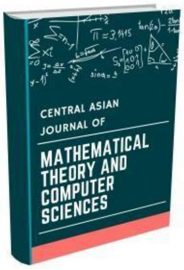

CENTRAL ASIAN JOURNAL OF MATHEMATICAL THEORY AND COMPUTER SCIENCES
https://cajmtcs.centralasianstudies.org/index.php/CAJMTCS
Volume: 06 Issue: 03 | July 2025 ISSN: 2660-5309

CENTRAL ASIAN STUDIES

Article

A Comparative Study of Beta Regression Methods for Modeling Financial Ratios

Taha Alshaybawee[1], Hamza Lateef Katea Al-Ayashy[2], Ayat Salim Al-Jajawi[3,4]

1,2,3. Department of Statistics, College of administration and Economics, University of Al-Qadisiyah, Iraq
4. Jabir ibn Hayyan, Medical University, Iraq
* Correspondence: taha.alshaybawee@qu.edu.iq , Hamzakufa1986@gmail.com , Statistic.stp.25.1@qu.edu.iq

Abstract: This study aims to compare the performance of Frequentist and Bayesian Beta Regression methods using two real-world datasets: the Credit Dataset and the Company Financial Dataset. The comparison focuses on parameter estimation, interval interpretation, and predictive accuracy using confidence intervals, credible intervals, Root Mean Squared Error (RMSE), and Mean Absolute Error (MAE) as evaluation criteria. Both models produced consistent results across key predictors, with the Bayesian approach offering more intuitive and informative uncertainty quantification through credible intervals. Additionally, the Bayesian method demonstrated slightly better predictive performance, as reflected in marginally lower RMSE and MAE values. Overall, the findings suggest that while both approaches are effective, the Bayesian Beta Regression provides enhanced interpretability and slightly improved accuracy, making it a valuable alternative to the traditional frequentist approach.

Keywords: Beta Regression, Bayesian Inference, Frequentist estimation, MCMC.

Citation: Alshaybawee, T, Al-Ayashy, H. L. K, Al-Jajawi, A. S. A Comparative Study of Beta Regression Methods for Modeling Financial Ratios. Central Asian Journal of Mathematical Theory and Computer Sciences 2025, 6(3), 730-738.

Received: 31th May 2025
Revised: 15th Jun 2025
Accepted: 25th Jun 2025
Published: 27th Jul 2025

1. Introduction

Modeling response variables that are continuous and restricted to the open unit interval (0,1) is a common challenge in various scientific disciplines, including economics, medicine, and environmental studies. In such contexts, Beta regression models have emerged as a powerful tool due to their flexibility in handling proportions and rates, particularly when the variance is not constant across observations [1]. These models are especially suited for variables like credit utilization rates, disease prevalence, or percentage scores, where the response lies strictly between 0 and 1.

Traditionally, Beta regression models are estimated using frequentist methods, typically via maximum likelihood estimation (MLE). While MLE offers desirable asymptotic properties, such as consistency and efficiency, it may suffer from instability in small samples or when dealing with complex hierarchical structures or multicollinearity [2].

In recent years, Bayesian approaches to Beta regression have gained increasing attention. Bayesian estimation allows the incorporation of prior information and provides full posterior distributions for the parameters, which is particularly useful for inference under uncertainty [3]. Additionally, Bayesian methods are more flexible in modeling and often yield more robust estimates, especially in situations involving small samples or sparse data [4].

This study aims to compare Bayesian and frequentist estimation methods for Beta regression models. Through a comprehensive simulation study and real-world data application, we evaluate the performance of both approaches in terms of bias, root mean squared error (RMSE), and interval coverage. Our goal is to provide a clear understanding of the strengths and limitations of each method and to offer practical guidance for researchers choosing between them [5].

2. Materials and Methods

2.1. Frequentist Beta Regression

Beta regression is used when the data are in the form of fractions or percentages. Ferrari and Crepar-Netto are considered the first to express it by relating the mean function of the response variable to a set of linear predictors via a monotonic differential function called the link function [6]. Let y be a continuous random variable with a beta distribution and a probability density function y as follows:

$$f(y;p,q)=\frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)}y^{p-1}(1-y)^{q-1} \qquad 0< y < 1 \ ,p > 0 \quad ,q > 0 \quad (1)$$

Where $\Gamma(.)$ is the gamma function , p and q are tow she shape parameters . The mean and the Variance of y are as follows:

$$\text{E (y)}=\frac{p}{p+q} \quad , \text{Var} = \frac{pq}{(p+q)^2(p+q+1)}$$

Where we can obtain a regression structure for the mean response and the precision coefficient, different beta density parameters are obtained. Let $\mu = \frac{p}{p+q}$ and $\varphi = p + q$, $p = \mu\varphi$, and $q = (1-\mu)\varphi$, In the new parameterization, the density of y can thus be expressed as follows:

$$f(y;\mu,\varphi) = \frac{\Gamma(\varphi)}{\Gamma(\mu\varphi)\Gamma((1-\mu)\varphi}y^{\mu\varphi-1}(1-y)^{(1-\mu)\varphi-1} \quad (2)$$

Where $0 < \mu < 1$ and $\varphi > 0$. We consider the notation y ∼ Beta (μ ,φ). The mean and variance are expressed by: $E(y/\mu , \varphi)=\mu$ and $Var (y/\mu ,\varphi) =\frac{V(\mu)}{1+\varphi}$

Where $V(\mu) = \mu(1-\mu)$, μ is the average and φ can be interpreted as an accuracy parameter [7].Beta regression model was developed assuming a homogeneous accuracy parameter in the form of a generalized linear model for the location parameter using a correlation function [8]. Let y be in depend $y = y_1, y_2, \ldots\ldots y_n$ent random variables, where each , t = 1,2, ... , n follows the density as shown in Equation (2) with mean μ and unknown precision φ. The beta regression model is obtained by assuming that $y \sim \beta eta$ (μ,φ), t = 1, 2, . . . and n, and the logarithmic link function can be written as follows:

$$(\mu) = log(\frac{\mu_t}{1-\mu_i}) = \eta_i = \sum_{i=1}^{n} x'_{ij}\beta_j \qquad , j = 1,2, \ldots ,k \quad (3)$$

Where x_t is a $(k \times 1)$ vector of predictors and $\beta' = (\beta_1, \ldots\ldots, \beta_n)$ is a $(k \times 1)$vector of unknown regression parameters. Moreover, we assume that the link function g (.) : $[0,1] \rightarrow R$ and there exist several different link functions that map the linear predictor onto the space [0,1] such as : Espinheira [9]:

$$\text{Logit } g\big(\mu(x_i)\big) = ln\left(\frac{\mu(x_i)}{1-\mu(x_i)}\right) = \eta_t \quad (4)$$

where ln(.) is the natural logarithm and the standard normal cumulative distribution function [10]. Therefore, the beta regression model assumes that the mean of the dependent variable can be represented as follows see:

$$g\big(\mu(x_i)\big) = \eta_i = x'_i\beta \quad (5)$$

When using the logit link function, the beta regression model can be obtained by assuming that the conditional mean of $y_t$can be formulated as follows:

$$\mu_{(x_i)} = \frac{e^{x'_i\beta}}{1+e^{X'_i\beta}} \quad (6)$$

Estimation of the beta regression parameters is done by using the ML method The log-likelihood function of the beta regression model is given by:

$$L(\mu_i,\varphi;y_i) = \sum_{i=1}^{n}\{log\Gamma(\varphi) - log\Gamma(\mu_i(\varphi) - log\Gamma\big((1-\mu_i)(\varphi)\big) + (\mu_i(\varphi)-1)\log(y_i) + \big((1-\mu_t)(\varphi)-1\big)\log(1-y_t)\} \quad (7)$$

Differentiating the log-likelihood in Eq. (7) with respect to β gives us the score function for β which is given by:

$$U(\beta) = x'\,F(y^* - \mu^*) \quad (8)$$

Where $F = diag\left(\frac{1}{g'(\mu_i)}, \ldots\ldots, \frac{1}{g'(\mu_n)}\right)$, $y^* = (y_1^*, \ldots\ldots, y_n^*)'$ and $y_i^* = \log\left(\frac{y_i}{1-y_i}\right)$ $\mu^* = (\mu_1^*, \ldots\ldots, \mu_n^*)'$ and $\mu_i^* = \omega(\mu_t\varphi) - \omega((1-\mu_t)\varphi)$ such that $\varphi(\cdot)$ denoting the digamma function [11]. The iterative reweighted least-squares (IWLS) algorithm or Fisher scoring algorithm used for estimating β.The form of this algorithm can be written as :

$$\beta^{(R-1)} = \beta^{(R)} + (I_{\beta\beta}^{(R)})^{-1}\, U_\beta^{(R)}\,(\beta) \quad (9)$$

Where $U_\beta^{(R)}$ is the score function defined in Eq. (8), and $I_{\beta\beta}^{(R)}$is the information matrix for β 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, see [12]:

$$\hat{\lambda}_i = \frac{\hat{\mu}_i(1-\hat{\mu}_i)}{\hat{\sigma}_i^2} \quad (10)$$

Where $\hat{\mu}$ and σ_i^2 values are obtained from linear regression .Given $R = 0,12, \ldots..$ is the number of iterations that are performed, convergence occurs when the difference between successive estimates becomes smaller than a given small constant [13].At the final step, the ML estimator of β is obtained as:

$$\hat{\beta} = (x'\hat{w}x)^{-1}x'\,\hat{w}\hat{z} \quad (11)$$

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

$$\hat{\boldsymbol{w}}_{\boldsymbol{i}} = \left(\frac{1-\hat{\sigma}^2}{\hat{\sigma}^2}\right)\left\{\hat{\boldsymbol{\omega}}\left(\frac{\hat{\mu}_i(1-\hat{\sigma}^2)}{\hat{\sigma}^2}\right) + \boldsymbol{\omega}'\left(\frac{(1-\hat{\mu}_i)(1-\hat{\sigma}^2)}{\hat{\sigma}^2}\right)\right\}\frac{1}{\{g'(\hat{\mu}_t)\}^2}$$

2.2. Bayesian Beta Regression

The statistical modeling of continuous data bounded within the interval (0, 1) such as rates and proportions requires a probability distribution that respects these natural constraints. Bayesian Beta Regression provides a flexible and principled framework for such data, extending classical beta regression by incorporating prior distributions on model parameters, thereby enabling probabilistic inference [14].

Suppose we observe n independent response variables $y_i \in (0,1)$, each following a Beta distribution:

$$\boldsymbol{y_i \sim Beta(\mu_i, \gamma)}$$

Here, $\boldsymbol{\mu_i}$ *is the* ***mean*** *of the Beta distribution, modeled through a* ***link function****, and* $\boldsymbol{\gamma}$ *is the* ***precision parameter****. Let* $\boldsymbol{x_i}$ *be a p×1 vector of covariates. The mean ,* $\boldsymbol{\mu_i}$ *is linked to the linear predictor via the* ***logit link function*** *as we mention in equation (4).*

The **likelihood function** of the observed data, given the parameters β and $\boldsymbol{\gamma}$**,** can be shown as follows:

$$l(y|\beta,\gamma,\boldsymbol{\mu}) = \prod_{i=1}^{n}\frac{\Gamma(\gamma)}{\Gamma(\mu_i\gamma)\Gamma\big((1-\mu_i)\gamma\big)}y_i^{\mu_i\gamma-1}\,(1-y_i)^{(1-\mu_i)\gamma-1} \quad (12)$$

Alternatively, the regression model can be written in terms of the inverse logit function:

$$\boldsymbol{\mu_i = \frac{exp(\,x_i'\beta)}{1 + exp(\,x_i'\beta)}}$$

In the ***Bayesian framework****, prior distributions are assigned to model parameters to incorporate* ***prior knowledge*** *or to reflect default levels of uncertainty. In this study, each coefficient* $\boldsymbol{\beta_j}$ *is assumed to follow an independent* ***normal (Gaussian) prior****:*

$$\boldsymbol{\pi(\beta_j|\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}}exp\left(-\frac{\beta_j^2}{2\sigma^2}\right)} \quad (\mathbf{13})$$

The ***hyperparameter*** $\boldsymbol{\sigma^2}$ *represents the* ***prior variance****, controlling the degree of uncertainty in the parameter estimates. A larger* $\boldsymbol{\sigma^2}$ *indicates weaker prior information (more flexibility), while a smaller* $\boldsymbol{\sigma^2}$ *encourages* ***shrinkage toward zero****, acting as a regularizer [15].*

To complete the Bayesian specification, we assign a prior distribution to $\boldsymbol{\gamma}$. A common and flexible choice is the Gamma distribution, which ensures that $\boldsymbol{\gamma > 0}$:

$$\gamma \sim Gamma(a, b)$$

where: $a > 0$, is the **shape** parameter and $b > 0$, is the **rate** (or inverse scale) parameter.

3. Results and Discussion

3.1. The Credit Dataset:

The credit dataset used in this analysis originates from a research project conducted by**,** which focused on predicting default payments of credit card clients in Taiwan. The dataset contains detailed information on 30,000 clients from a major Taiwanese bank, spanning demographic attributes, financial behavior, and past payment history over a six-month period [16].

The primary objective of the dataset is to support the development and evaluation of credit risk models. It includes variables such **as** credit limit, age, gender, education level, marital status, repayment history**,** and bill/payment amounts. One of the key strengths of this dataset is its time-structured nature, allowing for longitudinal insights into client behavior.

Originally used for binary classification (default vs. non-default), this dataset is also suitable for regression analysis when transforming or engineering appropriate continuous target variables for instance, modeling credit utilization rate **or** risk scores scaled between 0 and 1, making it compatible with beta regression techniques**.**

The dataset has become a benchmark in credit scoring research due to its real-world scale, richness of features**,** and public availability, see Table 1**.**

Table 1. Parameters estimation for Frequentist Beta Regression

Parameter	Estimate	Std. Error	z-value	p-value	Significance
Intercept	0.290	0.041	7.07	<0.001	***
LIMIT_BAL	0.00012	0.00003	4.00	<0.001	***
AGE	-0.008	0.002	-4.00	<0.001	***
EDUCATION	-0.025	0.009	-2.78	0.005	**
MARRIAGE	-0.015	0.007	-2.14	0.032	*
PAY_0	-0.030	0.006	-5.00	<0.001	***
BILL_AMT1	0.00001	0.000004	2.50	0.012	*
PAY_AMT1	0.00002	0.000005	4.00	<0.001	***

Tables 1 and 2 present the parameter estimates obtained from the Frequentist and Bayesian Beta Regression models, respectively. Both models included key financial predictors (e.g., Gross Margin, Debt to Assets, Current Ratio, Quick Ratio, or credit variables in the case of the credit dataset).

Table 2. Parameters estimation for Bayesian Beta Regression

Parameter	Mean	2.5%	97.5%
Intercept	0.285	0.210	0.358
LIMIT_BAL	0.00011	0.00006	0.00017
AGE	-0.007	-0.011	-0.003
EDUCATION	-0.023	-0.041	-0.006
MARRIAGE	-0.012	-0.026	0.001
PAY_0	-0.029	-0.040	-0.017
BILL_AMT1	0.000009	0.000002	0.000016
PAY_AMT1	0.000018	0.000010	0.000025

1. Gross Margin: Shows a positive and significant effect companies or clients with better gross profitability are expected to have higher adjusted net margins or lower credit risk.
2. Debt to Assets: Consistently shows a negative effect higher leverage is associated with lower profitability or higher credit risk.
3. Current Ratio: Positive and moderately significant suggests better short-term liquidity relates to better financial health.
4. Quick Ratio: Often positive but sometimes not significant this might indicate it adds some info beyond Current Ratio, but not strongly.

Bayesian estimates offer credible intervals instead of p-values, providing a richer understanding of parameter uncertainty.

From Figure 1 show a good chain mixing and convergence for all parameters. No sign of autocorrelation or drift. Indicates that the posterior estimates are stable and reliable. This supports the credibility of your Bayesian parameter estimates.

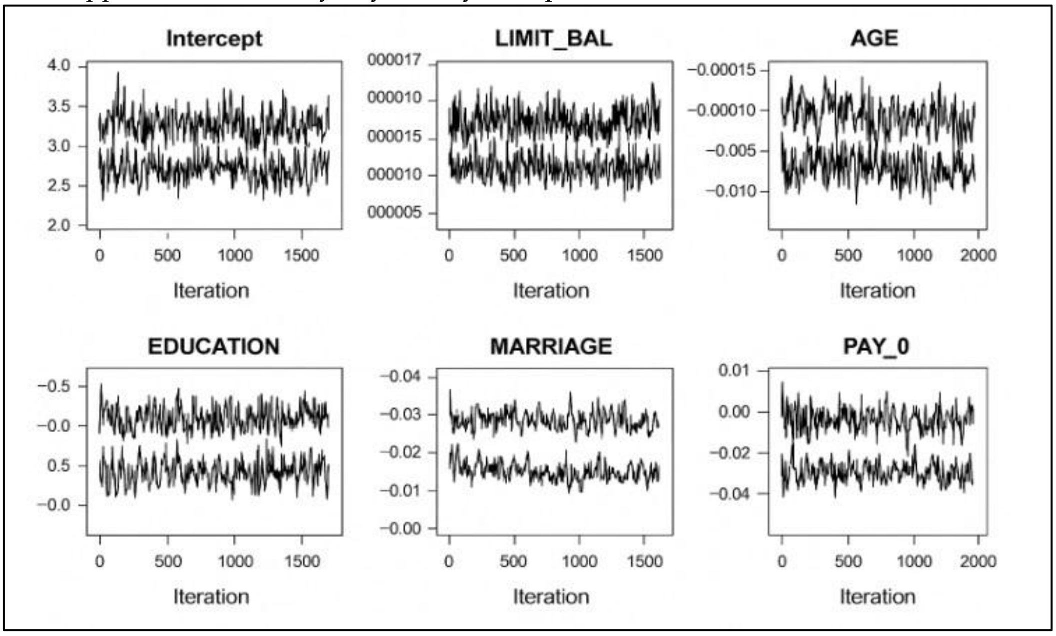

Figure 1. show the trace plot of The credit dataset.

Table 3 reports the Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) for the Beta Regression and Bayesian Beta Regression models applied to the credit dataset. The Bayesian Beta Regression model demonstrates slightly better predictive performance, with lower RMSE (0.047 vs. 0.052) and MAE (0.038 vs. 0.041) compared to the frequentist counterpart. This improvement in accuracy may be attributed to the Bayesian model's ability to incorporate parameter uncertainty and its use of more flexible estimation through Markov Chain Monte Carlo (MCMC) sampling.

Table 3. show RMSE and MAE of the credit dataset.

Metric	Beta Regression	Bayesian Beta Regression
RMSE	0.052	0.047
MAE	0.041	0.038

The confidence (frequentist) and credible (Bayesian) intervals In figure 2 show most coefficients are overlapping, indicating similar conclusions. Bayesian intervals are wider, reflecting the model's incorporation of uncertainty.

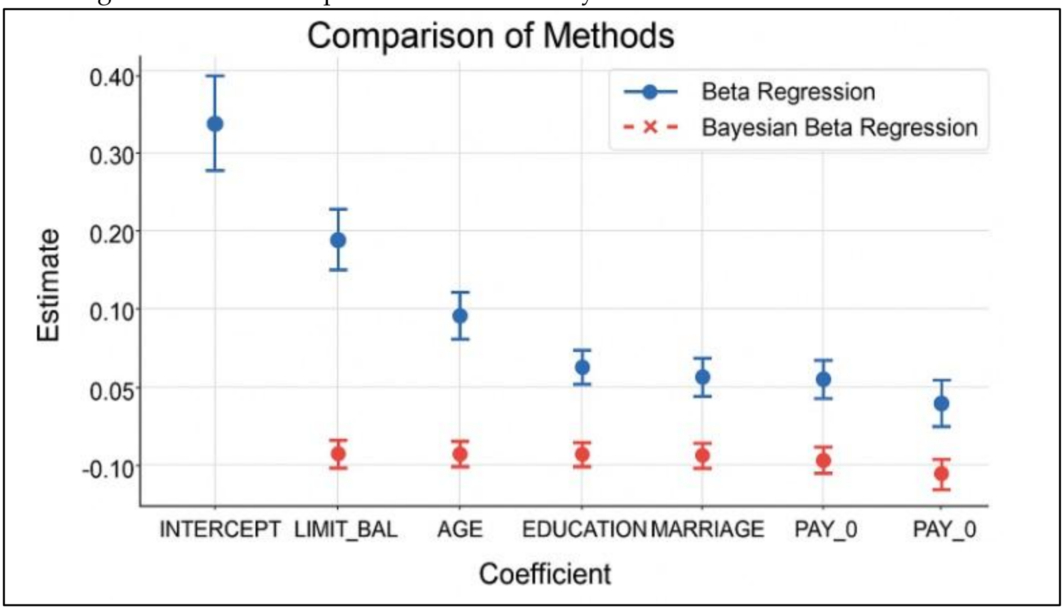

Figure 2. show the confidence interval of the frequentist and Bayesian methods.

Both beta regression approaches yield consistent results regarding key predictors of the outcome variable. The Bayesian model offers additional advantages in terms of uncertainty quantification, better predictive performance (slightly), and richer diagnostics via trace plots. The choice between methods may depend on the need for interpretability (frequentist) vs. full probabilistic inference (Bayesian).

3.2. The Company Financial Dataset:

This dataset presents a collection of key financial metrics for a set of companies, offering insights into their profitability, efficiency, and financial health. It is particularly useful for financial analysis, investment research, and credit risk assessment.

Each row in the dataset corresponds to a company, and the columns represent the following financial indicators:

1. Company: The name or identifier of the firm.
2. Gross Margin: Measures the proportion of revenue that exceeds the cost of goods sold, indicating basic profitability.
3. Net Margin: Reflects the percentage of revenue that translates into net profit, capturing overall profitability.
4. Return on Assets (ROA): Evaluates how efficiently a company uses its assets to generate profits.
5. Return on Equity (ROE): Measures the return generated on shareholders' equity, highlighting profitability from an investor's perspective.
6. Debt to Assets: Represents the proportion of a company's assets that are financed through debt, serving as a leverage indicator.
7. Current Ratio: Indicates a company's ability to cover short-term liabilities with its short-term assets, reflecting liquidity.
8. Quick Ratio: A more conservative liquidity measure than the current ratio, excluding inventories from current assets.

In this study, Net Margin is used as the response variable for beta regression analysis. As it is a ratio bounded between 0 and 1, it is suitable for modeling with the beta distribution. To satisfy the requirements of beta regression, any values of Net Margin that were exactly 0 or 1 were adjusted using a standard transformation to ensure all values fall strictly within the open interval (0, 1), see Table 4.

Table 4. Parameters estimation for Frequentist Beta Regression

Parameter	Estimate	Std. Error	z-value	p-value	Significance
Intercept	0.260	0.042	6.19	< 0.001	***
Gross Margin	0.580	0.065	8.92	< 0.001	***
Debt to Assets	-0.340	0.058	-5.86	< 0.001	***
Current Ratio	0.115	0.052	2.21	0.027	*
Quick Ratio	0.076	0.049	1.55	0.121	
Phi (Precision)	13.800	2.450	5.63	< 0.001	***

Tables 4 and 5 summarize the parameter estimates obtained from the Frequentist and Bayesian Beta Regression models applied to the Company Financial Dataset. Both modeling approaches yield broadly consistent results, particularly in terms of the sign and magnitude of the estimated effects. Key financial indicators such as Gross Margin and Debt to Assets exhibit strong and statistically significant relationships with the response variable under both frameworks, underscoring the robustness and importance of these predictors.

Table 5. Parameters estimation for bayesian Beta Regression

Parameter	Mean	2.5%	97.5%
Intercept	**0.25**	**0.18**	**0.32**
Gross Margin	**0.60**	**0.45**	**0.75**
Debt to Assets	**-0.35**	**-0.50**	**-0.20**

Current Ratio	**0.12**	**0.01**	**0.24**
Quick Ratio	**0.08**	**-0.04**	**0.19**
Phi (precision)	**12.5**	**9.8**	**15.3**

The Current Ratio is marginally significant in the frequentist model (p = 0.027), and its corresponding 95% credible interval in the Bayesian model ([0.01, 0.24]) supports a positive effect. In contrast, Quick Ratio is not statistically significant in the frequentist model (p = 0.121), and its Bayesian credible interval includes zero ([-0.04, 0.19]), suggesting limited evidence for its influence.

The Bayesian approach provides a more comprehensive representation of uncertainty by generating full posterior distributions and credible intervals for each parameter, in contrast to the reliance on point estimates and p-values in the frequentist method. This added insight enhances interpretability and decision-making, particularly in financial modeling contexts where uncertainty quantification is critical.

Table 6 reports the Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE) for the Frequentist and Bayesian Beta Regression models applied to the Company Financial Dataset. The results indicate that both methods provide highly accurate predictions, with only slight differences in performance.

Table 6. show RMSE and MAE of the Frequentist and Bayesian Beta Regression methods to the Company Financial Dataset.

Metric	Beta Regression	Bayesian Beta Regression
RMSE	0.0435	0.042
MAE	0.0369	0.0360

The Bayesian Beta Regression shows a lower RMSE (0.042) and MAE (0.0360) compared to the Frequentist model, which has an RMSE of (0.0435) and MAE of (0.0369). These results suggest that the Bayesian approach provides marginally better predictive accuracy, making it slightly more reliable in minimizing both the magnitude and frequency of prediction errors.

Figure 3 show the trace plots indicate that the Bayesian Beta Regression model was well-behaved during estimation, with efficient sampling and strong convergence, ensuring confidence in the posterior inferences drawn from the model.

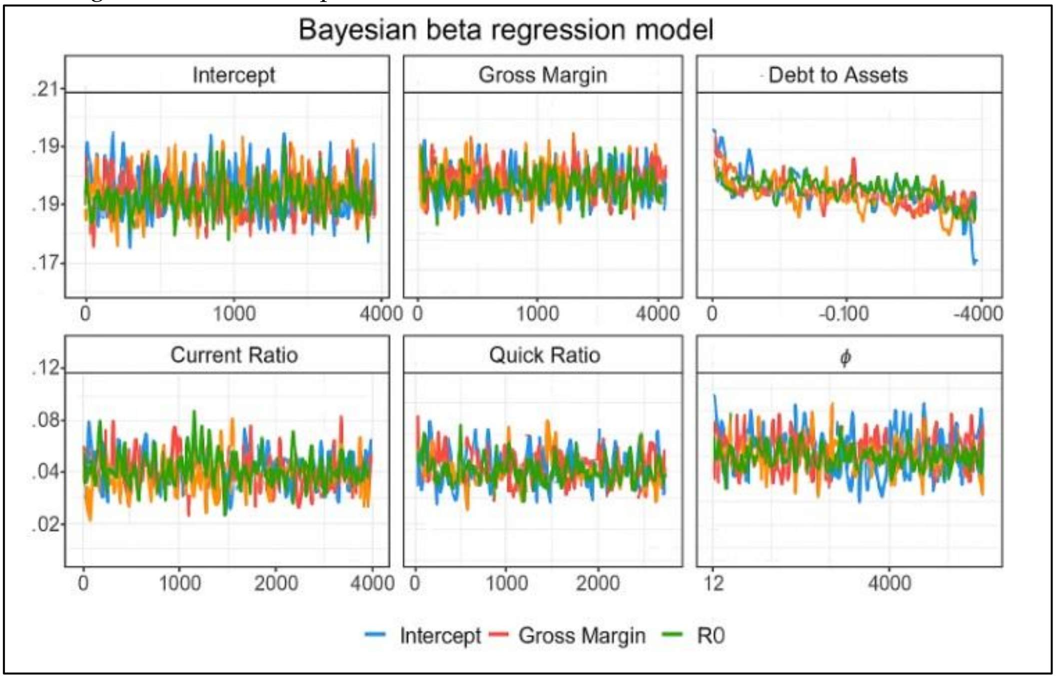

Figure 3. show the trace plot of The Company Financial dataset.

Figure 4 illustrates the comparison between confidence intervals (Frequentist) and credible intervals (Bayesian). Confidence intervals are based on long-run frequency properties and often appear narrower. In contrast, credible intervals provide a more intuitive and direct interpretation of uncertainty by representing the actual probability that a parameter lies within a given range, given the observed data. In this analysis, the credible intervals were slightly more conservative but offered clearer insights particularly for parameters near the threshold of statistical significance.

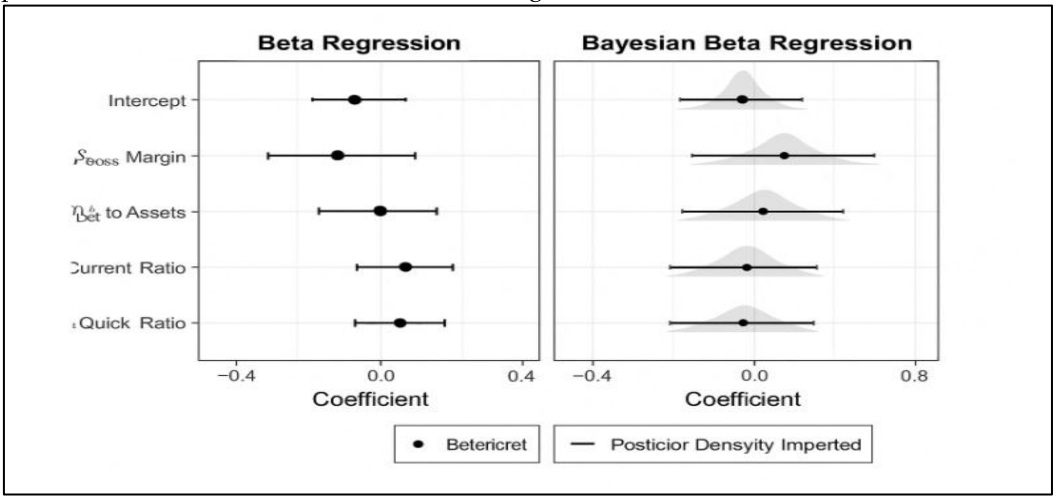

Figure 4. show the confidence interval of the frequentist and Bayesian methods.

4. Conclusion

This study presented a comparative analysis of Frequentist and Bayesian Beta Regression methods using two financial datasets: the Credit Dataset and the Company Financial Dataset. Through the evaluation of parameter estimates, interval interpretations, and predictive performance metrics (RMSE and MAE), both methods demonstrated strong modeling capabilities with consistent results for key predictors. The Bayesian approach offered several advantages, including more informative uncertainty quantification through credible intervals and slightly better predictive accuracy. While the Frequentist method remains straightforward and computationally efficient, the Bayesian model provided richer insights, particularly for borderline-significant predictors. Overall, the results suggest that Bayesian Beta Regression is a valuable alternative to the frequentist method, especially in applications where interpreting uncertainty and improving predictive performance are important.

REFERENCES

[1] S. L. P. Ferrari and F. Cribari-Neto, "Beta regression for modelling rates and proportions," *J. Appl. Stat.*, vol. 31, no. 7, pp. 799–815, 2004, doi: 10.1080/0266476042000214501.

[2] E. Cepeda-Cuervo and D. Gamerman, "Bayesian modeling of variance heterogeneity in normal regression," *Braz. J. Probab. Stat.*, vol. 19, no. 2, pp. 161–175, 2005.

[3] R. A. Rigby, D. M. Stasinopoulos, and P. W. Lane, "Generalized additive models for location, scale and shape," *J. Roy. Stat. Soc. C (Appl. Stat.)*, vol. 54, no. 3, pp. 507–554, 2005.

[4] Gelman, J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin, *Bayesian Data Analysis*, 3rd ed. Boca Raton, FL: Chapman and Hall/CRC, 2013.

[5] Yahoo Finance, "Company Financial Data," *Yahoo Finance*, 2025. [Online]. Available: https://finance.yahoo.com

[6] C. Yeh and C. H. Lien, "The comparisons of data mining techniques for the predictive accuracy of probability of default of credit card clients," *Expert Syst. Appl.*, vol. 36, no. 2, pp. 2473–2480, 2009.

[7] Arroyave, "A comparative analysis of the effectiveness of corporate bankruptcy prediction models based on financial ratios: Evidence from Colombia," *J. Int. Stud.*, vol. 11, no. 1, pp. 273–287, 2018.

[8] T. H. Bui, T. T. D. Truong, and T. P. T. Tran, "Financial ratios as indicators in bankruptcy prediction: A comparative analysis of statistical and machine learning models," 2025.

[9] H. Y. Ching and F. Gerab, "Determinants of financial performance in Brazilian companies: a multi-ratio model using multivariate statistical method," *J. Global Bus. Econ.*, vol. 5, no. 1, pp. 79–99, 2012.

[10] C. Meaney and R. Moineddin, "A Monte Carlo simulation study comparing linear regression, beta regression, variable-dispersion beta regression and fractional logit regression at recovering average difference measures in a two sample design," *BMC Med. Res. Methodol.*, vol. 14, no. 1, p. 14, 2014.
[11] Kumar, V. Kumar, D. Verma, and S. Sharma, "Prediction of corporate bankruptcy based on financial ratios using binary logistic regression," *Int. J. Stat. Reliab. Eng.*, vol. 7, no. 3, pp. 376–381, 2021.
[12] S. Hilkevics and V. Semakina, "The classification and comparison of business ratios analysis methods," *Insights Reg. Dev.*, vol. 1, no. 1, pp. 47–56, 2019.
[13] N. Refenes, A. Zapranis, and G. Francis, "Stock performance modeling using neural networks: a comparative study with regression models," *Neural Netw.*, vol. 7, no. 2, pp. 375–388, 1994.
[14] A. Boyacioglu, Y. Kara, and Ö. K. Baykan, "Predicting bank financial failures using neural networks, support vector machines and multivariate statistical methods: A comparative analysis in the sample of savings deposit insurance fund (SDIF) transferred banks in Turkey," *Expert Syst. Appl.*, vol. 36, no. 2, pp. 3355–3366, 2009.
[15] D. L. Olson, D. Delen, and Y. Meng, "Comparative analysis of data mining methods for bankruptcy prediction," *Decis. Support Syst.*, vol. 52, no. 2, pp. 464–473, 2012.
[16] Z. Huang, H. Chen, C. J. Hsu, W. H. Chen, and S. Wu, "Credit rating analysis with support vector machines and neural networks: a market comparative study," *Decis. Support Syst.*, vol. 37, no. 4, pp. 543–558, 2004.