A New Two-parameter Estimator for the Gamma Regression Model

Yasin Asar ¹, Zakariya Yahya Algamal ²,*

¹Department of Mathematics and Computer Sciences, Necmettin Erbakan University, Turkey
²Department of Statistics and Informatics, University of Mosul, Iraq

Abstract In this paper, we propose a new two-parameter biased estimator in gamma regression models when there is collinearity among the regressors. We investigate the mean squared error properties of the newly proposed estimator. Moreover, we provide some theorems to compare the new estimators to the existing ones. We conduct a Monte Carlo simulation study to compare the estimators under different designs of collinearity in the sense of mean squared error. Moreover, we provide a real data application to show the usefulness of the new estimator. The simulations and real data results show that the proposed estimator beats other competitor estimators.

Keywords Gamma regression model, two-parameter estimator, collinearity, Monte Carlo simulation

AMS 2010 subject classifications 62J05, 62J07

DOI: 10.19139/soic-2310-5070-822

1. Introduction

Gamma regression model is one of the widely applied models for studying several real data problems, such as medical science, health-care economics, and automobile insurance claims,(see [12], [13] and [23]). The gamma regression model is used when the values of the response variable under the study is positively skewed following gamma distribution ([1], [35]).

As in linear regression model, in gamma regression model, it is assumed that there is no correlation among the explanatory variables. In practice, however, this assumption often not holds, which leads to the problem of multicollinearity. In the presence of multicollinearity, the maximum likelihood (ML) estimator of the gamma regression coefficients are usually become unstable with high variance, and therefore low statistical significance [13] and [23].

Several remedial methods have been proposed to overcome the problem of multicollinearity. The ridge regression method was proposed by [15] and has been consistently demonstrated to be an attractive and alternative to the ML estimation method. The ridge estimator was considered in the generalized linear models (GLM) by [33]. Moreover, [33] and [32] considered the ridge estimator in logistic regression. [25] and [26] were also adapted the ridge estimator in Poisson regression and negative binomial regression models, respectively. The well-known Liu estimator [20] has been recently generalized to GLM and application of gamma distributed response variable was demonstrated by [19]. We also refer to the following papers for Liu regression: [36], [27], [34]. Another solution to the collinearity problem is the two-parameter estimator [28]. This estimator was also well studied in the literature and generalized to some models which are members of GLMs ([16], [11]). On the other hand, several estimators have been proposed for dealing with the issue of multicollinearity in gamma regression model [24, 29, 8, 21, 9, 6, 7, 3].
The organization of the paper is as follows: In Section 2, we propose the gamma two-parameter estimator and investigate its mean squared error (MSE) properties. In Section 3, we conduct a Monte Carlo simulation study to evaluate the MSE performances of used estimators in the presence of multicollinearity. Moreover, a real data application is provided to illustrate the benefits of the new estimation technique in Section 4. Finally, a conclusion is given in Section 5.

2. Theory and Method

Let \( Y_1, Y_2, \ldots, Y_n \) be independent random variables and \( y_1, y_2, \ldots, y_n \) be the corresponding observations from the gamma distribution having the following probability density function

\[
f(y_i) = \frac{y_i^{v-1}e^{-y_i/\tau}}{\Gamma(v)} \quad (\tau, y_i \geq 0)
\]

where \( v \) is the non-negative shape parameter and \( \tau \) is the scale parameter such that \( E(Y_i) = v\tau = \theta_i \) which is also known as the canonical parameter and \( \text{Var}(Y_i) = v\tau^2 = 1/(\theta_i^2) \). \( \theta_i = \exp(x_i^T \beta) \) where \( x_i = (x_{i1}, x_{i2}, \ldots, x_{ip})^T \), \( i = 1, 2, \ldots, n \) and \( j = 1, 2, \ldots, p \) where \( n \) is the sample size and \( p \) is the number of explanatory variables \( (n > p) \). Generally, the maximum likelihood estimation is used to obtain the parameters. To do so, the following log-likelihood function should be maximized with respect to \( \beta \)

\[
\ell(\beta) = \sum_{i=1}^{n} [(v - 1) \log(y) - y/\tau - v \log(\tau) - \log(\Gamma(v))].
\]

Since the obtained equations are nonlinear in \( \beta \), we should use some iterative methods to get the solutions. Therefore, by using the Fisher Scoring method, the following iterations can be defined

\[
\beta^{t+1} = \beta^t - \left\{ \frac{\partial \ell(\beta)}{\partial \beta} \right\}_{\beta = \beta^t}^{-1} \left\{ E[H_l(\beta)]_{\beta = \beta^t} \right\}
\]

where \( H_l(\beta) = -\frac{1}{\phi}X^TWX \) is the Hessian matrix such that \( \phi = 1/v \) is the dispersion parameter and

\[
\frac{\partial \ell(\beta)}{\partial \beta} = \phi \sum_{i=1}^{n} \left[ \frac{y_i}{(x_i^T \beta)^2} - 1 \right] x_i.
\]

Therefore, Equation (3) can be written as

\[
\beta^{t+1} = \beta^t - \left\{ \left( X^T \tilde{W}X \right)^{-1} X^T \tilde{W}z \right\}_{\beta = \beta^t}
\]

where \( \tilde{W} = \text{diag}(\theta_i^2) \) and the ith element of the vector \( \tilde{z} \) becomes \( \tilde{z}_i = \tilde{\theta}_i + \frac{y_i - \tilde{\theta}_i}{\tilde{\theta}_i} \). This iterative process continues until the successive estimates converges to, say, \( \hat{\beta}_{\text{MLE}} \), then we obtain \( \hat{\beta}_{\text{MLE}} = \left( X^T \tilde{W}X \right)^{-1} X^T \tilde{W} \tilde{z} \) where \( \tilde{W} \) and \( \tilde{z} \) are computed at the final iteration.

It is well-known that the covariance matrix of \( \hat{\beta}_{\text{MLE}} \), \( \text{Cov}(\hat{\beta}_{\text{MLE}}) = \phi \left( X^T \tilde{W}X \right)^{-1} \), may be ill-conditioned so that the variance of the regression coefficients is inflated (see [33], [22]). The mean squared
error (MSE) of MLE is given by

\[
\text{MSE}(\hat{\beta}_{\text{MLE}}) = E\left( (\hat{\beta}_{\text{MLE}} - \beta) ^\top (\hat{\beta}_{\text{MLE}} - \beta) \right) = \text{tr} \left[ \phi \left( X^\top W X \right)^{-1} \right]
\]

\[
= \phi \sum_{j=1}^{p} \frac{1}{\lambda_j}
\]

(6)

where \( \lambda_j \) is the jth eigenvalue of the matrix \( C = X^\top W X \) and \( \text{tr}(\cdot) \) is the trace of a matrix. Moreover, the eigenvalue decomposition of the matrix \( C \) is also considered as follows: \( C = Q \Lambda Q^\top \) such that \( Q \) is the orthogonal matrix consisting of the eigenvectors corresponding to the eigenvalues of \( C \) such that \( \Lambda = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_p) \). It is easy to see that if one or some of the eigenvalues are close to zero, then the MSE of MLE becomes inflated and thus the regression coefficients are affected negatively from this situation.

2.1. Gamma Ridge Estimator

The well-known ridge estimator was proposed by [33] in the generalized linear models. [2] adapted the ridge estimator to the gamma regression models. The author defined the gamma ridge estimator (GRE) as follows

\[
\hat{\beta}_k = (C + kI)^{-1} X^\top W \hat{\varepsilon}
\]

\[
b_{\text{GRE}} = C_k^{-1} C \hat{\beta}_{\text{MLE}}
\]

(7)

where \( k > 0 \), \( C = X^\top W X \) and \( C_k = (C + kI) \). The covariance matrix and bias vector of GRE can be obtained respectively by

\[
\text{Cov}(\hat{\beta}_k) = \phi C_k^{-1} C \text{Cov}(\beta) C_k^{-1}
\]

\[
b_{\text{GRE}} = \text{bias}(\hat{\beta}_k) = -k C_k \beta
\]

(8)

(9)

Thus, matrix MSE (MMSE) and MSE of GRE are obtained as

\[
\text{MMSE}(\hat{\beta}_k) = \text{Cov}(\hat{\beta}_k) + b_{\text{GRE}} b_{\text{GRE}}^\top
\]

\[
= \phi C_k^{-1} C \text{Cov}(\beta) C_k^{-1} + k^2 C_k^{-1} \beta \beta^\top C_k^{-1}
\]

(10)

\[
\text{MSE}(\hat{\beta}_k) = \text{tr} \left[ \text{MMSE}(\hat{\beta}_k) \right]
\]

\[
= \phi \sum_{j=1}^{p} \frac{\lambda_j}{(\lambda_j + k)^2} + k^2 \sum_{j=1}^{p} \frac{\alpha_j^2}{(\lambda_j + k)^2}
\]

(11)

where \( \alpha = Q^\top \beta \).

2.2. Gamma Liu Estimator

Another popular estimator which is known as Liu estimator has been adopted to the generalized linear models by [19] and the authors considered the gamma dependent variable to study the performance of Liu gamma estimator via Monte Carlo simulation study and real data application. The gamma Liu estimator (GLE) is defined as

\[
\hat{\beta}_d = F_d \hat{\beta}_{\text{MLE}}
\]

(12)
where \( F_d = (C + I)^{-1}(C + dI) \) and \( 0 < d < 1 \). The covariance matrix and bias vector of GLE can be obtained respectively by

\[
\text{Cov} \left( \hat{\beta}_d \right) = \phi F_d C^{-1} F_d^\top
\]
(13)

\[
b_{\text{GLE}} = \text{bias} \left( \hat{\beta}_d \right) = -(1 - d) (C + I)^{-1} \beta.
\]
(14)

Using the covariance and bias of GLE, one can obtain the following MMSE and MSE functions respectively

\[
\text{MMSE} \left( \hat{\beta}_d \right) = \text{Cov} \left( \hat{\beta}_d \right) + b_{\text{GLE}} b_{\text{GLE}}^\top
\]
\[
= \phi F_d C^{-1} F_d^\top + (1 - d)^2 (C + I)^{-1} \beta \beta^\top (C + I)^{-1}
\]
(15)

\[
\text{MSE} \left( \hat{\beta}_d \right) = \text{tr} \left[ \text{MMSE} \left( \hat{\beta}_d \right) \right]
\]
\[
= \phi \sum_{j=1}^{p} \frac{(\lambda_j + d)^2}{\lambda_j (\lambda_j + 1)^2} + (1 - d)^2 \sum_{j=1}^{p} \frac{\alpha_j^2}{(\lambda_j + 1)^2}.
\]
(16)

2.3. Gamma Two-parameter Estimator

In this paper, we propose to adopt the estimator defined by [28] in the gamma regression model, we call this estimator as gamma two-parameter estimator (GTPE) which is defined as follows

\[
\hat{\beta}_{(k,d)} = C_k^{-1} C_{kd} \hat{\beta}_{\text{MLE}}
\]
\[
= F_{kd} \hat{\beta}_{\text{MLE}}
\]
(17)

where \( 0 < d < 1, k > 0, C_k = C + kI, C_{kd} = C + kdI \) and \( F_{kd} = C_k^{-1} C_{kd} \). We obtain the covariance matrix and bias vector of GTPE as

\[
\text{Cov} \left( \hat{\beta}_{(k,d)} \right) = \phi F_{kd} C^{-1} F_{kd}^\top
\]
(18)

\[
b_{\text{GTPE}} = \text{bias} \left( \hat{\beta}_{(k,d)} \right) = k(d - 1)C_k^{-1} \beta.
\]
(19)

Therefore, MMSE and MSE functions of GTPE are respectively computed as

\[
\text{MMSE} \left( \hat{\beta}_{(k,d)} \right) = \text{Cov} \left( \hat{\beta}_{(k,d)} \right) + b_{\text{GTPE}} b_{\text{GTPE}}^\top
\]
\[
= \phi F_{kd} C^{-1} F_{kd}^\top + k^2 (d - 1)^2 C_k^{-1} \beta \beta^\top C_k^{-1}
\]
(20)

\[
\text{MSE} \left( \hat{\beta}_{(k,d)} \right) = \text{tr} \left[ \text{MMSE} \left( \hat{\beta}_{(k,d)} \right) \right]
\]
\[
= \phi \sum_{j=1}^{p} \frac{(\lambda_j + kd)^2}{\lambda_j (\lambda_j + k)^2} + k^2 (d - 1)^2 \sum_{j=1}^{p} \frac{\alpha_j^2}{(\lambda_j + k)^2}.
\]
(21)

2.4. Theoretical Comparisons Between Estimators

In this subsection, we provide some theorems comparing the MSE and MMSE functions of the listed estimators. To do so, we consider the MSE and MMSE differences and investigate under which conditions they are positive definite. If a matrix \( A \) is positive definite, then we write \( A > 0 \).

Now, we will present three lemmas and use them to prove some of the theorems given in this section.

**Lemma 1**

[14] Suppose that \( M \) be a positive definite matrix, namely \( M > 0 \), \( \alpha \) be some vector, then \( M - \alpha \alpha^\top \geq 0 \) if and only if \( \alpha^\top M^{-1} \alpha \leq 1 \).

Lemma 2
[31] Let $M > 0$, $N > 0$, then $M > N$, if and only if $\eta_{\text{max}} (NM^{-1}) < 1$, where $\eta_{\text{max}} (A)$ is the maximum eigenvalue of some matrix $A$.

Lemma 3
[31] Let $\hat{\beta}_j = A_j y, j = 1, 2$ be two competing estimator of $\beta$. Assume that $\Delta = \text{Cov} (\hat{\beta}_1) - \text{Cov} (\hat{\beta}_2) > 0$, then $\text{MSEM} (\hat{\beta}_1) - \text{MSEM} (\hat{\beta}_2) > 0$ if and only if $u_2^T (\Delta + u_1 u_1^T)^{-1} u_2 \leq 1$, where $u_j$ denotes the bias of $\hat{\beta}_j$.

In the next theorem, we compare MLE and GTPE using the MMSE functions.

Theorem 1
When $\eta_{\text{max}} (F_{kd} C^{-1} F_{kd}^T) < 1$ the new estimator GTPE is superior to MLE in the sense of MMSE if and only if $b_{\text{GLTE}}^T D_1^{-1} b_{\text{GLTE}} < 1$, where $D_1 = \text{Cov} (\hat{\beta}_{\text{MLE}}) - \text{Cov} (\hat{\beta}_{(k,d)})$.

Proof
Let us consider the following MMSE difference
$$\Delta_1 = \text{MMSE} (\hat{\beta}_{\text{MLE}}) - \text{MMSE} (\hat{\beta}_{(k,d)}) = \phi (C^{-1} - F_{kd} C^{-1} F_{kd}) - b_{\text{GLTE}}^T b_{\text{GLTE}}^T$$
(22)

Since the matrices $C^{-1}$ and $F_{kd} C^{-1} F_{kd}^T$ are positive definite then by Lemma 2, $D_1 = C^{-1} - F_{kd} C^{-1} F_{kd} > 0$ if $\eta_{\text{max}} (F_{kd} C^{-1} F_{kd}) < 1$. Then, by Lemma 1, $\Delta_1 > 0$ if and only if $b_{\text{GLTE}}^T D_1^{-1} b_{\text{GLTE}} < 1$. Thus, the proof is finished.

Theorem 2
The new estimator GLTE is superior to MLE in the sense of MSE if $\min \left\{ \frac{2 \phi \lambda_j}{\lambda_j \alpha_j^2 - \phi} \right\}_{j=1}^p > k(1 - d)$ where $k > 0$ and $0 < d < 1$.

Proof
The MSE differences of MLE and GTPE is given as
$$\text{MSE} (\hat{\beta}_{\text{MLE}}) - \text{MSE} (\hat{\beta}_{(k,d)}) = \phi \sum_{j=1}^p \frac{1}{\lambda_j} - \phi \sum_{j=1}^p \frac{(\lambda_j + k d)^2}{\lambda_j (\lambda_j + k)^2} + k^2 (d - 1)^2 \sum_{j=1}^p \frac{\alpha_j^2}{\lambda_j (\lambda_j + k)^2}$$
$$= \sum_{j=1}^p \frac{1}{\lambda_j (\lambda_j + k)^2} \left\{ \phi \left[ (\lambda_j + k)^2 - (\lambda_j + kd)^2 \right] - k^2 (d - 1)^2 \lambda_j \alpha_j^2 \right\}$$
$$= \sum_{j=1}^p \frac{k (1 - d)}{\lambda_j (\lambda_j + k)^2} \left\{ 2 \phi \lambda_j + \phi k (1 - d) - k (1 - d) \lambda_j \alpha_j^2 \right\}$$
(23)

The Equation (23) becomes positive if $2 \phi \lambda_j + \phi k (1 - d) - k (1 - d) \lambda_j \alpha_j^2 > 0$ which is satisfied if we have $\min \left\{ \frac{2 \phi \lambda_j}{\lambda_j \alpha_j^2 - \phi} \right\}_{j=1}^p > k(1 - d)$. Thus, the proof is finished.

In the next two theorems, we provide the conditions that GTPE is superior to GRE in both MMSE and MSE sense respectively.

Theorem 3
When $\eta_{\text{max}} (F_{kd} C^{-1} C_k C^{-1} C_k) < 1$, the new estimator GTPE is superior to GRE in the sense of MMSE if and only if $b_{\text{GLTE}}^T D_2 + b_{\text{GRE}} b_{\text{GRE}}^T b_{\text{GLTE}} \leq 1$ where $D_2 = C_k C^{-1} C_k - F_{kd} C^{-1} F_{kd}$.
Proof
Now, consider the difference of MMSE functions of GRE and GTPE
\[
\Delta_2 = \text{MMSE} \left( \hat{\beta}_k \right) - \text{MMSE} \left( \hat{\beta}_{(k,d)} \right) \\
= \phi \left( C_k C_k - F_{kd} C^{-1} F_{kd}^\top \right) + b_{\text{GRE}} b_{\text{GRE}}^\top - b_{\text{GLTE}} b_{\text{GLTE}}^\top
\]
(24)
Since $C_k C_k$ and $F_{kd} C^{-1} F_{kd}^\top$ are positive definite, then by Lemma 2, when
\[
\eta_{\text{max}} \left( F_{kd} C^{-1} F_{kd} [C_k C_k]^{-1} \right) = \eta_{\text{max}} \left( F_{kd} C^{-1} C_k C^{-1} C_k \right) < 1,
\]
$D_2 = C_k C_k - F_{kd} C^{-1} F_{kd} > 0$. Therefore, by Lemma 3, $b_{\text{GLTE}}^\top [D_2 + b_{\text{GRE}} b_{\text{GRE}}^\top] b_{\text{GLTE}} \leq 1$ if and only if $\Delta_2 > 0$. The proof is completed.

Theorem 4
The new estimator GTPE is superior to GRE in the sense of MSE
\begin{itemize}
  \item if $d > 0$ and $\min \left\{ \frac{\phi (2\lambda_j - d)}{\lambda_j \alpha_j^2} \right\} > d + 2k$ or
  \item if $d < 0$ and $\max \left\{ \frac{\phi (2\lambda_j - d)}{\lambda_j \alpha_j^2} \right\} < d + 2k$.
\end{itemize}
where $k > 0$.

Proof
The MSE differences of GRE and GTPE is obtained as
\[
\text{MSE} \left( \hat{\beta}_k \right) - \text{MSE} \left( \hat{\beta}_{(k,d)} \right) = \sum_{j=1}^{p} \left\{ \frac{\phi \lambda_j}{(\lambda_j + k)^2} + \frac{k^2 \alpha_j^2}{(\lambda_j + k)^2} \right\} - \sum_{j=1}^{p} \left\{ \frac{\phi \lambda_j}{\lambda_j} + \frac{k^2 (1 - d)^2 \alpha_j^2}{(\lambda_j + k)^2} \right\}
\]
\[
= \sum_{j=1}^{p} \frac{1}{\lambda_j} \left\{ \phi \lambda_j^2 - \phi (\lambda_j + k)^2 + k^2 \lambda_j \alpha_j^2 - k^2 (1 - d)^2 \lambda_j \alpha_j^2 \right\}
\]
\[
= \sum_{j=1}^{p} \frac{1}{\lambda_j} \left\{ k^2 d \lambda_j \alpha_j^2 (2 - d) - \phi k (2 \lambda_j - k d) \right\}.
\]
(25)
Thus, Equation (25) becomes positive if $k > \max \left\{ \frac{2 \phi \lambda_j}{2 \lambda_j \alpha_j^2 + d (\phi - \lambda_j \alpha_j^2)} \right\}.$

Finally, in the next theorem, we obtain the conditions that GTPE is better than GLE in the sense of MMSE.

Theorem 5
When $\mu_{\text{max}} \left( F_{kd} C^{-1} F_{kd}^\top \left[ F_d C^{-1} F_d^\top \right]^{-1} \right) < 1$, the new estimator GTPE is better than GLE in MMSE sense if $b_{\text{GLTE}}^\top [D_3 + b_{\text{GLE}} b_{\text{GLE}}^\top] b_{\text{GLTE}} \leq 1$ where $D_3 = F_d C^{-1} F_d^\top - F_{kd} C^{-1} F_{kd}^\top$.

Proof
Consider the following MMSE difference
\[
\Delta_3 = \text{MMSE} \left( \hat{\beta}_d \right) - \text{MMSE} \left( \hat{\beta}_{(k,d)} \right)
\]
\[
= \phi \left( F_d C^{-1} F_d^\top - F_{kd} C^{-1} F_{kd}^\top \right) + b_{\text{GLE}} b_{\text{GLE}}^\top - b_{\text{GLTE}} b_{\text{GLTE}}^\top
\]
(26)
Since $F_dC^{-1}F_d^T$ and $F_kdC^{-1}F_kd^T$ are positive definite, then by Lemma 2, when

$$\mu_{\max}\left( F_kdC^{-1}F_kd^T \left[ F_dC^{-1}F_d^T \right]^{-1} \right) < 1$$

$D_3 = F_dC^{-1}F_d^T - F_kdC^{-1}F_kd^T > 0$. Therefore, by Lemma 3, $b_{GLTE}^T D_3 b_{GLTE} \leq 1$ if and only if $\Delta_3 > 0$. The proof is completed.

2.5. Selection of the biasing parameters k and d

In order to obtain less MSE values and efficient regression coefficients, one needs to estimate the parameters of GTPE accordingly. Following [15] and [10], the values of $k$ and $d$ may be chosen iteratively as follows: Firstly, taking derivative of Equation (21) with respect to the parameter $k$ and equating the resulting quantity to zero then one obtains the following

$$\frac{\partial \text{MSE} \left( \hat{\beta}(k,d) \right)}{\partial k} = 2 \sum_{j=1}^{p} \frac{\phi (\lambda_j + kd)(d - 1) + k(d - 1)^2 \lambda_j \alpha_j^2}{(\lambda_j + k)^3} = 0. \quad (27)$$

As in [15] and [17], it is possible to equate the numerator of Eq. (27) to zero and solve for each individual parameter as

$$k_j = \frac{\phi \lambda_j}{(1 - d) \lambda_j \alpha_j^2 - \phi d} \quad (28)$$

where $j = 1, 2, ..., p$. Since each $k_j$ should be positive, we can obtain a condition such that $0 < d < 1$ by

$$d < \min \left\{ \frac{\lambda_j \alpha_j^2}{\phi + \lambda_j \alpha_j^2} \right\}. \quad (29)$$

Now, after obtaining an estimate of $d$ using Eq. (29), it is easy to compute the individual parameters $k_j$’s. However, we only need an estimate of the parameter $k$. Therefore, following [10], we propose the following estimators of $k$:

- $k_1 = \frac{1}{p} \sum_{j=1}^{p} \left\{ \frac{\phi \lambda_j}{(1 - d) \lambda_j \alpha_j^2 - \phi d} \right\}$ which is the arithmetic mean of $k_j$’s.
- $k_2 = \text{median} \left\{ \frac{\phi \lambda_j}{(1 - d) \lambda_j \alpha_j^2 - \phi d} \right\}$
- $k_3 = \min \left\{ \frac{\phi \lambda_j}{(1 - d) \lambda_j \alpha_j^2 - \phi d} \right\}$
- $k_4 = \max \left\{ \frac{\phi \lambda_j}{(1 - d) \lambda_j \alpha_j^2 - \phi d} \right\}$
- $k_5 = \left\{ \prod_{j=1}^{p} \left\{ \frac{\phi \lambda_j}{(1 - d) \lambda_j \alpha_j^2 - \phi d} \right\} \right\}^{\frac{1}{p}}$ which is the geometric mean of $k_j$’s.

3. Monte Carlo Simulation

In this section, a Monte Carlo simulation experiment is used to examine the performance of GTPE with different degrees of multicollinearity.

3.1. Simulation design

The response variable of $n$ observations from gamma regression model is generated by

$$y_i \sim \text{Gamma}(\theta_i, 1.5), \quad (30)$$
where $\theta_i = \exp(x_i^T \beta)$, $\beta = (\beta_1, \ldots, \beta_p)$ with $\sum_{j=1}^p \beta_j^2 = 1$ and $\beta_1 = \beta_2 = \ldots = \beta_p$ [17]. The explanatory variables $x_i^T = (x_{i1}, x_{i2}, \ldots, x_{in})$ have been generated from the following formula

$$x_{ij} = (1 - \rho^2)^{1/2} w_{ij} + \rho w_{ip}, \quad i = 1, 2, \ldots, n, \quad j = 1, 2, \ldots, p,$$

where $\rho$ 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: 50, 100 and 150. 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, since we are interested in the effect of multicollinearity, in which the degrees of correlation considered to be more important, three values of the pairwise correlation are taken into account such that $\rho = \{0.90, 0.95, 0.99\}$. For a combination of these different values of $n, p$ and $\rho$ the generated data is repeated 1000 times. For each replication, the MSE is calculated as

$$\text{MSE}_i \left( \hat{\beta} \right) = \left( \hat{\beta} - \beta \right)^T \left( \hat{\beta} - \beta \right), \quad i = 1, 2, \ldots, 1000,$$

where $\hat{\beta}$ is defined as

- $\hat{\beta}_{\text{MLE}}$
- $\hat{\beta}_k$ with $k = \frac{\hat{\phi}}{\hat{\phi}^2 + d^2}$, $\hat{\alpha}_j = Q^T \hat{\beta}_{\text{MLE}}$ and $\hat{\phi}$ is the estimated dispersion parameter which is computed as $\hat{\phi} = (n - p)^{-1} \sum_{i=1}^n (y_i - \hat{\mu}_i)^2$, (Pearson residual [18]).
- $\hat{\beta}_d$ with an optimal $d$ value as in [27], $d = \max(0, \max(\lambda_j(\hat{\alpha}_j^2 - \hat{\phi})/(\hat{\phi} + \lambda_j \hat{\alpha}_j^2)))$.
- $\hat{\beta}_{(k,d)}$ with the optimal $d$ (Eq. (29)) after substituting $k = \hat{k}$, $\lambda_j = \hat{\lambda}_j$, $\alpha_j = \hat{\alpha}_j$, and $\phi = \hat{\phi}$.

The averaged performance from 1000 simulations are summarized in terms of the MSE. All the computations are done using R programming language [30].

3.2. Simulation results

The average MSE values of the MLE, GRE, GLE, and GTPE are presented in Tables 1-3 for $n = 50$, $n = 100$, and $n = 150$, respectively. We may observe that the MSE of GTPE in all situations of selecting $k$ is less than the other used estimators, which clearly shows that the GTPE outperforms the MLE, GRE, and GLE in all of the cases. It is clearly seen that the optimal selection $k_3$ in GTPE works fine comparing with $k_1, k_2, k_4$, and $k_4$ with respect to MSE.

Further, the MSE values clearly increase when $p$ increases. However, there is a clear advantage of the GTPE over the other competitor estimators. In addition, it is clear that increasing the $n$ values leading to decreasing in MSE. For the large $n$, however, there is a clear advantage of using the GTPE.

Regarding the degree of correlation, the MSE increases monotonously when the correlation between the explanatory variables increasing. The GTPE again outperforms the MLE, GRE, and GLE.

4. A Real Data Application

To further demonstrate the usefulness of the GLTE in real application, we present here a chemistry dataset with $(n, p) = (65, 15)$, where $n$ is representing the number of imidazo[4,5-b]pyridine derivatives, which are used as anticancer compounds. While, $p$ is denoting the number of molecular descriptors, which are treated as explanatory variables [4]. The response of interest is the biological activities ($IC_{50}$). Quantitative structure-activity relationship (QSAR) study has become a great deal of importance in chemometrics. The principle of QSAR is to model several biological activities over a collection of chemical
compounds in terms of their structural properties [5]. Consequently, using of multiple regression model is one of the most important tools for constructing the QSAR model. A description of the used explanatory variables is provided in Table 4. All the variables are numerical.

First, to check whether the response variable belongs to the gamma distribution, Chi-square test is used. The result of the test equals to 9.3657 with p-value equals to 0.4534. It is indicated from this result that the gamma distribution fits very well to this response variable.

Second, to check whether there is a relationship among the explanatory variables or not, Figure 1 displays the correlation matrix among the 15 explanatory variables. It is obviously seen that there are correlations greater than 0.90 among MW, SpMaxA_D, and AT8v (r = 0.96), between SpMax3_Bh(s) and AT8v (r = 0.92), and between Mor21v and Mor21e (r = 0.93).

Third, to test the existence of multicollinearity, the eigenvalues of the matrix $X^\top WX$ are obtained as $2.14 \times 10^9, 3.85 \times 10^6, 2.42 \times 10^5, 1.26 \times 10^4, 1.29 \times 10^3, 2.14 \times 10^9, 9.01 \times 10^2, 4.71 \times 10^2$, $1.71 \times 10^2, 5.93 \times 10^1, 3.24 \times 10^1, 2.77 \times 10^1, 1.78 \times 10^1, 9.56$, and 1.23. The determined condition number $CN = \sqrt{\lambda_{\text{max}}/\lambda_{\text{min}}}$ of the data is 41652.77 indicating that the severe multicollinearity issue exists.

The estimated gamma regression coefficients and MSE values for the MLE, GRE, GLE, and GTPE using $k_3$ estimators are listed in Table 5. According to Table 5, it is clearly seen that the GTPE shrinkages the value of the estimated coefficients efficiently. Additionally, in terms of the MSE, there is an important

Table 1. Averaged MSE values when $n = 50$

<table>
<thead>
<tr>
<th>$p$</th>
<th>$\rho$</th>
<th>MLE</th>
<th>GRE</th>
<th>GLE</th>
<th>GTPE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>$k_1$</td>
<td>$k_2$</td>
<td>$k_3$</td>
<td>$k_4$</td>
</tr>
<tr>
<td>4</td>
<td>0.90</td>
<td>4.311</td>
<td>1.834</td>
<td>1.138</td>
<td>1.126</td>
</tr>
<tr>
<td></td>
<td>0.95</td>
<td>4.567</td>
<td>1.954</td>
<td>1.159</td>
<td>1.411</td>
</tr>
<tr>
<td></td>
<td>0.99</td>
<td>6.993</td>
<td>2.283</td>
<td>1.181</td>
<td>1.439</td>
</tr>
<tr>
<td>8</td>
<td>0.90</td>
<td>4.348</td>
<td>1.874</td>
<td>1.245</td>
<td>1.488</td>
</tr>
<tr>
<td></td>
<td>0.95</td>
<td>4.891</td>
<td>1.987</td>
<td>1.253</td>
<td>1.569</td>
</tr>
<tr>
<td></td>
<td>0.99</td>
<td>10.417</td>
<td>2.775</td>
<td>1.255</td>
<td>1.605</td>
</tr>
</tbody>
</table>

Table 2. Averaged MSE values when $n = 100$

<table>
<thead>
<tr>
<th>$p$</th>
<th>$\rho$</th>
<th>MLE</th>
<th>GRE</th>
<th>GLE</th>
<th>GTPE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>$k_1$</td>
<td>$k_2$</td>
<td>$k_3$</td>
<td>$k_4$</td>
</tr>
<tr>
<td>4</td>
<td>0.90</td>
<td>4.281</td>
<td>1.814</td>
<td>1.118</td>
<td>1.359</td>
</tr>
<tr>
<td></td>
<td>0.95</td>
<td>4.547</td>
<td>1.934</td>
<td>1.14</td>
<td>1.391</td>
</tr>
<tr>
<td></td>
<td>0.99</td>
<td>6.973</td>
<td>2.264</td>
<td>1.161</td>
<td>1.419</td>
</tr>
<tr>
<td>8</td>
<td>0.90</td>
<td>4.328</td>
<td>1.854</td>
<td>1.225</td>
<td>1.468</td>
</tr>
<tr>
<td></td>
<td>0.95</td>
<td>4.871</td>
<td>1.967</td>
<td>1.233</td>
<td>1.549</td>
</tr>
<tr>
<td></td>
<td>0.99</td>
<td>10.398</td>
<td>2.755</td>
<td>1.235</td>
<td>1.585</td>
</tr>
</tbody>
</table>

Table 3. Averaged MSE values when $n = 150$

<table>
<thead>
<tr>
<th>$p$</th>
<th>$\rho$</th>
<th>MLE</th>
<th>GRE</th>
<th>GLE</th>
<th>GTPE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>$k_1$</td>
<td>$k_2$</td>
<td>$k_3$</td>
<td>$k_4$</td>
</tr>
<tr>
<td>4</td>
<td>0.90</td>
<td>4.228</td>
<td>1.761</td>
<td>1.066</td>
<td>1.306</td>
</tr>
<tr>
<td></td>
<td>0.95</td>
<td>4.495</td>
<td>1.881</td>
<td>1.087</td>
<td>1.339</td>
</tr>
<tr>
<td></td>
<td>0.99</td>
<td>6.921</td>
<td>2.211</td>
<td>1.108</td>
<td>1.366</td>
</tr>
<tr>
<td>8</td>
<td>0.90</td>
<td>4.275</td>
<td>1.802</td>
<td>1.172</td>
<td>1.415</td>
</tr>
<tr>
<td></td>
<td>0.95</td>
<td>4.818</td>
<td>1.914</td>
<td>1.182</td>
<td>1.497</td>
</tr>
<tr>
<td></td>
<td>0.99</td>
<td>10.345</td>
<td>2.702</td>
<td>1.182</td>
<td>1.532</td>
</tr>
</tbody>
</table>
Table 4. Description of the used explanatory variables

<table>
<thead>
<tr>
<th>Variable name’s</th>
<th>description</th>
</tr>
</thead>
<tbody>
<tr>
<td>MW</td>
<td>molecular weight</td>
</tr>
<tr>
<td>IC3</td>
<td>Information Content index (neighborhood symmetry of 3-order)</td>
</tr>
<tr>
<td>SpMaxA_D</td>
<td>normalized leading eigenvalue from topological distance matrix</td>
</tr>
<tr>
<td>ATS8v</td>
<td>Broto-Moreau autocorrelation of lag 8 (log function) weighted by van der Waals volume</td>
</tr>
<tr>
<td>MATS7v</td>
<td>Moran autocorrelation of lag 7 weighted by van der Waals volume</td>
</tr>
<tr>
<td>MATS2s</td>
<td>Moran autocorrelation of lag 2 weighted by I-state</td>
</tr>
<tr>
<td>GATS4p</td>
<td>Geary autocorrelation of lag 4 weighted by polarizability</td>
</tr>
<tr>
<td>SpMax8_Bh(p)</td>
<td>largest eigenvalue n. 8 of Burden matrix weighted by polarizability</td>
</tr>
<tr>
<td>SpMax3_Bh(s)</td>
<td>largest eigenvalue n. 3 of Burden matrix weighted by I-state</td>
</tr>
<tr>
<td>P_VSA_e_3</td>
<td>P_VSA-like on Sanderson electronegativity, bin 3</td>
</tr>
<tr>
<td>TDB08m</td>
<td>3D Topological distance based descriptors - lag 8 weighted by mass</td>
</tr>
<tr>
<td>RDF100m</td>
<td>Radial Distribution Function - 100 / weighted by mass</td>
</tr>
<tr>
<td>Mor21v</td>
<td>signal 21 / weighted by van der Waals volume</td>
</tr>
<tr>
<td>Mor21e</td>
<td>signal 21 / weighted by Sanderson electronegativity</td>
</tr>
<tr>
<td>HATS6v</td>
<td>leverage-weighted autocorrelation of lag 6 / weighted by van der Waals volume</td>
</tr>
</tbody>
</table>

Figure 1. Correlation matrix of the real data.

reduction in favor of the GTPE. Specifically, it can be seen that the MSE of the GTPE was about 46.11%, 31.09%, and 26.40% lower than that of MLE, GRE, and GLE, respectively.

5. Conclusion

In this paper, gamma two-parameter estimator is proposed to overcome the multicollinearity problem in the gamma regression model. According to Monte Carlo simulation studies, the GTPE has better performance than MLE, GRE, and GLE estimators, in terms of MSE. Additionally, a real data application is also considered to illustrate benefits of using the GTPE in the context of gamma regression model.
Table 5. The estimated coefficients and MSE values of the listed estimators

<table>
<thead>
<tr>
<th></th>
<th>MLE</th>
<th>GRE</th>
<th>GLE</th>
<th>GTPE (k_3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \hat{\beta}_{\text{MW}} )</td>
<td>1.0383</td>
<td>0.7773</td>
<td>0.8713</td>
<td>0.7703</td>
</tr>
<tr>
<td>( \hat{\beta}_{\text{IC3}} )</td>
<td>1.2733</td>
<td>1.0133</td>
<td>1.1063</td>
<td>1.0053</td>
</tr>
<tr>
<td>( \hat{\beta}_{\text{SpMaxA-D}} )</td>
<td>-1.0657</td>
<td>-1.3267</td>
<td>-1.2327</td>
<td>-0.8657</td>
</tr>
<tr>
<td>( \hat{\beta}_{\text{AT85v}} )</td>
<td>-1.3427</td>
<td>-1.6037</td>
<td>-1.5097</td>
<td>-1.1427</td>
</tr>
<tr>
<td>( \hat{\beta}_{\text{MATS7v}} )</td>
<td>-1.1827</td>
<td>-1.4437</td>
<td>-1.3497</td>
<td>-0.9827</td>
</tr>
<tr>
<td>( \hat{\beta}_{\text{MATS2s}} )</td>
<td>-1.1787</td>
<td>-1.4397</td>
<td>-1.3457</td>
<td>-0.9787</td>
</tr>
<tr>
<td>( \hat{\beta}_{\text{GATS4p}} )</td>
<td>-1.2007</td>
<td>-1.4617</td>
<td>-1.3687</td>
<td>-1.0007</td>
</tr>
<tr>
<td>( \hat{\beta}_{\text{SpMax8_Bh(p)}} )</td>
<td>2.5423</td>
<td>2.2813</td>
<td>2.3753</td>
<td>2.7433</td>
</tr>
<tr>
<td>( \hat{\beta}_{\text{SpMax3_Bh(s)}} )</td>
<td>2.1053</td>
<td>1.8443</td>
<td>1.9383</td>
<td>2.3053</td>
</tr>
<tr>
<td>( \hat{\beta}_{\text{P_Vsa_e_3}} )</td>
<td>2.0373</td>
<td>1.7753</td>
<td>1.8693</td>
<td>2.2363</td>
</tr>
<tr>
<td>( \hat{\beta}_{\text{TDB08m}} )</td>
<td>-2.0667</td>
<td>-2.3287</td>
<td>-2.2337</td>
<td>-1.8667</td>
</tr>
<tr>
<td>( \hat{\beta}_{\text{RDF100m}} )</td>
<td>1.6073</td>
<td>1.3453</td>
<td>1.4393</td>
<td>1.8063</td>
</tr>
<tr>
<td>( \hat{\beta}_{\text{Mor21v}} )</td>
<td>-2.3977</td>
<td>-2.6587</td>
<td>-2.5647</td>
<td>-2.1987</td>
</tr>
<tr>
<td>( \hat{\beta}_{\text{Mor21e}} )</td>
<td>-2.3157</td>
<td>-2.5767</td>
<td>-2.4827</td>
<td>-2.1157</td>
</tr>
<tr>
<td>( \hat{\beta}_{\text{HATS6v}} )</td>
<td>2.2473</td>
<td>1.9863</td>
<td>2.0803</td>
<td>2.4473</td>
</tr>
<tr>
<td>MSE</td>
<td>4.075</td>
<td>3.187</td>
<td>2.984</td>
<td>2.196</td>
</tr>
</tbody>
</table>

The superiority of the GTPE based on the resulting MSE was observed and it was shown that the results are consistent with Monte Carlo simulation results. In conclusion, the use of the GTPE is recommended when multicollinearity is present in the gamma regression model.

REFERENCES

34. N. N. Urgan, and M. Tüz, Liu estimator in logistic regression when the data are collinear, In 29th EURO Mini Conference (pp. 323-327), 2008.