Journal of Information Systems Engineering and Management
2025, 10(33s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

New Robust Beta Regression Estimation to Overcome the Effect of High Leverage Points

Hamza Lateef Katea Al-Ayashy and Taha Alshaybawee

Department of Statistics, College of Administration and Economics, University of Al-Qadisiyah, Al-Qadisiyah, Iraq

Emails: Hamzakufa1986@gmail.com , taha.alshaybawee@qu.edu.iq

ARTICLE INFO

Received: 18 Dec 2024

Revised: 20 Feb 2025

Accepted: 25 Feb 2025

ABSTRACT

Beta regression models are widely used to model continuous data with a unit interval (0,1), such as ratios, fractions, and rates. The maximum likelihood method is typically used to estimate regression coefficients in beta regression models. However, the maximum likelihood estimator is highly sensitive to outliers. Several studies have proposed methods to address this problem, such as the M-Huber, S-estimator, LMS, and LTS estimators. However, these methods suffer from the problem of high leverage points (HLPoints) in the independent variables. The GM-estimator is one of the methods that address this problem. This study examines beta regression analysis, focusing on the effect of high leverage points (HLPoints) on parameter estimates. Monte Carlo simulations and real data are conducted to evaluate and compare the performance of the proposed robust method, Generalized M-Beta Regression (GMBr), with the existing beta regression method estimated using Maximum Likelihood (MLBr).

Keywords: Beta regression; Outliers; High leverage points (HLPoints); Generalized M, Robust estimation.

1- Introduction

Regression analysis is one of the most widely used statistical methods in various fields. It describes the relationship between a response variable and one or more independent variables, with the aim of determining the form of their association through a mathematical model. When the response variable (y) is expressed as fractions or percentages -- that is, when its values are limited to the interval (0, 1), beta regression is typically used. Beta regression is a well-known statistical model that finds applications in the natural sciences, medicine, finance, economics, environment, hydrology, psychology, and many other disciplines.

Maximum likelihood estimations (MLE) are commonly used to estimate the parameters of a beta regression model. (Ferrari & Cribari-Neto, 2004) proposed the beta regression model, and several studies have been conducted on different aspects of beta regression. For example, (Vasconcellos & Cribari-Neto, 2005) investigated the behavior of MLEs in a beta regression model where the distribution parameters are nonlinear functions of linear combinations of explanatory variables with unknown coefficients. (Ospina et al., 2006) conducted a study that involved point and interval optimization of a beta regression model. (Bayes et al., 2012a) presented a new regression model for proportions by incorporating the rectangular beta distribution proposed by (Bayes et al., 2012b) Hahn (2008). (Espinheira et al., 2015) studied separation measures in beta regression models, and (Bayer & Cribari-Neto, 2017) presented a study titled "Model selection criteria in beta regression with variable dispersion." (Espinheira et al., 2019) presented a study on model selection criteria for beta regression, and (Ghosh, 2019) conducted a study on developing a robust inference procedure for beta regression model. Additional studies have contributed to the development of beta regression methodologies.

Journal of Information Systems Engineering and Management
2025, 10(33s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/ **Research Article**

(Karlsson et al., 2020) proposed Liu shrinkage estimators for beta regression models, whereas (Abonazel, Dawoud, et al., 2022) proposed the Dawood-Kabria estimator for beta regression models. (Zhou & Huang, 2022) presented a study involving Bayesian empirical regression for limited responses with unknown support. (Abonazel & Taha, 2023) studied beta ridge regression estimators. (Ospina et al., 2023) proposed a robust semi-parametric inference for two-stage production models using a beta regression approach. (Wilcox, 2011) conducted a comparative study on robust estimation of a beta regression model in the presence of outliers. (Ribeiro & Ferrari, 2023) investigated robust estimation in beta regression using the Lq maximum likelihood method (Maluf et al., 2024) applied robust beta regression using logit transformation, and (Heng & Lange, 2025) conducted a study involving a preliminary estimate of the proportion of outliers in robust regression.

The success or failure of the estimator depends on the method used to estimate the model parameters. The ordinary least squares (OLS) method is known to be very sensitive to outliers as it has a breaking point close to (0), and the inference based on the maximum likelihood estimator lacks power when outliers are present. This sensitivity can lead to severe bias and misleading conclusions. Outliers represent one of the oldest and most important challenges in statistics. (Rosseeuw & Van Zomeren, 1990) classified outliers into vertical outliers that appear in the response variable (y), leverage points that appear in the independent variables (x), and outliers in both directions. When diagnosing the problem of the presence of outliers, many researchers treated the outliers by removing them and analyzing the remaining observations. Other studies have emphasized the importance of retaining these data, because any lack of observations can lead to the loss of valuable information and negatively affect the accuracy of the results. Instead, robust methods for dealing with outliers have been used, ensuring that the estimators and their associated test statistics are relatively insensitive to the presence of outliers. These robust methods, which include the M-Huber, S-estimator, LMS, and LTS estimators (Lim & Midi, 2016), generally provide more efficient results than traditional methods. However, even these robust methods can be affected by high leverage points (HLPoints) in the independent variables. In this study, we modeled the data using beta regression and used maximum likelihood estimation to estimate the model parameters. To address the problem of outliers, we used generalized M-estimators (GM estimators) that can be powerful tools for mitigating the effect of high leverage points (HLPoints) in the independent variables.

2- Beta Regression Model.

Practitioners typically use linear regression models to examine the relationship between a dependent variable and selected explanatory variables, as well as to analyze their effects. However, this approach is unsuitable when the dependent variable is constrained within the interval (0, 1), since it may lead to predicted values that exceed these boundaries. Consequently, inferences based on the normality assumption can be misleading.

The beta regression model was first introduced by (Ferrari & Cribari-Neto, 2004), linking the mean of the dependent variable to a set of linear predictors through a monotonic and differentiable link function. This model includes a precision parameter, the inverse of which is referred to as the dispersion scale. In the basic form of the beta regression model, the precision parameter is assumed to be constant across all observations. However, some studies—such as those by (Smithson & Verkuilen, 2006) and (Algamal et al., 2023) - have indicated that the precision parameter may vary across observations in certain cases.

Let y is a continuous random that follows a beta distribution with the following probability density function;

$$f(y;\mu,\lambda) = \frac{\Gamma(\lambda)}{\Gamma(\mu\lambda)\Gamma((1-\mu)\lambda)} y^{(\mu\lambda)-1}\,(1-y)^{(1-\mu)\lambda-1} \qquad , 0< y < 1\,; 0 < \mu < 1\,;\ \lambda > 0$$

(1)

Journal of Information Systems Engineering and Management
2025, 10(33s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/ Research Article

where $\Gamma(\cdot)$ is the gamma function, and λ is the precision parameter that can be written as in (Bayer & Cribari-Neto, 2017):

$$\lambda = \frac{1-\sigma^2}{\sigma^2} \qquad (2)$$

The mean and variance of the beta probability distribution are: $E(y) = \mu$, $var(y) = \mu(1-\mu)\sigma^2$. Using the logit link function , the model allows μ_i depending on covariates as follows

$$g(\mu_i) = \log(\frac{\mu_i}{1-\mu_i}) = x_i'\beta = \eta_i \qquad (3)$$

where g(.) be a monotonic differentiable link function used to relate the systematic component with the random component, $\beta = (\beta_1, \ldots \ldots, \beta_K)'$ is a $(k \times 1)$ vector of unknown parameters, $x_i = (x_{i1}, \ldots . . x_{ik})'$ is the vector of k regression, and η_i is the linear predictor. A particularly interesting specification of the model is obtained when the logit link function is used. In the case, the mean of y_i can be written as (El-Raoof et al., 2023) by:

$$\mu_i = \frac{e^{x_i'\beta}}{1+e^{x_i'\beta}} \qquad (4)$$

Where μ_i is the mean response function. since η depends on β and the mean response μ is a function of η ,the mean μ_1,μ_2,.......,μ_n are function of β (Abonazel, Algamal, et al., 2022).

Estimation of the beta regression parameters is done by using the ML method Espinheira et al., (2008) The log-likelihood function of the beta regression model is given by:

$$L(\mu_i, \lambda; y_i) = \sum_{i=1}^{n}\{log\Gamma(\lambda) - log\Gamma(\mu_i(\lambda) - log\Gamma\big((1-\mu_i)(\lambda)\big) + (\mu_i(\lambda) - 1)\log(y_i) + \big((1-\mu_i)(\lambda) - 1\big)\log(1-y_i)\} \qquad (5)$$

Differentiating the log-likelihood in Eq. (5) with respect to β gives us the score function for β which is given by:

$$U(\beta) = \lambda x' F(y^* - \mu^*) \qquad (6)$$

Where $F = diag\left(\frac{1}{g'(\mu_i)}, \ldots . . , \frac{1}{g'(\mu_n)}\right)$, $y^* = (y_1^*, \ldots \ldots, y_n^*)'$ and $y_i^* = \log(\frac{y_i}{1-y_i})$

$\mu^* = (\mu_1^*, \ldots \ldots, \mu_n^*)'$ and $\mu_i^* = \varphi(\mu_i\lambda) - \varphi((1-\mu_i)\lambda)$ such that $\varphi\ (\cdot)$ denoting the digamma function. The iterative reweighted least-squares (IWLS) algorithm or Fisher scoring algorithm used for estimating β (Espinheira et al., 2015) .The form of this algorithm can be written as :

$$\beta^{(t+1)} = \beta^{(t)} + (I_{\beta\beta}^{(t)})^{-1}\, U_\beta^{(t)}\ (\beta)$$

Where $U_\beta^{(r)}$ is the score function defined in Eq. (6), and $I_{\beta\beta}^{(r)}$ is the information matrix for β , see (Espinheira et al., 2019) and (Algamal et al., 2023) for more details. The initial value of β can be obtained by the least squares method, while the initial value for each precision parameter equals

$$\hat{\lambda}_i = \frac{\hat{\mu}_i(1-\hat{\mu}_i)}{\hat{\sigma}_i^2} \qquad (7)$$

Journal of Information Systems Engineering and Management
2025, 10(33s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/ **Research Article**

Where $\hat{\mu}$ and σ_i^2 values are obtained from linear regression .Given $r = 0,12, \dots ..$ is the number of iterations that are performed, convergence occurs when the difference between successive estimates becomes smaller than a given small constant .At the final step, the ML estimator of β is obtained as (Abonazel, Dawoud, et al., 2022):

$$\hat{\beta}_{GM} = (x'\widehat{w}x)^{-1}x'\,\widehat{w}\hat{z} \tag{8}$$

Where X is an $n \times p$ matrix of regressors, $\hat{z} = \hat{\eta} + \widehat{w}^{-1}\hat{F}\ (y^* - \mu^*)$, and $\widehat{w}$ = $diag(\widehat{w}_1, \dots \dots \widehat{w}_n)$

$$\widehat{w}_i = (\frac{1-\hat{\sigma}^2)}{\hat{\sigma}^2}\left\{\hat{\varphi}\left(\frac{\hat{\mu}_i(1-\hat{\sigma}^2)}{\hat{\sigma}^2}\right) + \varphi'\left(\frac{(1-\hat{\mu}_i)(1-\hat{\sigma}^2)}{\hat{\sigma}^2}\right)\right\}\frac{1}{\{g'(\hat{\mu}_i)\}^2}.$$

4- Generalized M-Estimator method (GM-estimator)

The generalized M-estimator (GM-estimator) is an extension of the M-estimator method introduced by Peter J. Huber in 1964. This method aims to provide robust estimates of parameters in statistical models, especially when there are extreme values or non-normal distributions of the data (Cheng & Van Ness, 1992). It was proposed by Schweppes, as described by Hill and Paul (1977), and was developed by several researchers, most notably Frank and Hampel in the 1970s and 1980s. The basic idea behind GM estimators is to minimize the effect of high leverage points, which is the main problem with M estimators (Hampel et al., n.d.) and (Midi et al., 2021). GM estimators can deal with this problem by taking advantage of some weight functions that minimize the presence of high leverage points. The GM estimator is the solution to the following equation (Wilcox, 2011).

$$\sum_{i=1}^{n} \tau_i \psi\left(\frac{(y_i - x_i'\hat{\beta})}{s\tau_i}\right) x_i \tag{9}$$

where τ_i is an initial weight function used to minimize the effects of leverage points, the GM-estimator in convergence can be written as

$$\hat{\beta}_{GM} = (x'\widehat{w}x)^{-1}x'\,\widehat{w}\,\mathrm{y} \tag{10}$$

where W is a diagonal weight matrix with elements w_i defined as

$$w_i = \frac{\psi[(y_i - x_i'\hat{\beta}_{GML})\,/\tau_i s]}{(y_i - x_i'\hat{\beta}_{GML}/\tau_i S)} \tag{11}$$

The initial weights of GM-estimators that minimize the effect of leverage points in (9) are computed based on the hat matrix values h_{ii} as (Wilcox, 2011)

$$\tau_i = \sqrt{(1 - hii)} \tag{12}$$

(h_{ii}) is found through a (hat matrix) which is a weight matrix, this matrix is used to identify the rows of existing observations (x) that contain extreme values, and is known as regression analysis to detect the presence of (HLPoints), and these rows can be written according to the following formula (Iglewicz & Hoaglin, 1993)

$$\mathrm{H}=x'(x'x)^{-1}x \tag{13}$$

The elements of the main diagonal of the (Hat Matrix) H are written as follows (Rousseeuw & Leroy, 2003):

$$h_{ii} = x'(x'x)^{-1}x_i \tag{14}$$

Journal of Information Systems Engineering and Management
2025, 10(33s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/ **Research Article**

These elements have desirable properties as their values range between (1 0). Also, the h_{ii} set equals p and its containment index of (HLPoints) for case j is a measure of the distance between the values of (x) for case j and the average values of (x) for all cases (n). Therefore, the large values (h_{ii}) indicate the containment index of case j, and thus it is far from the center of all observations of the variable (x). The average of this matrix can be written in the following formula (PJ & Leroy, 1987)

$$H\frac{p}{n}= \frac{\sum_1^n h_{ii}}{n} =$$

p: the number of independent variables , n: the sample size.

Both (Hoaglin & Welsch, 1978) suggested that the threshold value should be when there are (HLPoints) in one variable and not another when the value is $H_{tt} > \frac{2p}{n}$, and there is also a rule of three times the mean that was presented by (Velleman & Welsch, 1981) in order to diagnose (HLPoints) when it is $H_{tt} > \frac{3p}{n}$, and Huber ,2004, presented another cut-off point as a three-point interval for the data ($0.2 < H_{tt} < 0.5$) , and accordingly the data values should be avoided $(H_{tt} > o.5)$, Kutneret , 2005, noticed that the presence of a gap between the data for the stupidest cases and the abnormal values in an unusual way is conclusive evidence of the presence of abnormal value

6- Algorithm.

GM method is one of the important methods for treating outliers and leverage points as it can be used to reduce the effect of leverage points in independent variables (x) when estimating the parameters of beta regression model An algorithm for GM-estimator can be written in the following steps:

Step 1: Choosing an initial estimates $\beta^{(0)}$ from beta regression.

Step 2: For each iteration *r* compute $\mu_i^{(t-1)} = logistic\left(x_i\beta^{(t-1)}\right) = \frac{exp(x_i\beta^{(t-1)})}{1+exp(x_i\beta^{(t-1)})}$.

Step 3: Compute the residuals $r_i^{(t-1)} = y_i - \mu_i^{(t-1)}$ and scale $\hat{\sigma}^{(t-1)} = 1.4826 * median\ of$

$largest(n-p)\ of the \left|r_i^{(t-1)}\right|$, and then compute the standardized residuals $u_i^{(t-1)} = \frac{r_i^{(t-1)}}{\hat{\sigma}^{(t-1)}}$.

Step 4: Using the following form to compute the weight $\tau_i = \sqrt{(1-h_{ii})}$.

Step 5: Based on Huber weight function standardized residuals u_i use to compute the robust weights using the form $w_i^{(t-1)} = \frac{\psi\left[\left(y_i-\mu_i^{(t-1)}\right)/\tau_i\hat{\sigma}\right]}{\left(y_i-\mu_i^{(t-1)}\right)/\tau_i\hat{\sigma}}$.

Step 6: compute $\hat{\beta}^t$ using the form $\hat{\beta}^t = (X'W^{(t-1)}X)^{-1}X'W^{(t-1)}Y$.

Step 7: Steps (2-6) are repeated until convergence.

Simulation study

This section presents a Monte Carlo simulation to assess and compare the performance of the proposed robust method Generalized-M for Beta regression (GMBr) with the existing Beta regression method estimated using maximum likelihood (MLBr). In this study, we also propose a the performance criterion that considered by (Abonazel & Taha, 2023) , referred to as mean squares error for ($\hat{\beta}$) (MSE($\hat{\beta}$)) and mean absolute error for ($\hat{\beta}$) (MAE($\hat{\beta}$)), it can calculated as follows:

$$MSE(\hat{\beta}) = \frac{1}{T}\sum_{t=1}^{T}(\hat{\beta}_t - \beta^{true})'(\hat{\beta}_t - \beta^{true}), \quad MAE((\hat{\beta}) = \frac{1}{T}\sum_{t=1}^{T}|\hat{\beta}_t - \beta^{true}|$$

In this study, we generated the explanatory variables from a multivariate normal distribution MN(0,1) with four different sample sizes (n = 30, 50, 100, and 250). Additionally, we assume two scenarios for the coefficients and the number of explanatory variables:

- $(\beta_1 = 1,\ \beta_j = 0.25 \text{ for all } j = 2, \dots, 5)$,
- $(\beta_1 = 2,\ \beta_j = 0.50 \text{ for all } j = 2, \dots, 7)$

where β_1 is the intercept, the number of explanatory variables p set as 4 and 6 . Three levels of contamination for the explanatory variables are considered $(\delta = 5\%,\ 10\%, 20\%)$ and the precision parameter is set to as λ=2 and 4. The response variable will be generated as $y_i \sim Beta(\mu_i, \lambda)$, and the experiment will be repeated 500 times (T=500). Based on the package "*betareg*" introduced by (Cribari-Neto & Zeileis, 2010), the R language was employed to conduct our simulation study.

Table(1): the MSE ($\hat{\beta}$) and MAE ($\hat{\beta}$) at Three Different Contamination Levels and different Sample Sizes in the First Scenario When the Precision Parameter is λ=2.

Cont.	Sample	methods	MSE ($\hat{\beta}$)					MAE ($\hat{\beta}$)				
			$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\beta}_5$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\beta}_5$
5%	30	BR	0.9555	0.4846	1.4284	0.9629	1.5814	1.3856	1.0651	1.9868	1.6337	2.1601
		GMBR	0.9421	0.2132	1.1438	0.7103	1.3137	1.7075	0.8377	1.7391	1.4324	1.9273
	50	BR	0.9065	0.3402	1.2666	0.8499	1.4509	1.182	0.6911	1.583	1.3163	1.7989
		GMBR	0.831	0.124	1.0538	0.6273	1.2268	1.3256	0.4964	1.3962	1.1033	1.5898
	100	BR	0.8835	0.2654	1.1944	0.7661	1.3655	1.0576	0.4516	1.3483	1.0517	1.5381
		GMBR	0.7514	0.0626	0.9926	0.5629	1.1624	1.0448	0.2545	1.1499	0.8505	1.3376
	250	BR	0.774	0.125	1.0559	0.6262	1.2249	0.8953	0.1993	1.1012	0.8016	1.2872

Journal of Information Systems Engineering and Management
2025, 10(33s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/ **Research Article**

		GMBR	0.7026	0.025	0.9551	0.5251	1.1249	0.8741	0.1877	1.0002	0.7002	1.1878
10%	30	BR	0.9606	0.4002	1.4056	0.9678	1.5566	1.4167	0.9437	1.9812	1.6663	2.1349
		GMBR	0.9454	0.1991	1.1435	0.7133	1.3121	1.7135	0.8114	1.7403	1.4443	1.9256
	50	BR	0.9047	0.3327	1.2757	0.8366	1.4278	1.1737	0.6936	1.6264	1.2962	1.7742
		GMBR	0.8341	0.1245	1.0577	0.6247	1.2248	1.3312	0.4982	1.4049	1.0983	1.5868
	100	BR	0.8828	0.2647	1.1947	0.763	1.3644	1.057	0.4506	1.3517	1.0468	1.5381
		GMBR	0.7512	0.0627	0.9937	0.5623	1.1621	1.0446	0.2502	1.1509	0.8494	1.3371
	250	BR	0.8736	0.2246	1.1557	0.7258	1.3254	0.993	0.2974	1.201	0.901	1.3882
		GMBR	0.7025	0.0248	0.9551	0.5251	1.1251	0.874	0.0996	1.0001	0.7001	1.1881
20%	30	BR	0.9658	0.4776	1.3536	0.9458	1.5384	1.4065	1.1023	1.9126	1.6503	2.1233
		GMBR	0.9369	0.2145	1.1969	0.709	1.3071	1.6986	0.8435	1.7288	1.4333	1.9174
	50	BR	0.9254	0.4362	1.3243	0.7431	1.4482	1.2045	0.6987	1.6148	1.3125	1.81
		GMBR	0.8322	0.1242	1.1551	0.6954	1.2258	1.3274	0.4981	1.3999	1.1004	1.5893
	100	BR	0.8958	0.2763	1.2052	0.7787	1.3757	1.0658	0.4626	1.3606	1.0678	1.5492
		GMBR	0.7526	0.0624	0.9924	0.5628	1.1626	1.0469	0.2497	1.1497	0.8505	1.3403
	250	BR	0.8865	0.2379	1.1691	0.7379	1.338	1.0045	0.3125	1.2149	0.9125	1.4007
		GMBR	0.7026	0.0249	0.9552	0.5252	1.1251	0.8741	0.0998	1.0004	0.6999	1.1891

Table (1) show The GMBR method yields better results than BR because it achieves lower MSE and MAE values in most cases. Increasing the sample size reduces errors, which improves the accuracy of the model. Increasing the level of contamination increases the error values, but GMBR remains less affected than BR.

Journal of Information Systems Engineering and Management
2025, 10(33s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/ **Research Article**

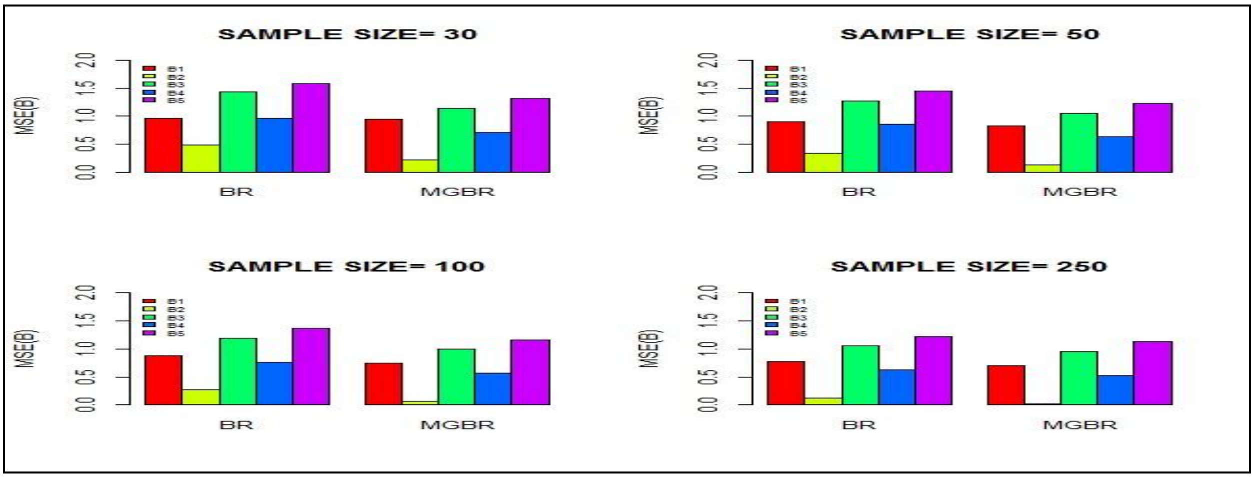

Figure (1): MSE ($\hat{\beta}$) for Different Sample Sizes at 5% Contamination in the First Scenario.

figure (1) show The GMBr method provides more accurate estimates and is less affected by contamination than the BR method. Increasing the sample size helps improve the accuracy of the estimate and reduce the MSE value for both methods, but GMBr remains the superior performer. Even at low contamination levels (5%), GMBr appears to be a more efficient method, making it a recommended choice when dealing with data that may contain anomalies or minor contamination.

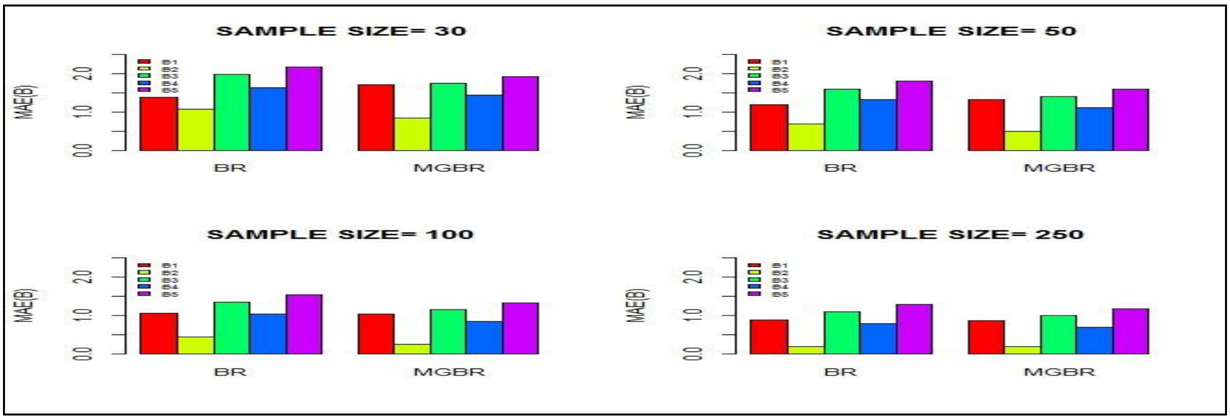

Figure (2): MAE ($\hat{\beta}$) for Different Sample Sizes at 5% Contamination in the First Scenario.

figure (2) show The GMBr method provides more accurate estimates and is less affected by contamination than the BR method .Increasing the sample size helps improve the accuracy of the estimate and reduces the MAE value for both methods, but GMBr remains the superior performer.

Even at low contamination levels (5%), GMBr appears to be a more efficient method, making it a recommended choice when dealing with data containing contamination or anomalies.

Journal of Information Systems Engineering and Management
2025, 10(33s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/ **Research Article**

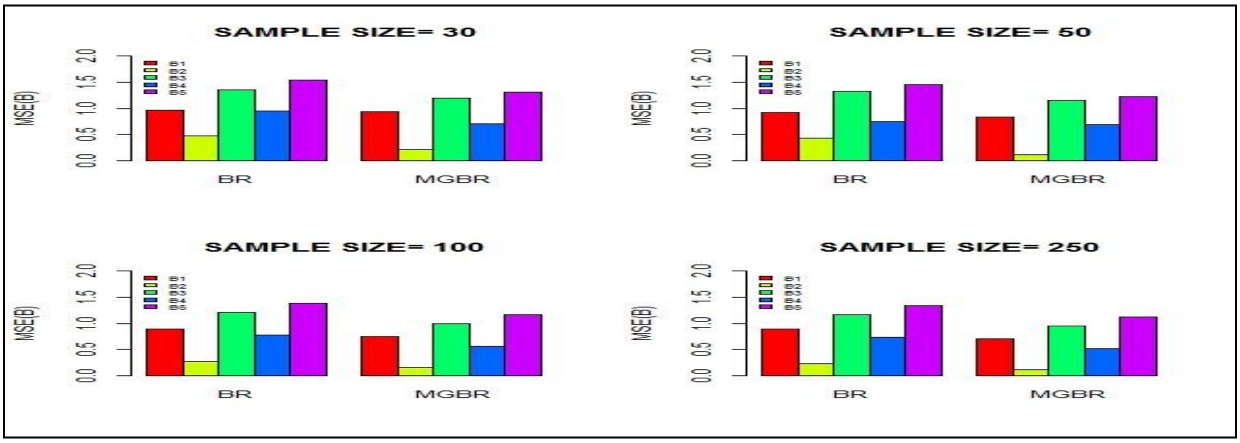

Figure (3): MSE ($\hat{\beta}$) for Different Sample Sizes at 20% Contamination in the First Scenario.

figure (3) show Increasing the contamination percentage (20%) significantly affects model performance, with higher MSE values for both methods, but the impact is greater for BR than for GMBr. GMBr provides more accurate estimates and is less susceptible to contamination, making it the preferred choice when dealing with contaminated data. As the sample size increases, MSE values decrease and the estimation accuracy improves, but GMBr still maintains an advantage over BR.

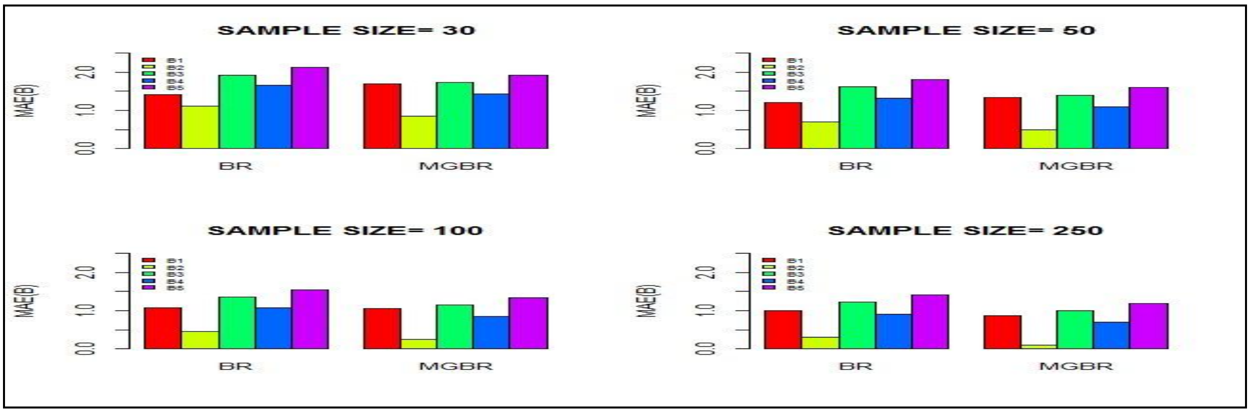

Figure (4): MAE ($\hat{\beta}$) for Different Sample Sizes at 20 % Contamination in the First Scenario.

figure (4) show Increasing the contamination percentage (20%) leads to higher errors in parameter estimation, making the model less accurate, but GMBr remains superior to BR. GMBr provides more stable estimates and is less affected by contamination, making it the preferred choice when dealing with contaminated data. As the sample size increases, MAE values decrease and estimation accuracy improves, but GMBr still maintains an advantage over BR, indicating that it is more robust against contaminated data.

Table (2): the MSE ($\hat{\beta}$) and MAE ($\hat{\beta}$) at Three Different Contamination Levels and different Sample Sizes in the First Scenario When the Precision Parameter is λ=4.

Cont.	Sample	methods	MSE ($\hat{\beta}$)					MAE ($\hat{\beta}$)				
			$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\beta}_5$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\beta}_5$
5%	30	BR	4.8451	2.4187	2.2479	1.3686	1.8454	5.9413	3.5971	3.5721	2.5427	3.0309

Journal of Information Systems Engineering and Management
2025, 10(33s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

		GMBR	1.2732	1.5053	1.2983	0.4251	0.945	2.906	2.7588	2.6607	1.6725	2.2102
	50	BR	3.2637	2.2034	2.0089	1.0988	1.6709	3.9395	2.9354	2.8624	1.798	2.4343
		GMBR	0.7617	1.3384	1.1251	0.2447	0.7852	1.7437	2.0983	1.9876	0.9864	1.5624
	100	BR	2.0652	2.0634	1.8421	0.9804	1.5099	2.4031	2.4458	2.319	1.3535	1.9086
		GMBR	0.375	1.2129	0.9945	0.1258	0.6558	0.8654	1.5994	1.4785	0.5011	1.0561
	250	BR	1.3213	1.9847	1.7673	0.898	1.4278	1.4622	2.1453	2.0264	1.0479	1.6024
		GMBR	0.1469	1.1379	0.92	0.0501	0.58	0.3427	1.2997	1.1799	0.2001	0.755
10%	30	BR	5.2973	3.0054	2.7945	1.9538	2.4225	6.4379	4.2203	4.1039	3.185	3.6212
		GMBR	1.2398	1.5081	1.2895	0.4267	0.9417	2.8672	2.7684	2.6461	1.6807	2.2057
	50	BR	3.9056	2.7798	2.5849	1.7351	2.2492	4.5584	3.5039	3.4317	2.4884	3.0117
		GMBR	0.753	1.3335	1.1204	0.2557	0.7809	1.7325	2.0895	1.9793	1.0097	1.5553
	100	BR	2.6781	2.6631	2.4456	1.5813	2.1091	3.0122	3.0441	2.926	1.9569	2.5065
		GMBR	0.3676	1.2126	0.9946	0.1259	0.6554	0.8569	1.5988	1.4789	0.5015	1.0556
	250	BR	1.9269	2.5883	2.3707	1.5001	2.0299	2.0667	2.75	2.6305	1.6494	2.2039
		GMBR	0.1435	1.1379	0.9201	0.05	0.5799	0.3387	1.2998	1.1801	0.1999	0.7548
20%	30	BR	5.7466	3.2556	3.0073	2.1314	2.6703	6.8156	4.5074	4.3087	3.3231	3.9012
		GMBR	1.2681	1.5144	1.2871	0.4144	0.9502	2.8998	2.7818	2.6421	1.6569	2.2236
	50	BR	4.1262	3.0395	2.809	1.927	2.4644	4.7811	3.8076	3.6635	2.6597	3.2286
		GMBR	0.7293	1.3412	1.1216	0.2478	0.7791	1.705	2.1056	1.9825	0.9949	1.5525
	100	BR	2.8935	2.8978	2.677	1.8039	2.3365	3.2317	3.2865	3.1611	2.1747	2.7353
		GMB	0.35	1.213	0.99	0.124	0.65	0.84	1.601	1.48	0.49	1.054

Journal of Information Systems Engineering and Management
2025, 10(33s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/ **Research Article**

		R	73	7	53	6	5	48	3	04	91	9
	250	BR	2.1611	2.8185	2.6008	1.7303	2.2593	2.2995	2.9806	2.8611	1.8802	2.4332
		GMBR	0.1417	1.1381	0.9201	0.05	0.5798	0.3366	1.3002	1.1802	0.2	0.7547

Table (2) show When using a higher precision factor (λ=4), the MSE and MAE values are larger than in the first table (λ=2), but the GMBR method still provides better estimates compared to BR. Increasing the sample size reduces the error, which means the model's accuracy in estimating the parameters improves. Increasing the level of contamination increases the MSE and MAE values, but GMBR remains less affected by contamination than BR. Compared to the first table, we see that the higher precision factor (λ=4) slightly increases the error, but does not change the overall trend of the results.

Table(3): the MSE ($\hat{\beta}$) and MAE ($\hat{\beta}$) at Three Different Contamination Levels and different Sample Sizes in the Second Scenario When the Precision Parameter is λ=2.

Cont.	Sample	methods	MSE ($\hat{\beta}$)							MAE ($\hat{\beta}$)						
			$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\beta}_5$	$\hat{\beta}_6$	$\hat{\beta}_7$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\beta}_5$	$\hat{\beta}_6$	$\hat{\beta}_7$
5 %	30	BR	9.0501	2.177	1.5706	2.7231	1.4948	2.4749	1.598	5.9205	2.8718	2.311	3.4587	2.2523	3.338	2.3386
		GMBR	6.5251	1.3913	0.8321	1.946	0.8203	1.7119	0.8382	4.6617	2.1937	1.6606	2.7627	1.6499	2.6508	1.6667
	50	BR	10.8816	2.2286	1.7063	2.2456	1.7175	2.6174	1.704	7.0422	3.1541	2.647	3.786	2.6678	3.674	2.6617
		GMBR	7.8243	1.5458	0.9998	2.1114	1.0004	1.8779	0.9988	5.5932	2.5182	1.9966	3.0966	1.9977	2.9849	1.9965
	100	BR	5.6941	1.7636	1.187	2.258	1.1836	2.0373	1.1798	2.6346	2.2279	1.6771	2.736	1.6731	2.636	1.6684
		GMBR	3.8863	1.0621	0.5014	1.5951	0.5013	1.3679	0.5002	2.7876	1.5366	1.0009	2.0826	1.7007	1.9773	0.9996
	250	BR	2.6999	1.4268	0.8691	1.9726	0.8705	1.7423	0.8725	1.9437	1.6004	1.0677	2.1593	1.0693	2.0508	1.0709
		GMBR	1.5553	0.7553	0.1994	1.3003	0.1999	1.0702	0.2005	1.1154	0.9302	0.3993	1.4882	0.3998	1.3801	0.4004
10	30	BR	17.3	3.4	2.8	3.9	2.8	3.7	2.9	11.	4.8	4.3	5.4	4.3	5.3	4.4

Journal of Information Systems Engineering and Management
2025, 10(33s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

%			64	022	053	484	482	287	245	3912	825	266	28	638	742	106
		GMBR	12.944	2.2277	1.667	2.7674	1.6746	2.5568	1.6886	9.2853	3.8609	3.3252	4.413	3.3334	4.3265	3.3459
	50	BR	11.0846	2.6536	2.0556	3.186	2.0587	2.8895	2.0554	7.3403	3.5882	3.0188	4.1295	3.017	3.9591	3.0166
		GMBR	7.7851	1.565	1.0012	2.1062	1.0033	1.8634	1.0005	5.5788	2.538	1.9992	3.0917	2.0011	2.9712	1.9985
	100	BR	6.1086	2.0598	1.523	2.6318	1.5349	2.3969	1.547	4.2044	2.5261	2.0156	3.1116	2.0276	2.9983	2.0394
		GMBR	3.8926	1.05	0.4991	1.6013	0.5022	1.3695	0.5056	2.7898	1.5246	0.9987	2.0889	1.0018	1.979	1.0052
	250	BR	3.0652	1.776	1.2215	2.3194	1.2213	2.0923	1.2201	2.2984	1.95	1.4204	2.5063	1.4203	2.4012	1.4191
		GMBR	1.5566	0.2	0.2002	1.2998	0.2	1.0704	0.1998	1.1159	0.9299	0.4001	1.4877	0.3999	1.3804	0.3997
20%	30	BR	17.4993	3.0613	3.1614	5.7413	3.9694	3.0351	4.5368	11.6157	6.0068	5.0281	4.6058	5.5305	7.3058	4.6866
		GMBR	12.89	1.6643	1.6966	4.3679	2.5704	1.6662	3.0803	9.2655	4.7247	3.6578	3.3293	4.2318	6.0279	3.3406
	50	BR	11.3147	2.3516	2.3371	5.007	3.1967	2.3874	3.6929	7.6024	4.7251	3.6091	3.2774	4.1706	6.0579	3.2588
		GMBR	7.7734	1.0061	1.0051	3.6954	1.8949	1.0145	2.3923	5.5745	3.4047	2.3035	1.9938	2.8935	4.713	1.9905
	100	BR	6.3397	1.7999	1.8106	4.5167	2.7052	1.8025	3.2018	4.4693	3.6949	2.6062	2.3097	3.1998	4.9963	2.2963
		GMBR	3.8755	0.499	0.5013	3.2024	1.401	0.499	1.8996	2.7837	2.3987	1.301	1.0021	1.9007	3.6987	0.9993
	250	BR	3.3305	1.5033	1.5018	4.1987	2.4024	1.5001	2.8987	2.5735	3.1025	2.001	1.6979	2.6016	4.3993	1.6978

Journal of Information Systems Engineering and Management
2025, 10(33s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/ **Research Article**

		GMBR	1.5472	0.2007	0.2002	2.8996	1.1003	0.2	1.5994	1.1125	1.8007	0.7002	0.3995	1.3003	3.1	0.3994

Table (3) show In the second scenario (when λ=2), the GMBR method still outperforms BR, achieving lower MSE and MAE values in most cases. Increasing the sample size significantly reduces the error, which means the model's accuracy in estimating parameters improves. Increasing the pollution level significantly increases the MSE and MAE values, but GMBR remains less affected by pollution than BR. Compared to the previous tables, the second scenario appears to contain more variables ($\hat{\beta}_1$ to $\hat{\beta}_7$), making estimation more complex, but GMBR still performs better than BR.

Table(4): the MSE ($\hat{\beta}$) and MAE ($\hat{\beta}$) at Three Different Contamination Levels and different Sample Sizes in the Second Scenario When the Precision Parameter is λ=4.

Cont.	Sample	methods	MSE ($\hat{\beta}$)							MAE ($\hat{\beta}$)						
			$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\beta}_5$	$\hat{\beta}_6$	$\hat{\beta}_7$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\beta}_5$	$\hat{\beta}_6$	$\hat{\beta}_7$
5 %	30	BR	6.1253	3.293	3.1901	3.6738	2.6699	3.5169	4.6596	7.1031	4.9069	4.9883	5.2091	4.1882	5.0554	5.9079
		GMBR	4.4687	2.3499	2.2728	2.7731	1.6912	2.5782	3.8458	5.4752	4.0983	4.2016	4.4209	3.3495	4.2656	5.2056
	50	BR	5.1788	2.5105	2.445	2.9485	1.8197	2.6665	4.028	5.0276	3.5596	3.6766	3.8993	2.7844	3.6696	4.6904
		GMBR	3.927	1.6753	1.6068	2.1077	1.0001	1.891	3.2012	3.3005	2.763	2.8747	3.0936	1.9981	2.9195	3.899
	100	BR	5.3028	1.9551	1.9002	2.3874	1.3098	2.1944	3.4936	4.372	2.5379	2.6625	2.8682	1.8013	2.7165	3.6858
		GMBR	3.9341	1.1684	1.1014	1.5982	0.5038	1.3998	2.6999	3.1412	1.758	1.8709	2.0858	1.0033	1.9293	2.8994
	250	BR	6.0317	1.7095	1.6357	2.1395	1.0396	1.9395	3.2421	3.0776	2.0483	2.1542	2.3761	1.2885	2.2183	3.1907
		GMBR	2.4664	0.9203	0.8494	1.3501	0.2501	1.1502	2.4505	1.5704	1.2602	1.3693	1.588	0.5	1.4301	2.4004
10 %	30	BR	5.5661	1.2244	1.1631	1.6899	0.5757	1.4582	2.7563	2.5905	1.5494	1.6657	1.9112	0.8078	1.7193	2.69
		GM	2.4	0.9	0.8	1.3	0.2	1.1	2.4	1.5	1.2	1.3	1.5	0.5	1.4	2.3

Journal of Information Systems Engineering and Management
2025, 10(33s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

		BR	639	179	5	566	53	485	487	693	57	69	936	019	274	977
	50	BR	6.3682	2.0251	1.9634	2.4563	1.3617	2.2487	3.5475	3.392	2.3582	2.477	2.6881	1.6036	2.5216	3.4901
		GMBR	2.4669	0.9203	0.8522	1.3505	0.2517	1.1487	2.4481	1.5704	1.26	1.3718	1.5881	0.5013	1.4283	2.3977
	100	BR	6.3656	2.0163	1.9493	2.459	1.3472	2.2503	3.5579	3.3928	2.3531	2.4664	2.6941	1.5944	2.5277	3.5051
		GMBR	2.4682	0.9188	0.8495	1.3519	0.2486	1.15	2.4518	1.571	1.2586	1.3693	1.5898	0.4985	1.4298	2.4016
	250	BR	6.3503	2.023	1.9483	2.4492	1.3534	2.2487	3.5463	3.3902	2.3619	2.4674	2.6863	1.6023	2.5277	3.4952
		GMBR	2.4636	0.9208	0.8497	1.3498	0.2509	1.1497	2.4492	1.5695	1.2607	1.3696	1.5878	0.5008	1.4297	2.3991
20%	30	BR	5.5923	1.2535	1.1835	1.679	0.5908	1.4735	2.7753	2.6108	1.5838	1.6947	1.9049	0.8284	1.7411	2.7144
		GMBR	2.4635	0.9213	0.8518	1.3497	0.2527	1.1486	2.4495	1.5691	1.2607	1.3713	1.587	0.502	1.4279	2.3988
	50	BR	5.7018	1.3515	1.2822	1.7888	0.6857	1.5886	2.8791	2.723	1.6872	1.7978	2.0227	0.9305	1.8641	2.824
		GMBR	2.4662	0.9195	0.8496	1.3512	0.2503	1.1514	2.449	1.5702	1.2593	1.3693	1.589	0.5	1.4311	2.3987
	100	BR	5.6783	1.353	1.2805	1.7888	0.6782	1.5816	2.881	2.7195	1.6908	1.7986	2.0246	0.9262	1.8597	2.829
		GMBR	2.4628	0.9204	0.8498	1.3517	0.2491	1.1499	2.45	1.5692	1.2602	1.3697	1.5896	0.499	1.4298	2.3998
	250	BR	6.1778	1.8646	1.7962	2.2942	1.1979	2.0933	3.3923	3.2312	2.2038	2.3155	2.5315	1.4471	2.3726	3.3415
		GMBR	2.4569	0.9202	0.8504	1.3501	0.2507	1.1495	2.4493	1.5674	1.2602	1.3703	1.5881	0.5007	1.4295	2.3992

Journal of Information Systems Engineering and Management
2025, 10(33s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/ **Research Article**

Table (4) show When using a higher precision factor (λ=4), the MSE and MAE values become larger than in Table 3 (λ=2), but GMBR still provides better estimates than BR. Increasing the sample size significantly reduces the error, which means the model's accuracy in estimating parameters improves. Increasing the level of contamination increases the MSE and MAE values, but GMBR remains less affected by contamination than BR. Compared to the previous tables, we see that a higher precision factor (λ=4) increases the error, but it does not change the overall trend of the results, as GMBR remains the most accurate and stable method.

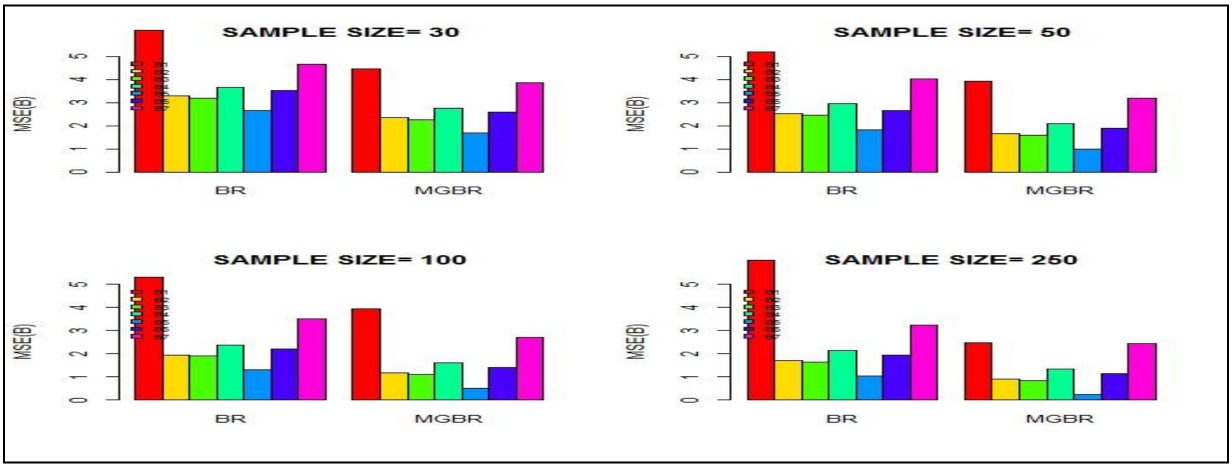

Figure (5): MSE ($\hat{\beta}$) for Different Sample Sizes at 5% Contamination in the Second Scenario.

figure (5) show The GMBr method provides more accurate estimates and is less susceptible to contamination than the BR method, even at low contamination levels (5%). Increasing the sample size significantly improves model accuracy and reduces the MSE value, but GMBr remains superior in all cases.The GMBr method is a better choice, especially when dealing with partially contaminated data or when using small samples.

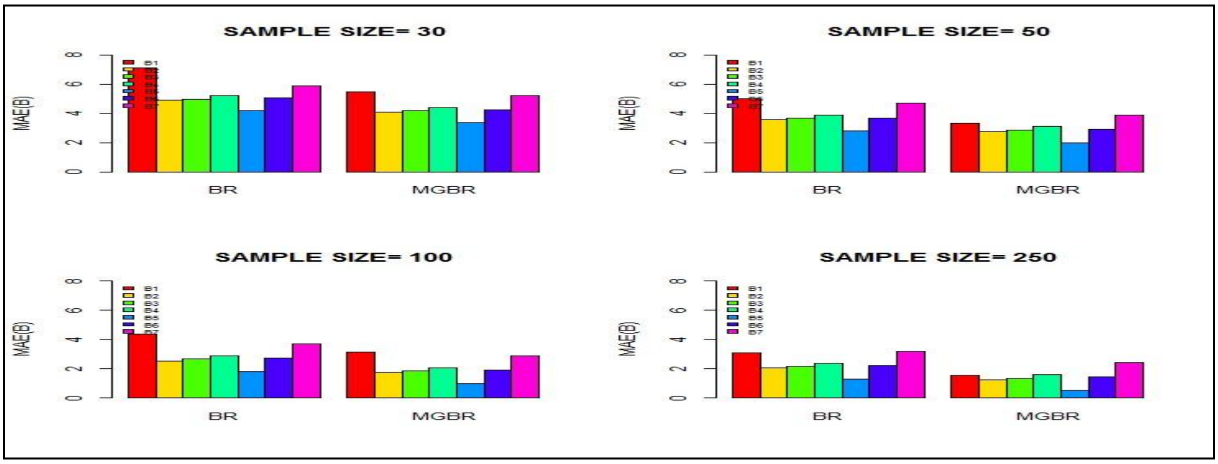

Figure (6): MAE ($\hat{\beta}$) for Different Sample Sizes at 5 % Contamination in the Second Scenario.

figure (6) show The GMBr method provides more accurate estimates and is less affected by contamination than the BR method, even at low contamination levels (5%). Increasing the sample size improves model accuracy and reduces the MAE value, but GMBr remains the best performer in all cases .The GMBr method is an ideal choice when stable and accurate estimates are needed, especially with partially contaminated data or small samples.

Journal of Information Systems Engineering and Management
2025, 10(33s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/ **Research Article**

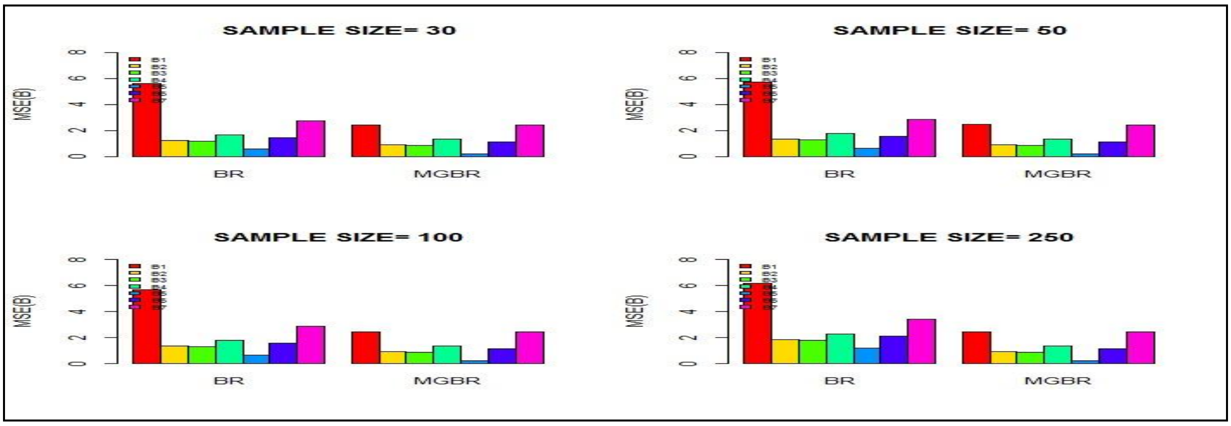

Figure (7): MSE ($\hat{\beta}$) for Different Sample Sizes at 20% Contamination in the Second Scenario.

figure (7) show Increasing the contamination percentage (20%) significantly affects model performance, with higher MSE values for both methods, but the effect is greater for BR than for GMBr . GMBr provides more accurate estimates and is less susceptible to contamination, making it the preferred choice when dealing with contaminated data. As the sample size increases, MSE values decrease and the estimation accuracy improves, but GMBr still maintains an advantage over BR, indicating that it is more robust against contaminated data.

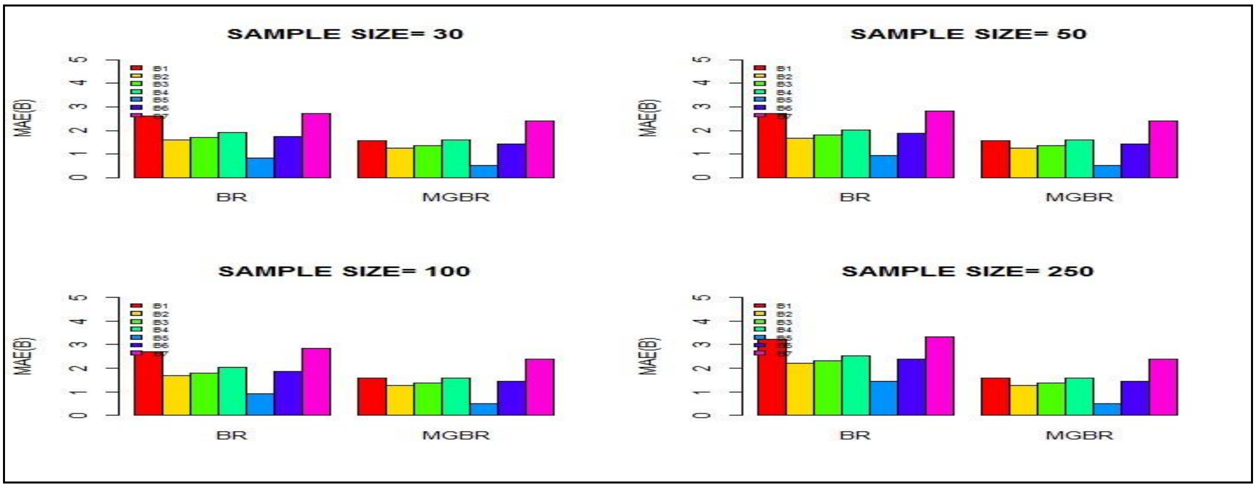

Figure (8): MAE ($\hat{\beta}$) for Different Sample Sizes at 20 % Contamination in the Second Scenario.

figure (8) show Increasing the contamination percentage (20%) significantly affects the estimation accuracy, as MAE values increase for both methods, but the effect is greater for BR than for GMBr.

GMBr provides more accurate estimates and is less susceptible to contamination, making it the preferred choice when dealing with contaminated data. As the sample size increases, MAE values decrease and estimation accuracy improves, but GMBr still maintains an advantage over BR, indicating that it is more robust against contaminated data.

Real Data

The Gasoline Yield dataset, originally gathered by Prater in 1956, focuses on the proportion of crude oil remaining after distillation and fractionation. Atkinson (1985) later analyzed the data using a linear regression model and observed that the error distribution appeared slightly asymmetrical, resulting in unusually large and small residuals. For our study, we introduced contamination by altering the first observations of the explanatory variables by 10%.

Journal of Information Systems Engineering and Management
2025, 10(33s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/ **Research Article**

Dataset Description

The Gasoline Yield dataset contains **32 observations** of **6 variables** related to gasoline reduction:

Variable Description

yield	Proportion of crude oil converted into gasoline (response variable)
gravity	API gravity of crude oil
pressure	Vapor pressure of crude oil
temp10	Temperature at which 10% of crude oil has vaporized
temp	Temperature at which 50% of crude oil has vaporized
temp90	Temperature at which 90% of crude oil has vaporized

Table(5): the coefficient estimated by BR and GMBR methods

methods	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\beta}_5$	$\hat{\beta}_6$
BR	-1.50568	1.69601	20.35790	-1.64932	2.87122	-23.13934
GMBR	-0.19659	0.58397	1.67494	-0.36112	0.40733	-2.28240

Table (5) show The GMBR method provides more stable and less extreme estimates than BR, making it more accurate and reliable. The BR method may be affected by outliers or contamination, leading to unstable or highly extreme estimates . GMBR is preferred when dealing with data that may contain significant noise or variance, as it provides more conservative and stable estimates.

Table (6): the MSE and MAE to the model that estimated by BR and GMBR.

	MSE	MAE
BR	0.30175	0.54891
MGBR	0.14614	0.29730

Table (6) show The GMBR method achieves lower MSE and MAE values compared to the BR method .This suggests that GMBR is more accurate and stable in model estimation.

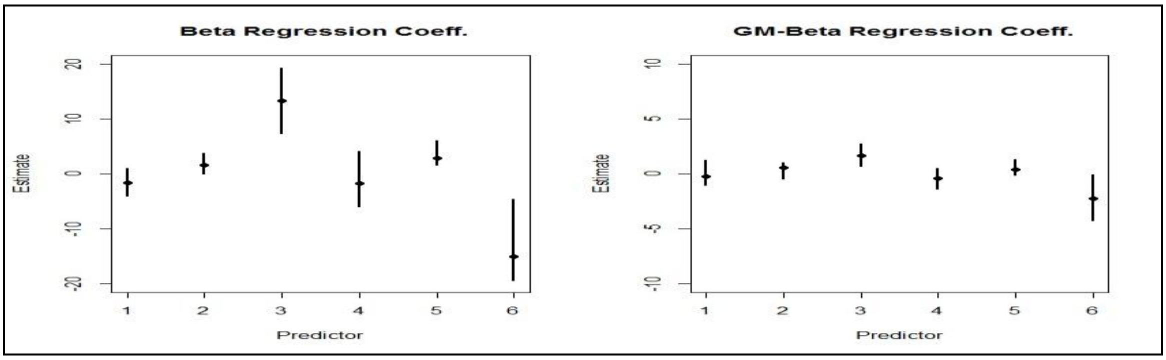

Figure (9): Confidence Intervals for the Estimated Coefficients Using BR and GMBR Methods.

figure (9) show The GMBr method achieves more stable estimates and is less susceptible to dispersion than the BR method, as demonstrated by the narrower confidence intervals. The BR method exhibits

Journal of Information Systems Engineering and Management
2025, 10(33s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/ **Research Article**

greater variance in parameter estimates, suggesting it may be less accurate, especially when dealing with data containing anomalies or contamination. The GMBr method is preferred when the goal is to obtain more reliable estimates that are less sensitive to contaminated data.

Conclusions:

This study examines beta regression analysis and its use in statistical modeling when the response values are between 0 and 1, focusing on high leverage points of parameter estimates. The maximum weighted estimation (MLE) method was used to estimate model parameters, but this method is sensitive to data anomalies, which can lead to skewed conclusions. Therefore, the generalized M-estimator (GM-estimator) method was proposed to reduce the influence of high leverage points and enhance model accuracy. The performance of the GMBr method was compared with the traditional beta regression (MLBr) method through simulation experiments and real data using criteria such as the mean squared error (MSE) and mean absolute error (MAE). The Generalized M-estimator (GMBr) method demonstrated superior performance over the traditional beta regression (MLBr) method, providing more accurate and stable estimates by reducing the mean squared error (MSE) and mean absolute error (MAE) across most sample sizes and contamination levels.

References:

[1] Abonazel, M. R., Algamal, Z. Y., Awwad, F. A., & Taha, I. M. (2022). A new two-parameter estimator for beta regression model: method, simulation, and application. *Frontiers in Applied Mathematics and Statistics*, *7*, 780322.

[2] Abonazel, M. R., Dawoud, I., Awwad, F. A., & Lukman, A. F. (2022). Dawoud–Kibria estimator for beta regression model: simulation and application. *Frontiers in Applied Mathematics and Statistics*, *8*, 775068.

[3] Abonazel, M. R., & Taha, I. M. (2023). Beta ridge regression estimators: simulation and application. *Communications in Statistics-Simulation and Computation*, *52*(9), 4280–4292.

[4] Algamal, Z. Y., Abonazel, M. R., & Lukman, A. F. (2023). Modified Jackknife ridge estimator for Beta regression model with application to chemical data. *International Journal of Mathematics, Statistics, and Computer Science*, *1*, 15–24.

[5] Bayer, F. M., & Cribari-Neto, F. (2017). Model selection criteria in beta regression with varying dispersion. *Communications in Statistics-Simulation and Computation*, *46*(1), 729–746.

[6] Bayes, C. L., Bazán, J. L., & García, C. (2012a). *A new robust regression model for proportions*.

[7] Bayes, C. L., Bazán, J. L., & García, C. (2012b). A new robust regression model for proportions. *Bayesian Analysis*, *7*(4), 841–866. https://doi.org/10.1214/12-BA728

[8] Cribari-Neto, F., & Zeileis, A. (2010). Beta regression in R. *Journal of Statistical Software*, *34*, 1–24.

[9] El-Raoof, Z. M. A., M El-Gohary, M., & G Yehia, E. (2023). Robust Estimation for Beta Regression Model in the Presence of Outliers: A Comparative Study. *التجارة والتمويل*, *43*(3), 380–406.

[10] Espinheira, P. L., da Silva, L. C. M., & Silva, A. de O. (2015). Prediction measures in beta regression models. *ArXiv Preprint ArXiv:1501.04830*.

[11] Espinheira, P. L., da Silva, L. C. M., Silva, A. de O., & Ospina, R. (2019). Model selection criteria on beta regression for machine learning. *Machine Learning and Knowledge Extraction*, *1*(1), 26.

[12] Espinheira, P. L., Ferrari, S. L. P., & Cribari-Neto, F. (2008). On beta regression residuals. *Journal of Applied Statistics*, *35*(4), 407–419.

[13] Ferrari, S., & Cribari-Neto, F. (2004). Beta regression for modelling rates and proportions. *Journal of Applied Statistics*, *31*(7), 799–815.

[14] Ghosh, A. (2019). Robust inference under the beta regression model with application to health care studies. *Statistical Methods in Medical Research*, *28*(3), 871–888. https://doi.org/10.1177/0962280217738142

[15] Heng, Q., & Lange, K. (2025). Bootstrap estimation of the proportion of outliers in robust regression. *Statistics and Computing*, *35*(1), 3.

[16] Hoaglin, D. C., & Welsch, R. E. (1978). The hat matrix in regression and ANOVA. *The American Statistician*, *32*(1), 17–22.

[17] Iglewicz, B., & Hoaglin, D. C. (1993). *Volume 16: how to detect and handle outliers*. Quality Press.

[18] Karlsson, P., Månsson, K., & Kibria, B. M. G. (2020). A Liu estimator for the beta regression model and its application to chemical data. *Journal of Chemometrics*, *34*(10), e3300.

[19] Lim, H. A., & Midi, H. (2016). Diagnostic Robust Generalized Potential based on Index Set Equality (DRGP (ISE)) for the identification of high leverage points in linear model. *Computational Statistics*, *31*, 859–877.

[20] Maluf, Y. S., Ferrari, S. L. P., & Queiroz, F. F. (2024). Robust beta regression through the logit transformation. *Metrika*, 1–21.

[21] Midi, H., Ismaeel, S. S., Arasan, J., & AMohammed, M. (2021). Simple and fast generalized-M (GM) estimator and its application to real data set. *Sains Malaysiana*, *50*(3), 859–867.

[22] Ospina, R., Baltazar, S. G. F., Leiva, V., Figueroa-Zúñiga, J., & Castro, C. (2023). Robust Semi-Parametric Inference for Two-Stage Production Models: A Beta Regression Approach. *Symmetry*, *15*(7), 1362.

[23] Ospina, R., Cribari-Neto, F., & Vasconcellos, K. L. P. (2006). Improved point and interval estimation for a beta regression model. *Computational Statistics & Data Analysis*, *51*(2), 960–981.

[24] PJ, P. J. R., & Leroy, A. M. (1987). *Robust regression and outlier detection*. John Wiley & Sons.

[25] Ribeiro, T. K. A., & Ferrari, S. L. P. (2023). Robust estimation in beta regression via maximum L q-likelihood. *Statistical Papers*, *64*(1), 321–353.

[26] Rosseeuw, P. J., & Van Zomeren, B. C. (1990). Unmasking multivariate outliers and leverate points. *Journal of the American Statistical Association*, *85*, 633–639.

[27] Rousseeuw, P. J., & Leroy, A. M. (2003). *Robust regression and outlier detection*. John wiley & sons.

[28] Smithson, M., & Verkuilen, J. (2006). A better lemon squeezer? Maximum-likelihood regression with beta-distributed dependent variables. *Psychological Methods*, *11*(1), 54.

[29] Vasconcellos, K. L. P., & Cribari-Neto, F. (2005). Improved maximum likelihood estimation in a new class of beta regression models. *Brazilian Journal of Probability and Statistics*, 13–31.

[30] Velleman, P. F., & Welsch, R. E. (1981). Efficient computing of regression diagnostics. *The American Statistician*, *35*(4), 234–242.

[31] Wilcox, R. R. (2011). *Introduction to robust estimation and hypothesis testing*. Academic press.

[32] Zhou, H., & Huang, X. (2022). Bayesian beta regression for bounded responses with unknown supports. *Computational Statistics & Data Analysis*, *167*, 107345.