BAREKENG: Journal of Mathematics and Its Applications
September 2026 Volume 20 Issue 3 Page 2085-2098
P-ISSN: 1978-7227 E-ISSN: 2615-3017
https://doi.org/10.30598/barekengvol20iss3pp2085-2098

DEVELOPMENT STUDY OF GLMM-GEE-TREE REGRESSION MODELLING FOR BETA DISTRIBUTION RESPONSE DATA (IMPLICATIONS OF GINI RATIO MODELING IN INDONESIA, 2018-2024)

Pardomuan Robinson Sihombing [1*], **Erfiani** [2], **Khairil Anwar Notodiputro** [3], **Anang Kurnia** [4]

[1,2,3,4] *School of Data Science, Mathematics and Informatics, IPB University*
Jln. Raya Darmaga Kampus IPB, Babakan, Kec. Dramaga, Kabupaten Bogor, Jawa Barat 16680, Indonesia

*Corresponding author's e-mail: * **robinsonpardomuan@apps.ipb.ac.id***

Article Info

Article History:
Received: 24th July 2025
Revised: 22nd November 2025
Accepted: 16th March 2026
Available online: 8th April 2026

Keywords:
Beta;
GEE;
Gini ratio;
GLMM;
Tree.

ABSTRACT

Economic inequality remains one of the most persistent challenges faced by Indonesia as a developing country. Previous studies have predominantly employed conventional models such as Ordinary Least Squares (OLS) or Panel Least Squares. However, these models are often inappropriate, as they fail to account for the bounded nature of inequality indices such as the Gini ratio, which ranges between 0 and 1. Beta regression offers a more appropriate alternative. In the context of panel data, Generalized Linear Mixed Models (GLMM) and Generalized Estimating Equations (GEE) are commonly used to handle correlated data; however, their integration with nonlinear models for longitudinal Beta-distributed responses remains limited. This study proposes a novel GLMM-GEE-Tree modeling approach for Beta-distributed response data. The proposed model combines GLMM (to capture individual random effects), GEE (to handle temporal correlation and provide robust marginal estimates), and Regression Trees (to address nonlinear relationships and complex interactions). The aim is to simultaneously tackle the challenges of proportional responses, panel structure, random effects, correlation, and nonlinearity. Empirical validation uses Gini ratio data from 34 Indonesian provinces spanning 2018 to 2024. The findings reveal that in this empirical data, the GLMM-GEE-Tree model outperforms alternative models, achieving an R^2 of 0.472 and a QIC of 13.435 and yielding the lowest AIC and BIC values.

How to cite this article:

P. R. Sihombing, Erfiani, K. A. Notodiputro and A. Kurnia., "DEVELOPMENT STUDY OF GLMM-GEE-TREE REGRESSION MODELLING FOR BETA DISTRIBUTION RESPONSE DATA (IMPLICATIONS OF GINI RATIO MODELING IN INDONESIA, 2018-2024)", *BAREKENG: J. Math. & App.*, vol. 20, no. 3, pp. 2085-2098, Sep, 2026.

Journal homepage: *https://ojs3.unpatti.ac.id/index.php/barekeng/*
Journal e-mail: *barekeng.math@yahoo.com; barekeng.journal@mail.unpatti.ac.id*

Research Article · Open Access

1. INTRODUCTION

The Gini Ratio (Gini) is a commonly used measure of income inequality within a population. It ranges from 0 to 1, where a value of 0 represents perfect equality (everyone has the same income), and a value of 1 represents maximum inequality (one individual or household has all the income while others have none). As measured by the Gini, income inequality in Indonesia has exhibited a fluctuating trend, with an upward movement observed in September 2024 following previous periods of decline. In September 2024, the national Gini reached 0.381, an increase from 0.379 recorded in March 2024. This figure indicates that income inequality remains moderate and shows signs of widening after a temporary improvement earlier in the year [1]. Considering the various factors influencing inequality, the government has undertaken multiple efforts to reduce its level. Since policy formulation is often grounded in academic studies, modelling inequality within academic research warrants careful and thorough attention.

Modeling the determinants of inequality is essential for formulating effective and equitable policies. By identifying the underlying causes of inequality through empirical approaches and appropriate econometric models, the government can design more targeted interventions to reduce inequality in Indonesia [2], [3]. To date, the modeling inequality determinants have predominantly relied on conventional Gaussian-based estimation methods, such as Ordinary Least Squares (OLS), Panel Regression (including Fixed Effects and Random Effects Models), and the Generalized Method of Moments (GMM) [4], [5], [6], [7], [8], [9]. Unfortunately, using Gaussian-based estimation methods is often challenged by violations of classical assumptions. Such violations can lead to biased estimates and invalid inferences, thereby compromising the reliability of the model's outcomes [10]. This challenge becomes particularly relevant when dealing with response variables inherently bounded between zero and one, such as ratios, proportions, or percentages, such as the Gini Ratio or the Village Development Index [11], [12]. In cases involving such data, the assumptions of normality and homoscedasticity are challenging to satisfy. As a result, Gaussian-based regression models become inefficient and inappropriate for analysis [13].

To address these limitations, Beta Regression has emerged as a statistically more appropriate alternative for modeling continuous response variables that are bounded within an open interval (0,1) [14], [15]. This model accommodates non-normal data distributions and the heteroskedasticity inherent in proportional data by modeling the Beta distribution's mean and precision parameters[16]. Nevertheless, large-sample inference in Beta Regression relies on first-order asymptotics, which tend to be inefficient when applied to small sample sizes [15] [14]. In addition, outliers can significantly affect the Maximum Likelihood Estimation (MLE) in Beta Regression models, although robust Bayesian solutions have been developed to mitigate this influence [14]. Research on Beta Regression modeling has grown consistently over the past decade, both in volume and scope. The increasing number of publications and the expansion of thematic applications reflect the model's growing relevance across various academic disciplines [9]. Therefore, the application of Beta Regression in modeling the determinants of economic inequality deserves further attention and exploration.

On the other hand, the development and application of panel data analysis, which combines cross-sectional and time series data, has progressed rapidly across various fields such as health, economics, and the social sciences [17], [18], [19]. In the context of panel data, Generalized Linear Mixed Models (GLMM) and Generalized Estimating Equations (GEE) are two popular paradigms used to extend classical regression models [20], [21]. GLMM enables the estimation of subject-specific or conditional parameters, making it well-suited for capturing individual-level heterogeneity and understanding mechanisms at the individual level [19], [20], [22]. However, the likelihood function in GLMM often involves high-dimensional integration, which makes it computationally challenging to maximize and prone to convergence issues, particularly in complex models or large datasets [20], [21]. In contrast, GEE focuses on modeling the population-averaged marginal mean and is well known for its robustness to misspecification of the working correlation structure, providing consistent estimates even when the assumed correlation structure is mis-specified [20], [21], [23]. Although GEE is analytically simpler than GLMM [22], its robustness to correlation misspecification may come at the cost of estimation efficiency [24]. Simulation studies have shown that GEE can consistently estimate correlation parameters, whereas GLMM often produces biased estimates of these parameters[25].

Although GLMM and GEE have been extensively applied in the analysis of longitudinal data, their utilization in modeling response variables following a Beta distribution remains limited in the existing literature [25], [26]. Several studies have attempted to compare the performance of GLMM and GEE in the context of longitudinal binary or count data, such as [18], [21], [22], [23], [27], [28]. However, direct comparisons between estimates from marginal models (GEE) and conditional models (GLMM) warrant further scrutiny, as the parameters are not directly comparable due to their differing interpretations [18]. In addition, the development of modeling approaches that simultaneously integrate GLMM and GEE methods remains largely underexplored in the current literatures [24]. Existing studies, such as those employing bivariate Beta models, remain limited to cases involving only two repeated measurements [29].

Further developments are not confined to modeling linear relationships between independent and dependent variables. Adopting machine learning approaches, such as regression trees, can overcome these limitations by capturing nonlinear relationships and complex interactions among predictors [30], [31]. Although previous studies have applied regression trees within the framework of generalized linear mixed models, such as GLMM trees for binary or binomial data [30], [31], their application to data with Beta-distributed responses remains a relatively underexplored area.

Based on the identified research background, this study proposes developing a novel method entitled "Development of the GLMM-GEE-Tree Modeling Framework for Beta-Distributed Response Data." This approach seeks to integrate the strengths of three distinct methodologies: Generalized Linear Mixed Models (GLMM) to account for random effects and individual-level heterogeneity, Generalized Estimating Equations (GEE) to incorporate the working correlation structure in time-series data, and Regression Trees to model nonlinear relationships. The novelty of this analysis lies in integrating these three components into a unified modeling framework designed explicitly for Beta-distributed.

2. RESEARCH METHODS

2.1 Beta Distribution

The Beta distribution is appropriate for modelling response variables constrained within the open interval (0,1), such as proportions or ratios. The Probability Density Function (PDF) of the Beta distribution is defined by two positive shape parameters that govern the distribution's form, namely α_0 and β_0

$$f(y \mid \alpha_0, \beta_0) = \frac{\Gamma(\alpha_0+\beta_0)}{\Gamma(\alpha_0)\Gamma(\beta_0)} y^{\alpha_0-1}(1-y)^{\beta_0-1}. \quad (1)$$

In the context of regression, the Beta distribution is often reparametrized in terms of mean ($\boldsymbol{\mu}$) and precision (ϕ). Parameter $\mu = E[Y] = \frac{\alpha_0}{\alpha_0+\beta_0}$ represents the expected value of the response variable, while $\phi = \alpha_0 + \beta_0$ controls variance's distribution; the larger ϕ, the smaller its variance. This reparameterization enables separate modeling of the mean and precision of the response. The likelihood function for a single observation y_i that follows a Beta distribution with mean parameter mean μ_i and precision ϕ_i is:

$$L(y_i \mid \mu_i, \phi_i) = \frac{\Gamma(\phi_i)}{\Gamma(\mu_i\phi_i)\Gamma((1-\mu_i)\phi_i)} y_i^{\mu_i\phi_i-1}(1-y_i)^{(1-\mu_i)\phi_i-1}, \quad (2)$$

and its log likelihood function is

$$\log L(y_i \mid \mu_i, \phi_i) = \log\Gamma(\phi_i) - \log\Gamma(\mu_i\phi_i) - \log\Gamma\big((1-\mu_i)\phi_i\big) + (\mu_i\phi_i - 1)\log(y_i) + \big((1-\mu_i)\phi_i - 1\big)\log(1-y_i). \quad (3)$$

For N independent observations, the total log-likelihood is the sum of individual log-likelihoods.

2.2 Beta Regression

Beta regression is an extension of the Generalized Linear Model (GLM) designed for response variables that follow a Beta distribution. This model allows both the mean parameter mean (μ_{it}) and the precision parameter (ϕ_{it}) to be modeled as functions of linear predictors.

$$\text{Model for mean } \mu_{it}: g(\mu_{it}) = \eta_{it} = X_{it}\beta, \quad (4)$$

where $g(\cdot)$ is a link function (such as logit, probit, cloglog, log, or loglog) that links μ_{it} With linear predictors η_{it}, and $X_{it}\beta$ is a linear combination of predictors. The logit function is a common choice because it maps (0,1). Note: to all real numbers. i denotes the observation index ($i = 1, 2, \dots, n$), and t represents the year or time period.

$$\text{Model for precision } \phi_{it}: h(\phi_{it}) = \lambda_{it} = W_{it}\gamma, \quad (5)$$

where $h(\cdot)$ is a link function (such as log, identity, or sqrt) which ensures ϕ_{it}>0, and $W_{it}\gamma$ is a linear combination of predictors for precision. Log functions are often used to ensure positive precision values. Parameter estimation (β and γ) uses the Maximum Likelihood (ML) method through numerical optimization algorithms such as Newton-Raphson or Fisher Scoring.

2.3 Generalized Linear Mixed Model (GLMM)

GLMM extends the Generalized Linear Model (GLM) by incorporating random effects for correlated or hierarchical data structures, such as repeated observations within the same individual (panel data). These random effects enable the modelling of heterogeneity across observational units.

GLMM model for response Y_{it} adding random effects u_i to linear predictor:

$$\eta_{it} = X_{it}\beta + \boldsymbol{Z_{it}}\boldsymbol{u_i}. \tag{6}$$

In where $\boldsymbol{Z_{it}}$ is a design matrix for random effects, and u_i is a random effect vector for an individual i, assumed to be normally distributed $u_i \sim N(0, \Sigma_u)$. Parameter estimation in GLMM is based on the marginal likelihood, which often lacks a closed-form solution and requires numerical approximation methods such as the Laplace Approximation or Adaptive Gaussian Quadrature.

2.4 Generalized Estimating Equations (GEE)

Generalized Estimating Equations (GEE) is a semi-parametric method for analyzing correlated data, primarily focusing on estimating population-averaged effects. GEE is well known for its robustness in the misspecification of the working correlation structure. Parameter estimation in GEE is achieved by solving quasi-likelihood estimating equations. The GEE framework employs the same mean and precision structures as those used in Generalized Linear Models (as shown in Eqs. (4) and (5)). The key difference lies in how the correlation among within-individual observations is handled during the estimation process. The key to GEE is the specification of the working correlation matrix. $(\mathbf{R}_i(\boldsymbol{\alpha}))$ for each $\boldsymbol{i}$. This matrix illustrates the assumed correlation pattern between $\boldsymbol{T_i}$ Observations of individuals $\boldsymbol{i}$. Parameter $\boldsymbol{\alpha}$ in $\mathbf{R}_i(\boldsymbol{\alpha})$ It is a correlation parameter.

For each $\boldsymbol{i}$, the working Covariance Matrix ($\mathbf{V}_i$) is developed. The size of $\mathbf{V}_i$ is $(\boldsymbol{T_i} \times \boldsymbol{T_i})$ and reflect both the marginal variance of the Beta response and the work correlation structure.

$$\mathbf{V}_i(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\alpha}) = \mathbf{A}_i^{1/2}\mathbf{R}_i(\boldsymbol{\alpha})\mathbf{A}_i^{1/2}, \tag{7}$$

in where: $\mathbf{A}\boldsymbol{i} = \mathbf{diag}\left(\mathbf{Var}(\boldsymbol{Yi}\mathbf{1}), \dots, \mathbf{Var}\left(\boldsymbol{Y}_{\boldsymbol{iT_i}}\right)\right)$: A diagonal matrix containing the marginal variance of each observation $\boldsymbol{Y_{it}}$. For Beta distributions, $\mathbf{Var}(\boldsymbol{Y_{it}}) = \frac{\boldsymbol{\mu_{it}}(\mathbf{1}-\boldsymbol{\mu_{it}})}{\mathbf{1}+\boldsymbol{\phi_{it}}}$

2.5 Tree-based Modelling

Tree-based modelling approaches, such as regression trees, recursively partition the predictor space into smaller sub-regions, with each terminal node fitted with its localized regression model. This structure enables the model to capture nonlinear relationships and complex interactions among predictors. In the context of Beta regression, model-based recursive partitioning facilitates the identification of subpopulations in which the association between predictors and the response variable varies, thereby enhancing both the flexibility and interpretive capacity of the model.

2.6 GLMM-GEE-Tree Model

The proposed GLMM-GEE-Tree model represents a synthesis of the strengths of GLMM, GEE, and tree-based approaches to address the complex characteristics of Beta-distributed response data, particularly in panel or longitudinal settings. The integration concept involves combining the GLMM framework to account for individual-level heterogeneity through random effects. At the same time, GEE is employed to model intra-subject correlation (such as time-series dependence in panel data) and provide robust estimates of population-averaged effects. Additionally, regression trees are utilized to adaptively identify data subgroups, allowing fixed-effect coefficients to vary across these subgroups and thereby capturing nonlinear relationships and complex interactions among predictors. The GLMM-GEE-Tree Beta model simultaneously models both the mean and precision parameters of the Beta distribution.

2.6.1. Mean (μ_{it}):

Linear predictors for mean $\boldsymbol{\mu_{it}}$ combines fixed effects, which vary based on the tree partition, and random effects: $\boldsymbol{g(\mu_{it}) = \eta_{it} = X_{it}\beta_{k(p_{it})} + Z_{it}u_i}$ in which:

- $\boldsymbol{g(\cdot)}$: Beta link function (such as, logit), $\boldsymbol{X_{it}}$: Vector variable predictor for fixed effects.
- $\boldsymbol{\beta_{k(p_{it})}}$: Coefficient vector for fixed effects, specific to the terminal node k where the observation (i,t) is located. Tree partition rules are defined by the $\boldsymbol{p_{it}}$.
- $\boldsymbol{Z_{it}}$: Design matrix for random effects
- $\boldsymbol{u_i}$: Random effect vector for individuals $\boldsymbol{i}$, assumed $\boldsymbol{u_i \sim N(0, \Sigma_u)}$.

2.6.2. Model Precision (ϕ_{it})

Parameter precision $\boldsymbol{\phi_{it}}$ modelled as a function of predictors. For the initial purpose, we can assume a precision model without a tree effect or random effect, however, this can be extended: $h(\phi_{it}) = \lambda_{it} = W_{it}\gamma$ in which $h(\cdot)$ is a link function (e.g., log).

2.6.3. GEE (Correlation of Work and Robustness) Aspects

The integration of GEE into the model will be implemented through the working correlation structure. During the estimation process, various working correlation structures (e.g., exchangeable, AR(1), unstructured) will be considered to capture the dependence among observations within clusters (individuals). This structure will influence the covariance matrix used in the estimating equations. Even if the selection of matrix correlation is mis specified, the model will be still robust. A sandwich estimator will be employed to calculate the standard errors of the fixed effects coefficients, ensuring robustness against misspecification of the correlation structure.

Parameter estimation in the GLMM-GEE-Tree model requires an iterative algorithm that integrates principles from model-based recursive partitioning with estimation procedures from both GEE and GLMM frameworks. The procedure consists of three steps, namely parameter estimation, theoretical evaluation, and practical evaluation.

1. Parameter Estimation: The estimation algorithm will be iterative (similar to EM algorithms or alternating optimization that is used to estimate parameters of the GLMM component in the presence of latent variables or incomplete data):
   a. Initialization: Set the initial value for all parameters (fixed effects coefficient $\boldsymbol{\beta_k}$, Precision Coefficient $\boldsymbol{\gamma}$, and random effect covariance matrix $\boldsymbol{\Sigma_u}$).
   b. Iteration (Loop): Tree Partition Step: The data is partitioned using a customized Beta regression tree algorithm with the current random effect as an offset. This model will result in a tree structure and an initial estimate of βk for each terminal node. The goal is to minimize variance/impurity of Y and detect instability in GLM coefficients.
   c. GLMM-GEE Estimation Steps: Based on the new tree structure, GLMM parameters ($\boldsymbol{\beta_k}$, $\boldsymbol{\gamma}$, and $\boldsymbol{\Sigma_u}$) are estimated using probability approximation methods (e.g., Laplace Approximation). In this step, the estimation equation will be modified to include information from the GEE working correlation structure.
   d. Correlation Parameter Update (if relevant): The working correlation structure parameters are updated based on the residual resulting from the GLMM-GEE estimate.
   e. Convergence Check: The iteration continues until the parameter changes between iterations are below the specified tolerance threshold. Challenges include stable convergence and computational efficiency, especially with complex data.
2. Theoretical Evaluation Criteria
   a. Consistency: The estimator should converge to the actual parameter values as the sample size increases.
   b. Efficiency: The variance of the proposed model's estimator should be competitive with, or superior to, that of existing models.
   c. Robustness: The model should exhibit resilience to misspecification of key assumptions, particularly regarding the working correlation structure and random effects distribution.
3. Practical Evaluation Metrics:

a. Prediction Accuracy: Assessed using metrics such as R-square and Root Mean Squared Error (RMSE) on the response scale.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}, \quad (8)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n}}, \quad (9)$$

y_i = actual (observed) value,
$\hat{y}_i$ = predicted value,
n = number of observations,
$\bar{y}$ = mean of the actual values.

RMSE measures the average magnitude of prediction errors, a smaller RMSE means the model is more accurate. RMSE reflects the average prediction error on the response scale; thus, lower RMSE values correspond to better predictive accuracy. R^2 measures the proportion of variance in the dependent variable explained by the model, with higher values indicating stronger explanatory power.

b. Goodness-of-Fit: Evaluated using criteria such as Deviance, Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), or Quasi-likelihood under the Independence Model Criterion (QIC) for GEE-based models.

$$\text{AIC} = 2\text{k} - 2\ln(\hat{\text{L}}), \quad (10)$$

$$\text{BIC} = \text{k}.\ln(\text{n}) - 2\ln(\hat{\text{L}}), \quad (11)$$

$$\text{QIC} = -2\text{Q}(\hat{\beta}) + 2.\text{trace}(\Omega_{\text{I}}^{-1}\Omega), \quad (12)$$

k = number of estimated parameters in the model,
n = number of observations,
$\hat{\text{L}}$ = maximum likelihood value,
$\text{Q}(\hat{\beta})$ = quasi-likelihood evaluated at estimated parameters $\hat{\beta}$,
Ω_{I}^{-1}= model-based covariance matrix under the independence working correlation assumption,
Ω = robust (empirical) covariance matrix.

AIC, BIC, and QIC are information criteria used to assess model fit while penalizing for complexity; lower values indicate a better trade-off between fit and parsimony. Taken together, these indicators offer a multidimensional view of model performance, capturing goodness-of-fit, generalizability, and predictive reliability.

c. Interpretability: Measured by the ease of understanding the tree partitioning rules, the significance of coefficients within each terminal node, and the contribution of random effects to the overall model structure.

2.7 Data and Variables

The data used to apply the model are obtained from BPS-Statistics Indonesia, comprising observations from 34 provinces over 7 years (2018–2024). The study includes the following research variables:

Table 1. Research Variables

Dependent Variable	Unit
Gini ratio	Point
Human Development Index (HDI)	Point
Poverty Rate	Percent
Unemployment Rate	Percent
Economic Growth	Percent

As independent variables, the key socioeconomic indicators used in this study include the Human Development Index (HDI), poverty rate, unemployment rate, and economic growth. HDI, developed by the United Nations Development Program (UNDP), is a composite index measuring average achievement in three basic dimensions of human development: health (life expectancy at birth), education (mean years of schooling for adults and expected years of schooling for children), and standard of living (gross national income per capita). The poverty rate is defined as the proportion of the population living below the national poverty line or international thresholds (e.g., \$1.90/day) as reported by official statistics agencies or the World Bank. The unemployment rate represents the share of the labor force that is without work but actively seeking employment, following the International Labor Organization (ILO) guidelines. Economic growth is measured as the annual percentage change in gross domestic product (GDP), reflecting the overall expansion of the economy, consistent with World Bank and national statistical definitions.

The relationship between the Gini Ratio and key socioeconomic indicators such as HDI, poverty rate, unemployment, and economic growth provide an ideal case to demonstrate the capabilities of the GLMM–GEE–Tree model for Beta-distributed data. The Gini Ratio is naturally bounded between 0 and 1, making it suitable for Beta regression, which properly handles proportion data and accounts for heteroskedasticity and non-normality. At the same time, these socioeconomic indicators may exhibit nonlinear effects, interactions, and hierarchical structures across regions and over time, which can be effectively captured by the combined GLMM, GEE, and tree partitioning framework.

3. RESULTS AND DISCUSSION

The Gini ratio is one of the most widely used measures for assessing economic inequality within a region. In Indonesia, the Gini ratio is calculated based on household consumption expenditure, as recorded through the National Socio-Economic Survey. The following presents an overview of the Gini ratios across Indonesian provinces in the year 2024:

Figure 1. Gini Ratio in Indonesia by Province, 2024
(Source: Calculated from BPS-Statistics Indonesia by Bing Map-Excel)

The Gini Ratio map for 2024 reveals significant regional disparities in income distribution across Indonesian provinces. Darker shades on the map indicate higher levels of income inequality, while lighter shades represent more equitable distributions. The most pronounced inequality is observed in West Java, which is depicted in the darkest color, suggesting that income is distributed most unevenly in this province relative to others. These spatial patterns suggest that income inequality is not confined to the more developed or urbanized regions. Instead, it is a complex phenomenon influenced by a range of structural and spatial factors. The findings underscore the need for geographically differentiated policy approaches that consider local development conditions and aim to promote more inclusive economic outcomes across the archipelago.

Furthermore, Table 2 presents the descriptive statistics of the research variables. The descriptive statistics presented in this table summarize data across all observed years. Consequently, differences between regions reflect aggregated outcomes over the study period, capturing overall regional variation rather than year-specific values. This approach allows for a more robust comparison of socioeconomic and inequality indicators across regions.

The average Gini Index is 0.348, with a minimum value of 0.236 observed in Bangka Belitung Province in 2022 and a maximum value of 0.449 recorded in Yogyakarta Province in 2023. The average HDI stands at 71.72, ranging from 60.06 in Papua Province in 2018 to a maximum of 83.08 in Jakarta Province in 2024. The average poverty rate is 10.34 percent, with the lowest recorded at 3.47 percent in Jakarta Province in 2019 and the highest at 27.74 percent in Papua Province in 2018. The average open unemployment rate is 5.01 percent, with a minimum of 1.40 percent in Bali Province in 2018 and a maximum of 10.95 percent in Jakarta Province in 2020. The average economic growth rate is 4.29 percent, with the lowest value of -15.74 percent in Papua Province in 2019 and the highest at 22.94 percent in North Maluku Province in 2022.

These figures indicates that the disparity of the indicators across Indonesia are still high. As a consequence, the results underscore the importance of region-specific policy measures. While national-level strategies are important, our findings suggest that regions with varying levels of HDI, poverty, unemployment, and economic growth face distinct inequality dynamics. Therefore, policymakers should design geographically differentiated interventions that consider local development conditions, ensuring that policies are targeted, effective, and responsive to regional heterogeneity.

Table 2. Descriptive Statistics

Variable	Gini	HDI	Poverty	Unemployment	Growth
Mean	0.348	71.72	10.34	5.01	4.29
Minimum	0.236	60.06	3.47	1.40	-15.74
Maximum	0.449	83.08	27.74	10.95	22.94
Std deviation	0.041	3.93	5.29	1.71	4.04

Subsequently, a traditional regression model was estimated using the Ordinary Least Squares (OLS) estimator and diagnostic tests for classical assumptions. Table 3 presents the results of classical assumption testing for the Ordinary Least Squares (OLS) regression model. The normality test yields a p-value of 0.033, indicating that the residuals are not normally distributed, thus violating one of the key assumptions of OLS. The homoscedasticity test, with a p-value of 0.060, suggests that the variance of the residuals is relatively constant across observations. The autocorrelation test shows a p-value of 0.000, implying the presence of autocorrelation in the residuals, which may compromise the efficiency of the estimators. Finally, the linearity test reports a p-value of 0.020, indicating that the relationship between the independent and dependent variables may not be adequately captured by a linear specification. Taken together, these diagnostic results suggest that the OLS model is likely mis specified and that alternative modeling approaches, particularly those that relax classical assumptions or accommodate nonlinearity and serial correlation, are more appropriate for the data at hand.

Table 3. Classical Assumption Testing

Test	Prob. Value	Conclusion
Normality	0.033	Residuals are not normally distributed
Homoscedasticity	0.060	Residual's variance is homogeneous
Autocorrelation	0.000	Autocorrelation of error between times occurs
Linearity	0.020	Non-linear models

Subsequent modeling was conducted using several approaches, including GLM-Beta, GLMM-Beta, GEE-Beta, Beta Tree, GLMM-Tree-Beta, and the proposed GLMM-GEE-Tree-Beta model. Table 4 provides a comparative evaluation of several regression models using five key performance indicators: Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), Quasi-likelihood under the Independence model Criterion (QIC), coefficient of determination (R^2), and Root Mean Squared Error (RMSE).

Table 4. Model Evaluation

Model	AIC	BIC	QIC	R2	RMSE
OLS	-889.211	-868.377	-	0.212	0.036
GLM	-888.649	-888.285	-	0.220	0.036
Tree Beta	-1013	-871	-	-	0.027
GLMM	-1322.033	-1297.728	-	0.167	0.010
GEE	-177.536	-156.703	15.55	0.22	0.162
GLMM Tree	-1245	-1206	-	-	0.0352
GLMM-GEE-Tree	-39.915	-29.489	**13.435**	**0.472**	0.0352

Among the evaluated models, the GLMM-GEE-Tree model demonstrates the strongest overall performance, as reflected by its highest R^2 value of 0.472 and a relatively low RMSE of 0.0352. This suggests that the model explains a substantially larger proportion of the variance in the data compared to other models, while maintaining strong predictive accuracy. Additionally, it records the lowest QIC value (13.435) among models where QIC is applicable, indicating superior model fit in the context of correlated data. Although AIC and BIC can be computed for GEE models, the QIC (Quasi-likelihood under the Independence model Criterion) is considered a more appropriate measure of model fit for GEE components, as it accounts for the quasi-likelihood estimation and the correlation structure. In addition, R^2 shows that the GLMM–GEE–Tree model effectively captures the contributions of independent variables, reflecting its explanatory power. By considering both explanatory power (R^2) and predictive performance (RMSE/QIC), the GLMM–GEE–Tree model provides a balanced approach that combines accuracy and robustness.

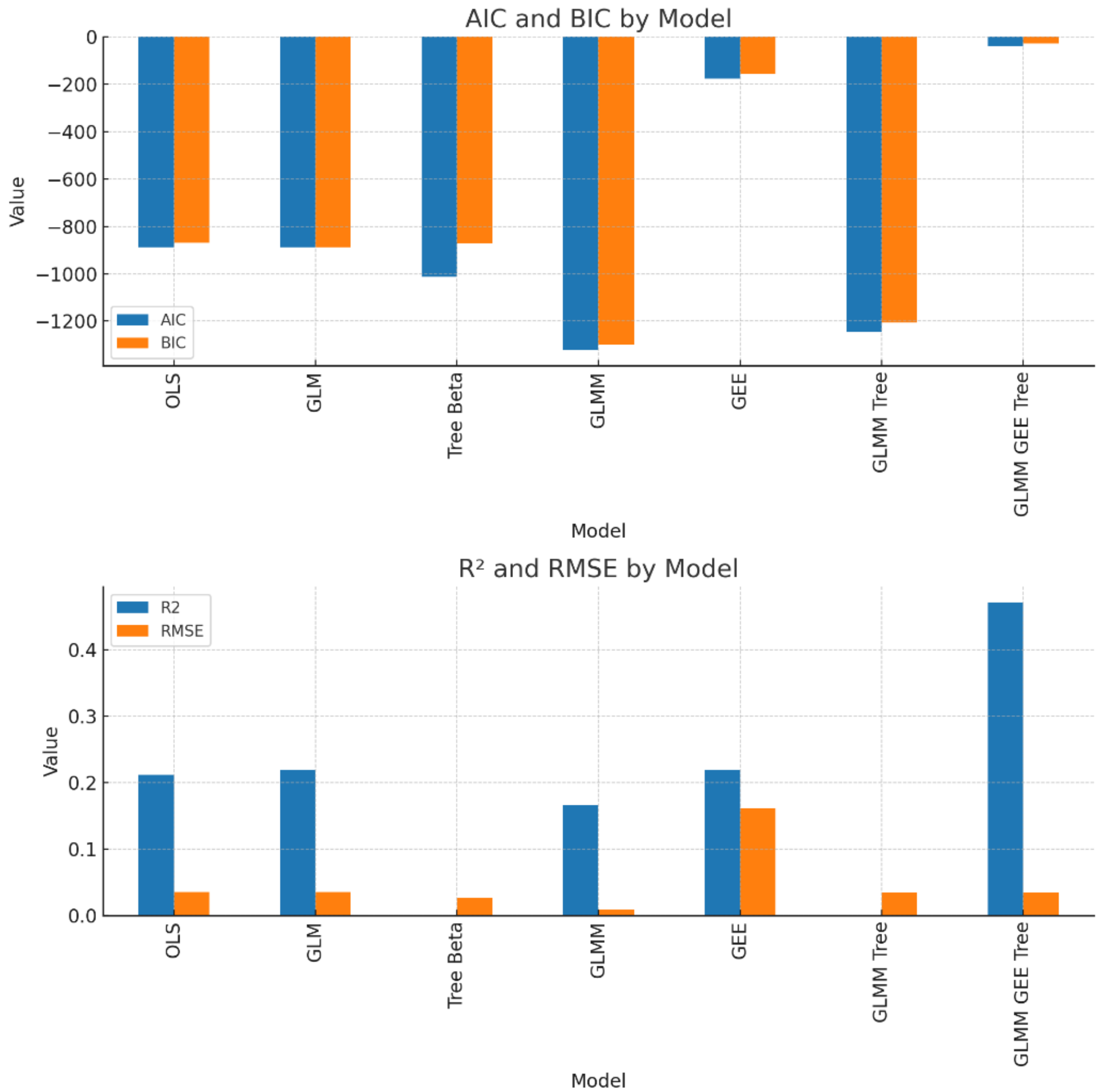

Figure 2. Model Evaluation
(Source: Author's calculation by R software)

In contrast, as in Fig. 2, while the GLMM model achieves the lowest AIC (-1322.033) and BIC (-1297.728) values, it underperforms in terms of R^2 (0.167) and RMSE (0.010), suggesting limited explanatory capacity despite better penalization scores for model complexity. Similarly, the Tree Beta model yields the lowest RMSE (0.027), yet lacks R^2 and QIC values, making its overall performance more difficult to assess in terms of explanatory adequacy. Traditional models such as OLS and GLM show relatively poor fit, with low R^2 values (0.212 and 0.220, respectively) and identical RMSEs (0.036), underscoring their limitations when applied to bounded and potentially non-linear data. The GEE model performs slightly better in terms of R^2 (0.22), but its higher RMSE (0.162) and QIC (15.55) indicate potential issues in capturing residual variability. Overall, the results highlight the advantages of hybrid and flexible modeling approaches, particularly the GLMM-GEE-Tree framework, in addressing complex data characteristics such as heterogeneity, correlation, and nonlinearity.

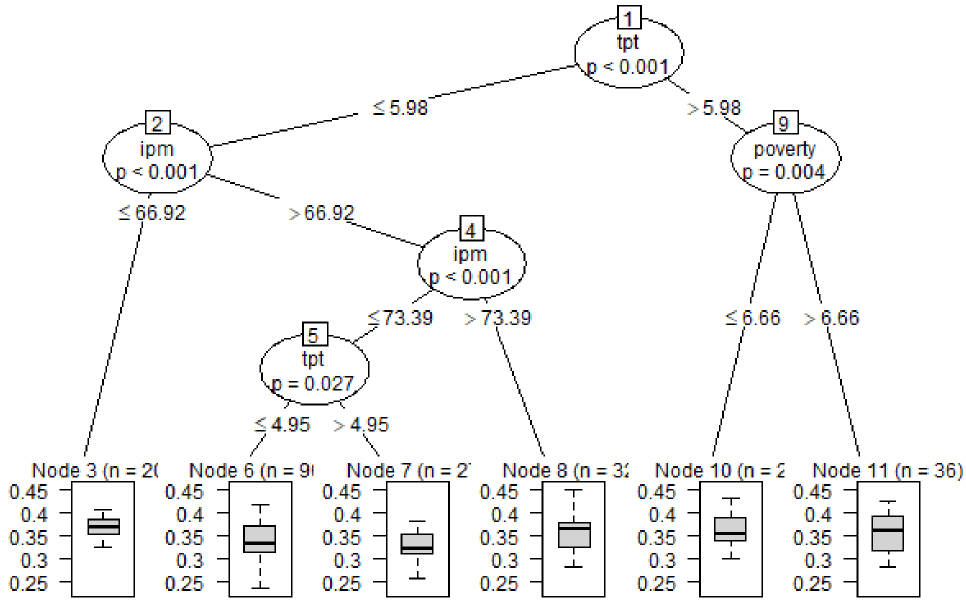

Figure 3. Tree Structure Visualization

Based on the results of Fig. 3, the analysis identified three key variables that significantly affect income inequality (Gini ratio): Unemployment, Human Development Index (HDI), and poverty rate. The model produced six distinct subgroups with heterogeneous inequality profiles, where Unemployment was the main predictor at the root node ($p < 0.001$) separating the data at a threshold of 5.98, followed by HDI emerging as a separator at three different levels (66.92, 73.39) with $p < 0.001$, and poverty as a differentiator for groups with high Unemployment ($p = 0.004$). Important findings show that the highest inequality (median Gini ~0.375) was found in areas with a combination of low Unemployment (≤5.98) and low HDI (≤66.92) in Node 3 (n=20), as well as in areas with high Unemployment (>5.98) but low poverty (≤6.66) in Node 10 (n=29), while the lowest inequality (median Gini ~0.340) occurred in Node 7 (n=21) with low-medium Unemployment (4.95-5.98) and medium HDI (66.92-73.39). Node 6 is the largest group (n=95), covering the majority of observations with very low poverty (≤4.95) and medium HDI, indicating the need for specific interventions to increase Unemployment in this group. The tree model proved effective in capturing spatial heterogeneity with all separations being highly statistically significant ($p < 0.05$), providing empirical justification for targeted and context-specific policies in reducing inequality based on a combination of regional economic characteristics.

Subsequent hypothesis testing was conducted for the proposed model. Table 5 presents the results of hypothesis testing for five explanatory variables: HDI, poverty rate, unemployment rate, economic growth, and constant term. These variables are evaluated across multiple modelling frameworks, each differing in the way they handle distributional assumptions, correlation structures, and heterogeneity. The GLMM-GEE-Tree model yields the most comprehensive and robust results. In this model, HDI is statistically significant at the 1 percent level, while poverty and unemployment rates are significant at the 5 percent level. All three coefficients are negative, suggesting that higher HDI, lower poverty, and lower unemployment are associated with reduced income inequality. The negative coefficient for HDI (-0.05723) indicates that improvements in education, health, and living standards are linked with more equitable income distribution. Similarly, the negative coefficients for poverty (-0.03838) and unemployment (-0.03159) may reflect redistribution mechanisms or uniform economic pressures across income groups that lead to a narrowing of income gaps.

In contrast, traditional models such as OLS and GLM yield positive and statistically significant coefficients for HDI and poverty, implying that higher values of these indicators are associated with greater inequality. These outcomes may result from the models' inability to fully account for complex data structures, including time dependence and regional-level variation. The GEE model, which addresses correlation but not heterogeneity, also reports positive relationships for HDI and poverty, although with smaller effect sizes.

Across all models, economic growth does not appear to have a significant effect on income inequality, indicating that growth alone is not sufficient to reduce disparities unless accompanied by inclusive development strategies. The results highlight the importance of employing advanced, flexible modelling frameworks, such as the GLMM-GEE-Tree model, to produce more accurate and policy-relevant insights in the context of regional inequality.

Table 5. Testing of Hypothesis

Variable	OLS	GLM	GLMM	GEE	GLMM-GEE-Tree
Constant	-0.0830	-2.50944***	0.311483	-2.5081***	4.1638*
HDI	0.005336***	0.02328***	-0.014369***	0.0231***	-0.05723**
Poverty	0.004664***	0.02044***	0.00772	0.0206***	-0.03838*
Unemployment	0.0004	0.0017	0.0004	0.0023	-0.03159*
Growth	-0.0006	-0.0025	0.0007	-0.0025	0.00161

Signif. codes: 0.001 '***' 0.01 '**' 0.05 '*'

Since the GLMM-GEE-Tree Beta model demonstrates the best performance in empirical analysis, interpretation is based on the coefficients from this model. Several insights regarding the relationship between the Gini Ratio and its predictor variables can be drawn. Empirical evidence strongly indicates that an increase in the HDI is associated with a decrease in the Gini Ratio across various regions in Indonesia. Improvements in the quality of life, reflected in better education, health, and income as captured by the HDI, consistently contribute to a reduction in income inequality in the country [4], [5], [6], [7], [8]. Meanwhile, the relationship between the poverty rate and the Gini Ratio appears negative. Although this may initially seem counterintuitive, such a pattern often emerges in specific regions or periods experiencing relatively high economic growth. In these contexts, rising income inequality may coincide with new economic opportunities for segments of the low-income population, reducing absolute poverty despite an increase in overall inequality [7]. This condition is typically temporary and often occurs during a developing economic phase, before implementing more comprehensive strategies for redistribution and social inclusion. The open unemployment rate also exhibits a negative relationship with income inequality. This negative correlation may also seem counterintuitive, as it is generally assumed that higher unemployment exacerbates inequality. However, this phenomenon can be explained in some instances when unemployment is widespread and relatively evenly distributed across social groups and regions. Low- and middle-income populations may be similarly affected, resulting in a more equal, though downward-shifted, income distribution [32]. This result may lead to a reduction in the Gini Ratio, as income disparities partially diminish due to all income groups being negatively affected relatively equally. Unlike the other factors in this study, economic growth does not significantly affect regional income inequality. Economic growth does not automatically lead to income redistribution. This research can occur when the benefits of expanding sectors are disproportionately captured by middle- to upper-income groups, while lower-income populations remain largely excluded from the gains. In other words, economic growth without inclusivity is insufficient to reduce inequality [33].

4. CONCLUSION

This study introduces the GLMM-GEE-Tree Beta model as a novel analytical framework for modeling bounded response variables in longitudinal data, with a specific application to income inequality across Indonesian provinces. By combining the strengths of Generalized Linear Mixed Models, Generalized Estimating Equations, and regression trees, the model accommodates individual-level heterogeneity, intra-unit correlation, and nonlinear interactions simultaneously. The empirical evidence shows that this integrated approach outperforms conventional methods in terms of model fit and explanatory power. Empirical evaluation confirms that this integrated framework provides a better statistical fit and improved predictive accuracy than conventional models, as shown by lower values of AIC, BIC, QIC, and RMSE, along with a higher R^2. These results highlight the advantages of using flexible, data-adaptive tools in the analysis of complex socio-economic phenomena, particularly when standard assumptions of classical models are not met.

Among the advantages of this model, GLMM-GEE-Tree Beta still has limitations. A notable limitation of the GLMM–GEE–Tree model lies in its computational demand. Due to the model's greater complexity, where GEE estimation is performed separately for each partitioned node, the computation time is longer than that required for simpler models such as GLMM or GEE alone. While this additional computational cost is justified by the model's improved balance between explanatory power and prediction accuracy, users should be aware of this trade-off when applying the approach to large datasets or time-sensitive analyses. In addition, the developed GLMM-GEE-Tree Model focuses on modeling the mean (μ) rather than the precision parameter (φ). Integrating the precision parameter into the tree structure or random effects would significantly increase the computational complexity.

The findings carry significant policy implications. Enhancing human development is crucial to reducing income inequality, while the counterintuitive effects of poverty and unemployment rates suggest the need for nuanced, region-specific interventions. The insignificant role of economic growth emphasizes that inclusive policies must accompany expansionary strategies to ensure equitable outcomes. Despite its contributions, this study has limitations. It does not yet explore interaction effects between predictors or extend the precision component with random or tree-based structures. Future research may refine the model by incorporating more granular administrative data, evaluating alternative estimation algorithms, or integrating spatial effects. Nonetheless, the proposed framework provides a solid foundation for informing targeted, evidence-based policy aimed at addressing persistent regional disparities.

Author Contributions

Pardomuan Robinson Sihombing: Conceptualization, Formal Analysis, Funding Acquisition, Investigation, Methodology, Project Administration, Resources, Software, Supervision, Writing - Review and Editing. Erfiani: Conceptualization, Formal Analysis, Funding Acquisition, Investigation, Methodology, Project Administration, Resources, Software, Supervision, Writing - Review and Editing. Khairil Anwar: Data Curation, Investigation, Resources, Software, Validation, Visualization, and Writing - Original Draft. Anang Kurnia: Investigation, Visualization, and Writing - Review and Editing. All authors discussed and approved the final manuscript.

Funding Statement

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Acknowledgement

The authors extend sincere thanks to School of Data Science, Mathematics and Informatics, IPB University for kind support and assistance that led to this article. Also thanks to Reviewers, and other parties that give some constructive suggestions for make this article more significant to contribute in academic things.

Declarations

The authors declare that there are no conflicts of interest associated with this research.

Declaration of Generative AI and AI-assisted Technologies

Generative AI tools (e.g., Grammarly and Quillbot) were used solely for language refinement (grammar, spelling, and clarity). The scientific content, analysis, interpretation, and conclusions were developed entirely by the authors. The authors reviewed and approved all final text.

REFERENCES

[1] Badan Pusat Statistik Indonesia, "TINGKAT KETIMPANGAN PENGELUARAN PENDUDUK INDONESIA SEPTEMBER 2024," Jakarta, 2025. [Online]. Available: https://www.bps.go.id/id/pressrelease/2025/01/15/2399/gini-ratio-september-2024-tercatat-sebesar-0-381.html

[2] R. Syaifudin, M. S. Alfarisi, G. S. R. Putri, M. A. Jabbar, A. Malik, and F. N. Zulfa, "DETERMINAN KETIMPANGAN WILAYAH DI INDONESIA TAHUN 2012–2022: PENDEKATAN ANALISIS PANEL DINAMIS," *Journal of Business and Economics Research (JBE)*, vol. 5, no. 2, pp. 129–137, 2024, doi: https://doi.org/10.47065/jbe.v5i2.5143.

[3] N. Fahmi and N. S. B. Maria, "ANALISIS DETERMINAN KETIMPANGAN WILAYAH DI PROVINSI DAERAH ISTIMEWA YOGYAKARTA (DIY) TAHUN 2013-2020," *Diponegoro Journal of Economics*, vol. 12, no. 2, pp. 45–56, 2023, doi: https://doi.org/10.14710/djoe.37965

[4] Anin Nabail Azim, Hady Sutjipto, and Rah Adi Fahmi Ginanjar, "DETERMINAN-KETIMPANGAN-PEMBANGUNAN-EKONOMI-ANTAR-PROVINSI-DI-INDONESIA," *Jurnal Riset Ilmu Ekonomi*, vol. Vol 2 (1), no. 1, pp. 1–16, 2022. https://doi.org/10.23969/jrie.v2i1.23

[5] D. Jati and S. D. Purnomo, "DETERMINAN KETIMPANGAN PENDAPATAN DALAM UPAYA PEMERATAAN PEMBANGUNAN DI PULAU JAWA," *Jurnal Penelitian Inovatif*, vol. 3, no. 3, pp. 739–748, 2023, doi: https://doi.org/10.54082/jupin.760

[6] M. R. Maghriza and M. I. Hasmarini, "THE INFLUENCE OF THE NUMBER OF POOR PEOPLE, MINIMUM WAGE, HUMAN DEVELOPMENT INDEX, AND OPEN UNEMPLOYMENT RATE ON INCOME INEQUALITY IN YOGYAKARTA (2010-2023)," *Economos : Jurnal Ekonomi dan Bisnis*, vol. 7, no. 3, pp. 304–311, 2024, doi: https://doi.org/10.31850/economos.v7i3.3451

[7] Ramadhansyah Rayyan Effendy, Verina Araminda Prinary, Cupian Cupian, and Ardi Apriliadi, "DAMPAK ZIS, KORUPSI, DAN PENDIDIKAN TERHADAP TINGKAT KETIMPANGAN PENDAPATAN DI NEGARA BERKEMBANG: STUDI KASUS DI INDONESIA," *SANTRI : Jurnal Ekonomi dan Keuangan Islam*, vol. 3, no. 1, pp. 103–115, 2025, doi: https://doi.org/10.61132/santri.v3i1.1241

[8] K. Raziq and L. L. N. El Hasanah, "ANALISIS DETERMINAN KETIMPANGAN PENDAPATAN DI PROVINSI DAERAH ISTIMEWA YOGYAKARTA," *Jurnal Kebijakan Ekonomi dan Keuangan*, vol. 2, no. 1, pp. 12–21, 2023, doi: https://doi.org/10.20885/JKEK.vol2.iss1.art2

[9] P. R. Sihombing, K. A. Notodiputro, and A. Kurnia, "BIBLIOMETRIC MAPPING AND TREND ANALYSIS OF BETA REGRESSION MODELING : A DECADE OF," *Sinkron : Jurnal dan Penelitian Teknik Informatika*, vol. 9, no. 3, pp. 1062–1072, 2025, doi: https://doi.org/10.33395/sinkron.v9i3.14949

[10] D. Gujarati, "BASIC ECONOMETRICS BY GUJARATI," 2004, *McGraw-Hill Inc., Singapura*.

[11] P. R. Sihombing, K. A. Notodiputro, and B. Sartono, "COMPARISON OF GEE AND GLMM METHODS FOR LONGITUDINAL DATA (CASE STUDY: DETERMINANTS OF THE PERCENTAGE OF POOR PEOPLE IN INDONESIA, 2015-2019)," *AIP Conf Proc*, vol. 2563, no. October, pp. 2015–2019, 2022, doi: https://doi.org/10.1063/5.0103254

[12] Suryadiningrat and P. R. Sihombing, "KOMPARASI PEMODELAN LOGIT, PROBIT DAN CLOG-LOG PADA REGRESI BETA," *Nusantara Journal of Behavioral and Social Sciences*, vol. 1, no. 4, pp. 101–104, 2022, doi: https://doi.org/10.47679/202215

[13] P. R. Sihombing *et al.*, "KOMPARASI PEMODELAN REGRESI OLS GAUSSIAN, BETA DAN REGRESI FRACTIONAL PADA DATA RASIO," *Jurnal Bayesian : Jurnal Ilmiah Statistika dan Ekonometrika*, vol. 3, no. 2, pp. 193–199, 2023.

[14] J. van Niekerk, A. Bekker, and M. Arashi, "BETA REGRESSION IN THE PRESENCE OF OUTLIERS – A WIELDY BAYESIAN SOLUTION," *Stat Methods Med Res*, vol. 28, no. 12, pp. 3729–3740, 2019, doi: https://doi.org/10.1177/0962280218814574

[15] 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: https://doi.org/10.1080/0266476042000214501

[16] C. J. Swearingen, M. S. M. Castro, and Z. Bursac, "MODELING PERCENTAGE OUTCOMES: THE %BETA_REGRESSION MACRO," *SAS Global Forum 2011*, pp. 1–12, 2011.

[17] M. Hunger, A. Döring, and R. Holle, "LONGITUDINAL BETA REGRESSION MODELS FOR ANALYZING HEALTH-RELATED QUALITY OF LIFE SCORES OVER TIME," *BMC Med Res Methodol*, vol. 12, no. 1, pp. 1–12, 2012, doi: https://doi.org/10.1186/1471-2288-12-144

[18] A. P. Alencar, J. M. Singer, and F. M. M. Rocha, "COMPETING REGRESSION MODELS FOR LONGITUDINAL DATA," *Biometrical Journal*, vol. 54, no. 2, pp. 214–229, 2012, doi: https://doi.org/10.1002/bimj.201100056

[19] N. Koper and M. Manseau, "GENERALIZED ESTIMATING EQUATIONS AND GENERALIZED LINEAR MIXED-EFFECTS MODELS FOR MODELLING RESOURCE SELECTION," *Journal of Applied Ecology*, vol. 46, pp. 590–599, 2009. https://doi.org/10.1111/j.1365-2664.2009.01642.x

[20] H. Zhang and W. Lu, "REGULARIZED REGRESSION BETA TREE FOR RARE EVENT PREDICTION," *IEEE Trans Neural Netw Learn Syst*, vol. 3, no. 16, pp. 2006–2017, 2020.

[21] Y. Li *et al.*, "ADJUSTING FOR BASELINE ON THE ANALYSIS OF REPEATED BINARY RESPONSES WITH MISSING DATA," *Contemp Clin Trials*, vol. 39, no. 1, pp. 993–1007, 2012, doi: https://doi.org/10.1016/j.sapharm.2015.03.004

[22] M. Mittal, D. L. Harrison, D. M. Thompson, M. J. Miller, K. C. Farmer, and Y.-T. Ng, "AN EVALUATION OF THREE STATISTICAL ESTIMATION METHODS FOR ASSESSING HEALTH POLICY EFFECTS ON PRESCRIPTION DRUG CLAIMS," *Research in Social and Administrative Pharmacy*, vol. 12, no. 1, pp. 29–40, 2016, doi: https://doi.org/10.1016/j.sapharm.2015.03.004

[23] M. C. Pardo and T. Pérez, "ANALYSIS OF HOUSING PRICES BY GEE AND GLMM METHODOLOGIES: A LONGITUDINAL STUDY," *Appl Stoch Models Bus Ind*, vol. 29, no. 5, pp. 552–563, 2013, doi: 1 https://doi.org/10.1002/asmb.1940

[24] P. R. Sihombing, Erfiani, K. A. Notodiputro, A. Kurnia, and A. Info, "CHOOSING THE RIGHT TOOL : PRACTICAL CONSIDERATIONS FOR GLMM AND GEE IN LONGITUDINAL STUDIES , WITH A FOCUS ON DATA CHALLENGES," vol. 9, no. 1, pp. 37–44, 2025, doi: https://doi.org/10.30829/zero.v9i1.24602

[25] N. A. Hein, "BETA REGRESSION MODELS FOR REPEATED MEASURES DATA ANALYSIS," Omaha, 2019.

[26] M. Hunger, A. Döring, and R. Holle, "LONGITUDINAL BETA REGRESSION MODELS FOR ANALYZING HEALTH-RELATED QUALITY OF LIFE SCORES OVER TIME," *BMC Med Res Methodol*, vol. 12, 2012, doi: https://doi.org/10.1186/1471-2288-12-144

[27] H. Zhang, Q. Yu, C. Feng, D. Gunzler, P. Wu, and X. M. Tu, "A NEW LOOK AT THE DIFFERENCE BETWEEN THE GEE AND THE GLMM WHEN MODELING LONGITUDINAL COUNT RESPONSES," *J Appl Stat*, vol. 39, no. 9, pp. 2067–2079, 2012, doi: https://doi.org/10.1080/02664763.2012.700452

[28] P. R. Sihombing, K. A. Notodiputro, and B. Sartono, "COMPARISON OF GEE AND GLMM METHODS FOR LONGITUDINAL DATA (CASE STUDY: DETERMINANTS OF THE PERCENTAGE OF POOR PEOPLE IN INDONESIA, 2015-2019)," in *AIP Conference Proceedings*, A. L., R. null, P. R., P. G.E., and S. T.Y., Eds., BPS-Statistics Indonesia, Jl. Salemba Tengah No.36, RT.2/RW.4, Paseban, Senen, Daerah Khusus Ibukota, Jakarta, 10440, Indonesia: American Institute of Physics Inc., 2022. doi: https://doi.org/10.1063/5.0103254

[29] J. I. Figueroa-Zúñiga, R. B. Arellano-Valle, and S. L. P. Ferrari, “MIXED BETA REGRESSION: A BAYESIAN PERSPECTIVE,” *Comput Stat Data Anal*, vol. 61, pp. 137–147, 2013, doi: https://doi.org/10.1016/j.csda.2012.12.002

[30] L. Fontana, C. Masci, F. Ieva, and A. M. Paganoni, “PERFORMING LEARNING ANALYTICS VIA GENERALISED MIXED-EFFECTS TREES,” *Data (Basel)*, vol. 6, no. 7, 2021, doi: https://doi.org/10.3390/data6070074

[31] M. Fokkema, N. Smits, A. Zeileis, T. Hothorn, and H. Kelderman, “DETECTING TREATMENT-SUBGROUP INTERACTIONS IN CLUSTERED DATA WITH GENERALIZED LINEAR MIXED-EFFECTS MODEL TREES,” *Behav Res Methods*, vol. 50, no. 5, pp. 2016–2034, 2018, doi: https://doi.org/10.3758/s13428-017-0971-x

[32] Gg. A. M. C. Rini, N. L. P. Suciptawati, and I. A. P. A. Utari, “IDENTIFIKASI FAKTOR YANG MEMENGARUHI GINI RATIO DI INDONESIA,” *E-Jurnal Matematika*, vol. 11, no. 3, p. 160, 2022, doi: https://doi.org/10.24843/MTK.2022.v11.i03.p376.

[33] A. Erlando, F. D. Riyanto, and S. Masakazu, “FINANCIAL INCLUSION, ECONOMIC GROWTH, AND POVERTY ALLEVIATION: EVIDENCE FROM EASTERN INDONESIA,” *Heliyon*, vol. 6, no. 10, p. e05235, 2020, doi: https://doi.org/10.1016/j.heliyon.2020.e05235