ridge parameter in linear regression modeling. Perveen and Suhail [19] studied the bootstrapped Liu estimator in Poisson regression model. Namely, in 2024, Dar and Chand [20] made use of the bootstrap quantile ridge estimator for regression.

In the paired bootstrap, the pairs bootstrap samples are produced by sampling n observations from $(y_1, x_1, x_2, \ldots, x_p), (y_2, x_1, x_2, \ldots, x_p), \ldots, (y_n, x_1, x_2, \ldots, x_p)$ with replacement and having equal selection probability. Then, the bootstrap sample is $(y_1^{boot}, x_1^{boot}, x_2^{boot}, \ldots, x_p^{boot}), (y_2^{boot}, x_1^{boot}, x_2^{boot}, \ldots, x_p^{boot}), \ldots, (y_n^{boot}, x_1^{boot}, x_2^{boot}, \ldots, x_p^{boot})$ and the bootstrap The bootstrap estimate of Eq.(6) and Eq.(7)

$$\widehat{\theta}_{boot} = \left(\acute{Z}Z\right)^{-1}\acute{Z}y \tag{15}$$

$$\widehat{\beta}_{OLS}^{boot} = R\widehat{\theta}_{boot} \tag{9}$$

To alleviate multicollinearity issue in DLM, the proposed bootstrapped Liu shrinkage parameter is then obtained by determining the bootstrapped d, d^{boot}. For the number of bootstrapped samples, B, d^{boot} can be defined as: $d_1^{boot} = \frac{1}{B}\sum_{i=1}^{B} d_{1,i}$, $d_2^{boot} = \frac{1}{B}\sum_{i=1}^{B} d_{2,i}$, and $d_3^{boot} = \frac{1}{B}\sum_{i=1}^{B} d_{3,i}$.

3. Simulation results

To evaluate the performance of the estimators used in DLM, we therefore undertook a simulation study for DLM. The dependent variable and the explanatory variables are produced based on the following study [5, 7, 9].

$$yt = \beta 0xt + \beta 1xt - 1 + \cdots + \beta pxt - p + u_t \tag{10}$$

$$x_1 = v_1 \tag{11}$$

$$x_t = \pi x_{t-1} + \sqrt{(1 - \pi^2)}\, v_t \quad for\ t \geq 2 \tag{12}$$

where's π denotes the correlation between the regressors and its values are chosen as follows: 0.90, 0.95, and 0.99. The MSE is minimized subject to the constraint $B'B = 1$. u_t and v_t aregen erated such that $u_t \sim N\left(0, \sigma^2\right)$ and $v_t \sim N(0, 1)$, respectively. The choices of σ^2 are 1, 10 and 15. An observation of T-p =60 and 100 with the lag length of 8 and 16 was evaluated. The experiment is repeated 1000 times. The mean squared error (MSE) is obtained as follows:

$$MSE\left(\hat{B}\right) = \frac{1}{1000}\sum_{j=1}^{1000}\left(\hat{B}_{ij} - B_i\right)'\left(\hat{B}_{ij} - B_i\right) \quad (13)$$

where $\hat{B}_{ij}$ denotes the estimate of the ith parameter in jth replication and B_i is the true parameter values. The MSE values are presented in Tables 1-6. The best obtained results are in bold font.

For all n and π, the MSE decreases progressively from OLS to the biased estimators d_1, d_2, and d_3 and further to their enhanced bootstrap versions d_1^{boot}, d_2^{boot}, and d_3^{boot}. Further, OLS estimator consistently has the highest MSE, especially under severe multicollinearity, π=0.99. Moreover, the Liu estimator with d_1, d_2, and $d_3$has intermediate MSE reduction, with d_3 outperforming d_2 and d_1. On the other hand, the Liu estimator with d_1^{boot}, d_2^{boot}, and d_3^{boot}achieved the lowest MSE, suggesting improved performance over their predecessors d_1, d_2, and d_3. Liu estimator with d_3^{boot} consistently yields the smallest MSE across all scenarios, indicating it balances bias-variance trade-offs most effectively.

Increasing the sample size from 60 to 100 reduces the MSE for all estimators across all correlation levels. This aligns with statistical theory that larger samples provide more information, leading to more precise estimates. Notably, the reduction in MSE is more pronounced for the biased estimators, especially the Liu estimator with d_3^{boot}, highlighting their improved efficiency with larger data.

Further, it is also seen that for fixed value of σ^2 the estimated MSE values of the estimators increase with the increase of the length of lag. Furthermore, the MSE values of the use estimator increases with the increase of σ^2 and the length of lag.

Table 1 shows MSE for OLS, Liu estimators (d1,d2,d3), and bootstrapped versions (dboot1-3) at n=60 and 100. OLS has highest MSE (e.g., 0.952 at n=60, ρ=0.99ρ=0.99); bootstrapped Liu (dboot3) lowest (0.435). Larger n reduces MSE across all, with biased estimators gaining most efficiency under high $\rho\rho$. dboot3 bolded as superior, confirming bootstrap enhances Liu's bias-variance balance.

Further, Table 2 shows Longer lags inflate MSE versus Table 1 (e.g., OLS 0.977 at n=60, ρ=0.99ρ=0.99). Pattern persists: OLS worst, dboot3 best (0.462). MSE rises with lag length due to intensified multicollinearity among more lags; sample increase to n=100 yields sharper relative gains for dboot estimators.

According to Table 3, higher variance elevates all MSE (e.g., OLS 0.989 at n=60, ρ=0.99ρ=0.99); dboot3 remains optimal (0.472). Bootstrap Liu outperforms non-bootstrap counterparts, especially severe multicollinearity, by stabilizing shrinkage parameter d via resampling.

Finally, from Table 4, high variance, long lags (OLS 0.997 at n=60, ρ=0.99ρ=0.99). dboot3 excels (0.483), with consistent hierarchy: OLS ¿ d1-d3 ¿ dboot1-3. Overall, simulations validate proposed bootstrapped Liu estimators minimize MSE, ideal for your DLM work with correlated lags in economic/epidemiological time series.

Table 1. MSE values when σ^2=1 and lag length=8

n	π	$\widehat{\beta}_{OLS}$	d_1	d_2	d_3	d_1^{boot}	d_2^{boot}	d_3^{boot}
60	*0.90*	*0.25*	*0.223*	*0.167*	*0.136*	*0.193*	*0.137*	***0.106***
	0.95	*0.765*	*0.365*	*0.309*	*0.278*	*0.335*	*0.279*	***0.248***
	0.99	*0.952*	*0.552*	*0.497*	*0.465*	*0.522*	*0.467*	***0.435***
100	*0.90*	*0.234*	*0.209*	*0.153*	*0.127*	*0.179*	*0.123*	***0.097***
	0.95	*0.577*	*0.307*	*0.251*	*0.222*	*0.277*	*0.221*	***0.191***
	0.99	*0.773*	*0.433*	*0.377*	*0.347*	*0.403*	*0.347*	***0.317***

Table 2. MSE values when $\sigma^2 = 1$ and lag length=16

n	π	$\widehat{\beta}_{OLS}$	d_1	d_2	d_3	d_1^{boot}	d_2^{boot}	d_3^{boot}
60	*0.90*	*0.275*	*0.248*	*0.192*	*0.161*	*0.218*	*0.162*	***0.131***
	0.95	*0.79*	*0.39*	*0.334*	*0.303*	*0.36*	*0.304*	***0.273***
	0.99	*0.977*	*0.577*	*0.522*	*0.491*	*0.547*	*0.492*	***0.462***
100	*0.90*	*0.259*	*0.234*	*0.178*	*0.152*	*0.204*	*0.148*	***0.122***
	0.95	*0.602*	*0.332*	*0.276*	*0.247*	*0.302*	*0.246*	***0.216***
	0.99	*0.798*	*0.458*	*0.402*	*0.372*	*0.428*	*0.372*	***0.342***

Table 3. MSE values when $\sigma^2 = 10$ and lag length=8

n	π	$\widehat{\beta}_{OLS}$	d_1	d_2	d_3	d_1^{boot}	d_2^{boot}	d_3^{boot}
60	0.90	0.287	0.26	0.204	0.173	0.23	0.174	**0.143**
	0.95	0.802	0.402	0.346	0.315	0.372	0.316	**0.285**
	0.99	0.989	0.589	0.534	0.502	0.559	0.504	**0.472**
100	0.90	0.271	0.246	0.19	0.164	0.216	0.16	**0.134**
	0.95	0.614	0.344	0.288	0.259	0.314	0.258	**0.228**
	0.99	0.81	0.47	0.414	0.384	0.44	0.384	**0.354**

Table 4. MSE values when $\sigma^2 = 10$ and lag length=16

n	π	$\widehat{\beta}_{OLS}$	d_1	d_2	d_3	d_1^{boot}	d_2^{boot}	d_3^{boot}
60	0.90	0.295	0.268	0.212	0.181	0.238	0.182	**0.151**
	0.95	0.81	0.41	0.354	0.323	0.38	0.324	**0.293**
	0.99	0.997	0.597	0.542	0.51	0.567	0.512	**0.483**
100	0.90	0.279	0.254	0.198	0.172	0.224	0.168	**0.142**
	0.95	0.622	0.352	0.296	0.267	0.322	0.266	**0.236**
	0.99	0.818	0.478	0.422	0.392	0.448	0.392	**0.362**

Table 5. MSE values when $\sigma^2 = 15$ and lag length=8

n	π	$\widehat{\beta}_{OLS}$	d_1	d_2	d_3	d_1^{boot}	d_2^{boot}	d_3^{boot}
60	0.90	0.391	0.364	0.308	0.277	0.334	0.278	0.247
	0.95	0.906	0.506	0.45	0.419	0.476	0.42	0.389
	0.99	1.093	0.693	0.638	0.606	0.663	0.608	0.576
100	0.90	0.375	0.35	0.294	0.268	0.32	0.264	0.238
	0.95	0.718	0.448	0.392	0.363	0.418	0.362	0.332
	0.99	0.914	0.574	0.518	0.488	0.544	0.488	0.458

Table 6. MSE values when $\sigma^2 = 15$ and lag length=16

n	π	$\widehat{\beta}_{OLS}$	d_1	d_2	d_3	d_1^{boot}	d_2^{boot}	d_3^{boot}
60	0.90	0.399	0.372	0.316	0.285	0.342	0.286	0.255
	0.95	0.914	0.514	0.458	0.427	0.484	0.428	0.397
	0.99	1.101	0.701	0.646	0.614	0.671	0.616	0.587
100	0.90	0.383	0.358	0.302	0.276	0.328	0.272	0.246
	0.95	0.726	0.456	0.4	0.371	0.426	0.37	0.34
	0.99	0.922	0.582	0.526	0.496	0.552	0.496	0.466

4. Applications

The Almon dataset is used to show the performances of each of the estimators as proposed by Almon [4]. The data in question covered the period from 1953 to 1967, in which quarterly expenditures were the response and appropriations the independent variable. In line with Gültay and Kaçıranlar [21] and Güler, Gültay and Kaçiranlar [22] considerations on VECM, the lag length is set 8 and the lag weight estimated to be of second order polynomial. The eigenvalues of $Z'Z$ has been computed to be 2.388929e±13, 1.628323e±9, 4.312742e±7, 4.782067e±0 respectively. Source: Computation The condition index of condition index is established to be a 2,235,083 which means that there is a high level of multicollinearity between the explanatory variable and its lags.

Table 7 contains the MSE values and estimated Almon coefficients for the OLS and the biased estimators d_1, d_2, and d_3 and further to their enhanced bootstrap versions d_1^{boot}, d_2^{boot}, and d_3^{boot}. Table 5 makes it abundantly evident that OLS estimator extremely has high MSE with 3397.92, indicating severe instability due to multicollinearity. While the Liu estimator with d_1, d_2, d_3, d_1^{boot}, d_2^{boot}, and d_3^{boot} has dramatically lower MSE (all <200), demonstrating their effectiveness in mitigating multicollinearity-induced variance inflation. On the other words, $d_1 \to d_2 \to d_3$: MSE decreases slightly (194.34 $\to$ 190.88 $\to$ 188.63), suggesting incremental improvements from adjustments to biasing parameters. On the other hand, Liu estimator with $d_1^{boot} \to d_2^{boot} \to d_3^{boot}$has further MSE reduction (179.57 $\to$ 171.48 $\to$ 162.31), likely due to enhanced techniques of bootstrap. The Liu estimator with d_3^{boot} achieves the lowest MSE outperforming OLS by 95.2%.

Further, the OLS MSE=3397.92, unstable from variance inflation. Non-bootstrap Liu reduces to 194.34 (d1), 190.88 (d2), 188.63 (d3). Bootstrapped versions further improve: dboot1=179.57, dboot2=171.48, dboot3=162.31 (95.2% MSE reduction vs. OLS). dboot3 optimal, leveraging paired bootstrap for robust d selection in dependent time series data. In addition, shrunk magnitudes (e.g., x1 from 0.0962 to 0.0057) stabilize fits while preserving hump-shaped lag pattern realistic for delayed spending responses. For your research, dboot3 enhances precision in multicollinear DLMs like environmental forecasting.

Table 7. The estimated coefficients and MSE values for the used estimators

coefficients	$\widehat{\beta}_{OLS}$	d_1	d_2	d_3	d_1^{boot}	d_2^{boot}	d_3^{boot}
x1	*0.0962*	*0.0105*	*0.0097*	*0.0088*	*0.0071*	*0.0068*	*0.0057*
x2	*0.0320*	*0.02607*	*0.02521*	*0.02488*	*0.02387*	*0.02371*	*0.02344*
x3	*-0.0052*	*-0.00439*	*-0.00401*	*-0.00384*	*-0.00359*	*-0.00342*	*-0.00328*
MSE	*3397.92*	*194.34*	*190.88*	*188.63*	*179.57*	*171.48*	*162.31*

5. Conclusion

To solve the multicollinearity issue in DLM, the Liu estimator is introduced in this paper. For the first time, a new estimator designed with bootstrapping was introduced in the DLM to fix the multicollinearity issue. This was accomplished by using a simple way to select the biasing parameter d. The strategy was assessed for its performance as multiple affecting factors were present in the data. Simulation results indicate that d_1^{boot}, d_2^{boot}, and d_3^{boot}, which are the proposed methods, work better and have smaller MSE compared to the standard methods used. In real data testing, it was found that the estimators called d_1^{boot}, d_2^{boot}, and d_3^{boot}worked better than the other methods did.

REFERENCES

1. M. Belloumi, *The relationship between trade, FDI and economic growth in Tunisia: An application of the autoregressive distributed lag model*, Economic Systems, vol. 38, no. 2, pp. 269–287, 2014.
2. H. Erdal, G. Erdal, and K. Esengun, *An analysis of production and price relationship for potato in Turkey: a distributed lag model application*, Bulgarian Journal of Agricultural Science, vol. 15, no. 3, pp. 243–250, 2009.
3. A. M. Rushworth, et al., *Distributed lag models for hydrological data*, Biometrics, vol. 69, no. 2, pp. 537–544, 2013.
4. S. Almon, *The distributed lag between capital appropriations and expenditures*, Econometrica: Journal of the Econometric Society, pp. 178–196, 1965.
5. A. F. Lukman and G. B. M. Kibria, *Almon-KL estimator for the distributed lag model*, Arab Journal of Basic and Applied Sciences, vol. 28, no. 1, pp. 406–412, 2021.
6. A. Majid, et al., *Robust estimation of the distributed lag model with multicollinearity and outliers*, Communications in Statistics - Simulation and Computation, vol. 53, no. 8, pp. 3933–3947, 2022.
7. N. Özbay, *Two-Parameter Ridge Estimation for the Coefficients of Almon Distributed Lag Model*, Iranian Journal of Science and Technology, Transactions A: Science, vol. 43, no. 4, pp. 1819–1828, 2018.
8. A. Erkoç, et al., *The beta Liu-type estimator: simulation and application*, Hacettepe Journal of Mathematics and Statistics, vol. 52, no. 3, pp. 828–840, 2023.
9. N. Özbay and S. Kaçıranlar, *The Almon two parameter estimator for the distributed lag models*, Journal of Statistical Computation and Simulation, vol. 87, no. 4, pp. 834–843, 2017.
10. 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.
11. K. Månsson, *Developing a Liu estimator for the negative binomial regression model: method and application*, Journal of Statistical Computation and Simulation, vol. 83, no. 9, pp. 1773–1780, 2013.
12. K. Månsson, et al., *Improved Liu estimators for the Poisson regression model*, International Journal of Statistics and Probability, vol. 1, no. 1, pp. 1–6, 2012.
13. G. J. Babu, *Resampling methods for model fitting and model selection*, Journal of Biopharmaceutical Statistics, vol. 21, no. 6, pp. 1177–1186, 2011.

14. O. Baser, W. H. Crown, and C. Pollicino, *Guidelines for selecting among different types of bootstraps*, Current Medical Research and Opinion, vol. 22, no. 4, pp. 799–808, 2006.
15. R. J. Tibshirani and B. Efron, *An introduction to the bootstrap*, Monographs on Statistics and Applied Probability, vol. 57, no. 1, pp. 1–436, 1993.
16. F. Rabbi, et al., *Model selection in linear regression using paired bootstrap*, Communications in Statistics - Theory and Methods, vol. 50, no. 7, pp. 1629–1639, 2021.
17. N. S. Hawa, et al., *Bootstrap Liu-type estimator for Conway-Maxwell-Poisson regression model*, Communications in Statistics - Simulation and Computation, pp. 1–12, 2023.
18. Y. P. Chaubey, M. Khurana, and S. Chandra, *Confidence intervals based on resampling methods using Ridge estimator in linear regression model*, New Trends in Mathematical Sciences, vol. 6, no. 4, 2018.
19. I. Perveen and M. Suhail, *Bootstrap Liu estimators for Poisson regression model*, Communications in Statistics - Simulation and Computation, vol. 52, no. 7, pp. 2811–2821, 2021.
20. I. S. Dar and S. Chand, *Bootstrap-quantile ridge estimator for linear regression with applications*, PLoS One, vol. 19, no. 4, p. e0302221, 2024.
21. B. Gültay and S. Kaçıranlar, *Mean Square Error Comparisons of the Alternative Estimators For The Distributed Lag Models*, Hacettepe Journal of Mathematics and Statistics, vol. 44, no. 35, pp. 1–1, 2014.
22. H. Güler, B. Gültay, and S. Kaçiranlar, *Comparisons of the alternative biased estimators for the distributed lag models*, Communications in Statistics - Simulation and Computation, vol. 46, no. 4, pp. 3306–3318, 2016.
23. 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.
24. 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.
25. Kahya, M. A., Altamir, S. A., & Algamal, Z. Y. *Improving whale optimization algorithm for feature selection with a time-varying transfer function* , Numerical Algebra, Control and Optimization, vol. 11, no. 1 , p. 87-98 , 2020.
26. 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.
27. 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.
28. 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.
29. 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.
30. Ismael, O. M., Qasim, O. S., & Algamal, Z. Y. *Improving Harris hawks optimization algorithm for hyperparameters estimation and feature selection in v-support vector regression based on opposition-based learning* , Journal of Chemometrics, vol 34 , no. 11, e3311, 2020.
31. 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.
32. 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.
33. 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.
34. 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.
35. 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.
36. Mahmood, S. W., Basheer, G. T., & Algamal, Z. Y. *Quantitative Structure–Activity Relationship Modeling Based on Improving Kernel Ridge Regression* , Journal of Chemometrics, vol. 39, no. 5, p. e70027, 2025.
37. Mahmood, S. W., Basheer, G. T., & Algamal, Z. Y. *Improving kernel ridge regression for medical data classification based on meta-heuristic algorithms* , Kuwait Journal of Science, vol. 52, no. 3 , p. 100408 , 2025
38. 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
39. 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
40. 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
41. 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
42. 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
43. 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

44. Ewees, A. A., Algamal, Z. Y., Abualigah, L., Al-Qaness, M. A., Yousri, D., Ghoniem, R. M., & Abd Elaziz, M. *A cox proportional-hazards model based on an improved aquila optimizer with whale optimization algorithm operators*, Mathematics, vol. 10, no.(8), p.1273, 2022
45. 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
46. 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
47. Almishlih1, Zaynab Ayham, Qasim, Omar Saber, Algamal, Zakariya Yahya *Binary Arithmetic Optimization Algorithm Using a New Transfer Function for Fusion Modeling*, Fusion: Practice and Applications, vol. 18, no.2, p. 157-168, 2025
48. Kahya, M. A., Altamir, S. A., & Algamal, Z. Y. *Improving whale optimization algorithm for feature selection with a time-varying transfer function* , Numerical Algebra, Control and Optimization, vol. 11, no. 1 , p. 87-98 , 2020.
49. 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.
50. 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.
51. 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.
52. Ewees, A. A., Al-Qaness, M. A., Abualigah, L., Algamal, Z. Y., Oliva, D., Yousri, D., & Elaziz, M. A. *Enhanced feature selection technique using slime mould algorithm: A case study on chemical data* , Neural Computing and Applications, vol. 35, 3307-3324, 2023.
53. 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.
54. Algamal, Z. Y. *Variable selection in count data regression model based on firefly algorithm* , Statistics, Optimization & Information Computing, vol. 7, 520-529, 2019.
55. Kahya, M. A., Altamir, S. A., & Algamal, Z. Y. *Improving firefly algorithm-based logistic regression for feature selection*, Journal of Interdisciplinary Mathematics, vol. 22, 1577-1581, 2019.
56. Yousif, H. M., & Algamal, Z. Y. *Bandwidth Selection in Geographically Weighted Poisson Regression Model Using Firefly Optimization Algorithm with Application to Cancer Rate Data*, European Journal of Statistics, vol. 5, 10, 2025.
57. AL-Taie, F. A. Y., Qasim, O. S., & Algamal, Z. Y. *Improving kernel semi-parametric regression model based on a bat optimization algorithm*, AIP Conference Proceedings, vol. 3036, p. 040003, 2024.
58. Andu, Y., Lee, M. H., & Algamal, Z. Y. *Generalized dynamic principal component for monthly nonstationary stock market price in technology sector*, Journal of Physics: Conference Series, vol. 1132, p. 012076, 2018.
59. Andu, Y., Lee, M. H., & Algamal, Z. Y. *Generalized dynamic principal component for monthly nonstationary stock market price in technology sector*, Journal of Physics: Conference Series, vol. 1132, p. 012076, 2018.
60. Almishlih, Z. A., Qasim, O. S., & Algamal, Z. Y. *Design and evaluation of a new tent-shaped transfer function using the Polar Lights Optimizer algorithm for feature selection*, Informatyka, Automatyka, Pomiary w Gospodarce i Ochronie Środowiska, vol. 15, p. 27-31, 2025.
61. Al-Thanoon, N. A., Qasim, O. S., & Algamal, Z. Y. *Improving the binary tree growth algorithm with fuzzy mutual information for feature selection*, AIP Conference Proceedings, vol. 3318, p. 030008, 2025.
62. 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.
63. 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.
64. 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.
65. 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.
66. 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.