

MODEL REGRESI GLMM-GEE-TREE UNTUK RESPON BERSEBARAN BETA DAN PENERAPANNYA DALAM ANALISIS DATA RASIO GINI DI INDONESIA

PARDOMUAN ROBINSON SIHOMBING



PROGRAM STUDI DOKTOR STATISTIKA DAN SAINS DATA
SEKOLAH SAINS DATA, MATEMATIKA, DAN INFORMATIKA
INSTITUT PERTANIAN BOGOR
BOGOR
2026



@Hak cipta milik IPB University

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.



PERNYATAAN MENGENAI DISERTASI DAN SUMBER INFORMASI SERTA PELIMPAHAN HAK CIPTA

Dengan ini, saya menyatakan bahwa disertasi dengan judul “Model Regresi GLMM-GEE-Tree untuk Respon Bersebaran Beta dan Penerapannya dalam Analisis Data Rasio Gini di Indonesia” merupakan hasil karya saya yang disusun di bawah bimbingan dosen pembimbing dan belum pernah diajukan dalam bentuk apa pun kepada perguruan tinggi lainnya. Semua sumber informasi yang digunakan, baik yang diterbitkan maupun tidak diterbitkan telah disebutkan dalam teks dan dicantumkan dalam Daftar Pustaka pada bagian akhir disertasi ini. Dengan ini saya melimpahkan hak cipta dari karya tulis saya kepada Institut Pertanian Bogor.

Bogor, Juli 2026

Pardomuan Robinson Sihombing

G1601202008

RINGKASAN

PARDOMUAN ROBINSON SIHOMBING. Model Regresi GLMM-GEE-Tree untuk Respon Bersebaran Beta dan Penerapannya dalam Analisis Data Rasio Gini di Indonesia. Dibimbing oleh ERFIANI, KHAIRIL ANWAR NOTODIPUTRO, dan ANANG KURNIA.

Pemodelan data respon berupa proporsi atau rasio yang ada pada selang terbuka (0,1) dengan struktur longitudinal menghadapi kompleksitas ganda berupa heterogenitas tidak teramati antarsubjek dan korelasi galat temporal. Pendekatan konvensional seperti *Generalized Linear Mixed Models* (GLMM) seringkali sensitif terhadap asumsi sebaran, sementara *Generalized Estimating Equations* (GEE) mengabaikan prediksi spesifik pada level subjek. Metode hubungan linier parametrik global seringkali gagal menangkap pola interaksi nonlinier yang umum ditemukan dalam data empiris. Penelitian ini bertujuan mengembangkan model GLMM-GEE-Tree untuk data respon bersebaran beta yang mengintegrasikan kekuatan prediksi GLMM, ketangguhan inferensi GEE, dan fleksibilitas algoritma pohon regresi.

Metodologi penelitian ini dilaksanakan melalui empat tahapan utama. Pertama, formulasi analitik model hibrida GLMM-GEE melalui pendekatan linearisasi *pseudolikelihood* untuk menduga parameter efek tetap dan efek acak. Kedua, penggabungan mekanisme penduga *sandwich* empiris untuk mengoreksi galat baku agar kekar (*robust*) terhadap salah spesifikasi struktur korelasi. Ketiga, integrasi algoritma partisi rekursif (*tree*) yang dibangun dari galat *pseudo-linear* yang telah terkalibrasi untuk menangkap hubungan non-linier. Keempat, validasi kinerja model melalui simulasi Monte Carlo dan penerapannya pada data empiris rasio gini di Indonesia periode 2018-2024.

Hasil kajian simulasi menunjukkan bahwa model GLMM-GEE-Tree menunjukkan kinerja yang baik. Model ini mampu mempertahankan akurasi prediksi (*Root Mean Square Error*) rendah setara dengan GLMM. Selain itu, model ini juga menjamin validitas inferensi (*Coverage Probability* mendekati level nominal 95%) setara dengan GEE, terutama pada kondisi heterogenitas dan autokorelasi tinggi sedangkan model standar lainnya mengalami penurunan performa signifikan. Algoritma pohon yang dikembangkan juga terbukti efektif dalam mengidentifikasi struktur interaksi kovariat secara tepat dan konsisten.

Penerapan model pada data rasio gini provinsi di Indonesia mengungkap adanya struktur interaksi non-linier dan efek ambang batas yang signifikan antara indikator pembangunan makroekonomi terhadap ketimpangan. Model mengidentifikasi beberapa kelompok wilayah dengan pola ketimpangan yang berbeda. Dampak peubah seperti pertumbuhan ekonomi dan tingkat kemiskinan terhadap rasio gini bersifat kondisional terhadap karakteristik spesifik wilayah tersebut. Temuan ini memberikan kontribusi teoritis bagi pengembangan statistika komputasional serta kontribusi praktis dalam penyusunan kebijakan intervensi yang lebih terlokalisasi dan berbasis data (*evidence-based policy*).

Kata kunci: Sebaran Beta, GEE, GLMM, *Regression Tree*, Rasio Gini.

SUMMARY

PARDOMUAN ROBINSON SIHOMBING. GLMM-GEE-*Tree* Regression Model for Beta *Distribution* Response and Its Application in the Analysis of Gini Ratio Data in Indonesia. Supervised by ERFIANI, KHAIRIL ANWAR NOTODIPUTRO, AND ANANG KURNIA.

Modeling of response data in the form of proportions or ratios at open intervals (0,1) with longitudinal structures faces double complexity in the form of unobserved heterogeneity between subjects and temporal error correlations. Conventional approaches such as Generalized Linear Mixed Models (GLMM) are often sensitive to distribution assumptions, while Generalized Estimating Equations (GEE) ignore specific predictions at the subject level. In addition, global linear-parametric relationships often fail to capture the non-linear interaction patterns commonly found in real data. This study aims to develop a GLMM-GEE-*Tree* model for beta dispersed response data that integrates the predictive power of GLMM, the robustness of GEE inference, and the flexibility of regression tree algorithms.

The methodology of this research was carried out through four main stages. First, the analytical formulation of the GLMM-GEE hybrid model through the Pseudo-Likelihood Linearization approach to estimate the parameters of fixed effects and random effects. Second, the incorporation of the Empirical Sandwich Estimator mechanism to correct standard error to be *robust* against the mis-specification of the correlation structure. Third, the integration of a recursive partition algorithm (*Tree*) built on a pseudo-linear residue that has been calibrated to capture non-linear relationships. Fourth, validation of model performance through Monte Carlo simulations and applications to empirical data on the Gini Ratio in Indonesia for the 2018-2024 period.

The results of the simulation study show that the GLMM-GEE-*Tree* model showed favorable performance under the evaluated scenarios. This model is able to maintain prediction accuracy (*low Root Mean Square Error*) equivalent to GLMM and ensure inference validity (*Coverage Probability* close to the nominal level of 95%) equivalent to GEE, especially in conditions of high heterogeneity and autocorrelation where other standard models experience significant performance degradation. The developed tree algorithm has also proven to be effective in accurately and consistently identifying the structure of covariate interactions.

The application of the model to the Gini Ratio data of provinces in Indonesia reveals the existence of a non-linear interaction structure and a significant threshold effect between macroeconomic development indicators on inequality. This model successfully identifies several different inequalities, where the impact of variables such as economic growth and poverty rates on the Gini Ratio is conditional on the specific characteristics of the region. These findings make a theoretical contribution to the development of computational statistics as well as a practical contribution to the formulation of more localized and evidence-based intervention policies.

Keywords: Beta Distribution, GEE, GLMM, *Regression Tree*, Gini Ratio.



Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

© Hak Cipta milik IPB, tahun 2026
Hak Cipta dilindungi Undang-Undang

Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan atau menyebutkan sumbernya. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik, atau tinjauan suatu masalah, dan pengutipan tersebut tidak merugikan kepentingan IPB.

Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apa pun tanpa izin IPB

MODEL REGRESI GLMM-GEE-TREE UNTUK RESPON BERSEBARAN BETA DAN PENERAPANNYA DALAM ANALISIS DATA RASIO GINI DI INDONESIA

PARDOMUAN ROBINSON SIHOMBING

Disertasi

Sebagai salah satu syarat untuk memperoleh gelar Doktor pada
Program Studi Statistika dan Sains Data

**PROGRAM STUDI DOKTOR STATISTIKA DAN SAINS DATA
SEKOLAH SAINS DATA, MATEMATIKA, DAN INFORMATIKA
INSTITUT PERTANIAN BOGOR
BOGOR
2026**



Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

Penguji Luar Komisi Pembimbing pada Ujian Tertutup Disertasi:

1. Prof. Dr. I Made Sumertajaya, M.Si. _____
2. Dr. Bagus Sartono, S.Si., M.Si. _____

Promotor Luar Komisi Pembimbing pada Sidang Promosi Terbuka Disertasi:

1. Dr. Bagus Sartono, S.Si., M.Si. _____
2. Dr. Ardi Adji, S.Si., M.Si. _____



Judul Disertasi : Model Regresi GLMM-GEE-Tree untuk Respon Bersebaran Beta dan Penerapannya dalam Analisis Data Rasio Gini di Indonesia
Nama : Pardomuan Robinson Sihombing
NIM : G1601202008

Disetujui oleh

Pembimbing 1:
Dr. Ir. Erfiani, M.Si.

Pembimbing 2:
Prof. Dr. Ir. Khairil Anwar Notodiputro, MS

Pembimbing 3:
Prof. Dr. Anang Kurnia, S.Si., M.Si.

Diketahui oleh

Ketua Program Studi:
Dr. Kusman Sadik, S. Si, M. Si
NIP. 196909121997021001

Dekan Sekolah Sains Data, Matematika, dan Informatika:
Prof. Dr. Ir. Agus Bueno, M. Si, M. Kom
NIP. 196607021993021001

Tanggal Ujian: 16 Juli 2026

Tanggal Lulus:

PRAKATA

Puji dan syukur penulis panjatkan kepada Tuhan Yang Maha Esa atas segala kasih dan karunia-Nya sehingga disertasi ini berhasil diselesaikan. Disertasi ini berjudul “Model Regresi GLMM-GEE-Tree untuk Respon Bersebaran Beta dan Penerapannya dalam Analisis Data Rasio Gini di Indonesia”. Disertasi ini penulis susun sebagai salah satu syarat untuk mendapatkan gelar Doktor Statistika dan Sains Data pada Sekolah Sains Data, Matematika, dan Informatika, Institut Pertanian Bogor.

Disertasi ini dapat diselesaikan berkat doa, bimbingan, masukan, saran dan bantuan dari banyak pihak. Oleh karena itu, dalam kesempatan ini penulis menyampaikan penghargaan dan terima kasih kepada:

1. Komisi pembimbing yang telah banyak memberikan bimbingan, arahan, masukan dan motivasi kepada penulis dengan penuh kesabaran: Ibu Dr. Ir. Erfiani, M.Si, Bapak Prof. Dr. Ir. Khairil Anwar Notodiputro, M.S, dan Bapak Prof. Dr. Anang Kurnia, S.Si., M.Si.
2. Penguji luar komisi Ujian Tertutup, Bapak Prof. Dr. I Made Sumertajaya, M.Si. dan Bapak Dr. Bagus Sartono, S.Si., M.Si. serta Penguji luar komisi Ujian Terbuka, Bapak Dr. Bagus Sartono, S.Si., M.Si. dan Bapak Dr. Ardi Adji, S.Si., M.Si. yang telah memberikan masukan serta saran perbaikan disertasi.
3. Jajaran pimpinan Badan Pusat Statistik (BPS) yang telah memberikan kesempatan kepada penulis untuk melanjutkan pendidikan S3 di Institut Pertanian Bogor.
4. Istri tercinta Ade Marsinta Arsani yang telah sabar dan selalu memberikan dukungan serta doa dalam proses penyelesaian studi, sedangkan anak-anak, Anastasia Kanyadiva Sihombing dan Alexander William Seamaruli Sihombing yang selalu menjadi penyemangat dalam proses penyelesaian studi.
5. Sekolah Sains Data, Matematika, dan Informatika IPB beserta jajarannya yang telah membantu memfasilitasi studi penulis.
6. Ketua Program Studi Doktor Statistika dan Sains Data IPB, serta Bapak dan Ibu Dosen yang telah banyak memberikan ilmu, pencerahan dan arahan selama masa pendidikan, serta sekretariat Program Studi Doktor Statistika dan Sains Data IPB yang telah banyak membantu administrasi dan informasi.
7. Teman-teman S3 angkatan 2020, dan teman-teman S3 dan S2 Program Studi Statistika dan Sains Data lainnya atas ketulusan doa, bantuan, semangat, dan kerja sama kepada penulis.
8. Keluarga besar penulis yang senantiasa mendukung, memotivasi, dan mendoakan kelancaran studi penulis dengan penuh ketulusan dan kesabaran.

Disertasi ini masih memiliki keterbatasan dan kekurangan. Namun demikian, penulis berharap penelitian ini dapat bermanfaat bagi semua pihak.

Bogor, Juli 2026

Pardomuan Robinson Sihombing

DAFTAR ISI

DAFTAR TABEL	xiii
DAFTAR GAMBAR	xiii
DAFTAR LAMPIRAN	xiv
I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Tujuan dan Manfaat	3
1.3 Kerangka Kerja Penelitian	4
1.4 Kebaruan (<i>Novelty</i>)	5
1.5 Batasan Penelitian	6
II TINJAUAN PUSTAKA	8
2.1 Pendekatan Regresi pada Data Longitudinal, Proporsi, dan Nonlinear	8
2.1.1 Model Linear Terampat dan Regresi Beta	8
2.1.2 Pendekatan Metode <i>Generalized Linear Mixed Model</i> (GLMM)	10
2.1.3 Pendekatan Metode Generalized Estimating Equations (GEE)	11
2.1.4 Pendekatan Metode Regresi Pohon (<i>Tree</i>)	13
2.2 Model Pohon Regresi untuk Data Respon Bersebaran Beta Berbasis GEE, GLMM, dan GLMM-GEE	15
2.3 Faktor Yang Mempengaruhi Rasio Gini Suatu Wilayah	17
2.4 Perkembangan Terkini Regresi Beta dalam Bibliometrik	18
2.5 Tinjauan Literatur Sistematis (<i>Systematic Literature Review</i> /SLR) Perbandingan Metode GLMM dan GEE Dalam Studi Longitudinal	21
III PERBANDINGAN KINERJA GLMM DAN GEE BERBASIS SEBARAN BETA DALAM PEMODELAN KETIMPANGAN EKONOMI	30
3.1 Pendahuluan	30
3.2 Data dan Sumber Data	32
3.3 Metode Penelitian	32
3.4 Hasil dan Pembahasan	34
3.5 Kesimpulan	38
IV MODEL REGRESI GLMM-GEE UNTUK RESPON BERSEBARAN BETA: KAJIAN ANALITIK DAN SIMULASI	39
4.1 Pendahuluan	39
4.2 Spesifikasi Model Sebaran Beta pada Data Panel	40
4.3 Penurunan Model GLMM Beta (Pendekatan Spesifik-Subjek)	41
4.4 Integrasi Inferensi Kekar (robust) (Tahap GEE)	43
4.5 Kajian Simulasi	45
4.6 Hasil dan Pembahasan Evaluasi Model Simulasi	48
4.7 Kesimpulan	51

V	PENGEMBANGAN MODEL GLMM-GEE-TREE PADA DATA RESPON BETA DENGAN VALIDASI ANALITIK DAN SIMULASI	53
5.1.	Landasan Filosofis dan Kerangka Integrasi	53
5.2.	Spesifikasi Model Probabilistik	54
5.3.	Konstruksi Model Hibrida GLMM-GEE-Tree	55
5.4.	Penurunan Analitik Penduga Parameter (<i>Pseudo-Likelihood</i>)	58
5.5.	Inferensi Statistik: Prosedur Iteratif GEE dan Derivasi Matriks Sandwich Robust	60
5.5.1	Prosedur Iteratif Koreksi GEE	60
5.5.2	Inferensi Statistik: Derivasi Matriks <i>Sandwich Robust</i>	61
5.6.	Properti Asimtotik Penduga Model Hibrida GLMM-GEE-Tree	62
5.7	Kajian Simulasi Model Pengembangan GLMM-GEE-Tree Pada Respon Bersebaran Beta	65
5.8	Hasil Simulasi Model Pengembangan GLMM-GEE-Tree Pada Respon Bersebaran Beta	67
5.9	Kesimpulan	71
VI	IMPLEMENTASI MODEL REGRESI GLMM-GEE-TREE PADA DATA EMPIRIS	72
6.1	Pendahuluan	72
6.2	Metodologi	74
6.3	Hasil dan Pembahasan	75
6.4.	Evaluasi Ketahanan Model terhadap Bias Data dan Guncangan Pandemi	80
6.5	Kesimpulan	81
6.6	Implikasi	81
VII	PEMBAHASAN UMUM	83
VIII	SIMPULAN DAN SARAN	89
8.1	Simpulan	89
8.2	Saran	89
	DAFTAR PUSTAKA	91
	LAMPIRAN	99
	RIWAYAT HIDUP	117

DAFTAR TABEL

1	Indikator Bibliometrik dan Teknik Analisis	18
2	Database Artikel Scopus dengan Kata Kunci GLMM dan GEE	25
3.	Deskriptif Pemilihan 10 Provinsi Terpilih di Indonesia	32
4	Peubah Bebas Penelitian	32
5	Perbandingan Metrik Akurasi RMSE dan R^2 antar Model Regresi Beta Pembanding	36
6	Estimasi Diagnostik Efek Acak dan Korelasi Sisaan Temporal	37
7	Skenario Simulasi Dalam Model Pembanding	47
8	Evaluasi Kinerja Lima Model pada Empat Skenario Simulasi Monte Carlo	49
9	Analisis Ragam Dua-Arah R^2 dan RMSE menurut Model dan Skenario	50
10	Uji Lanjut Model Hibrida GLMM-GEE terhadap Model Pembanding	50
11	Statistik Deskriptif R^2 dan RMSE Berdasarkan Model	67
12	Hasil ANOVA terhadap R^2	68
13	Hasil ANOVA terhadap RMSE	68
14	Uji Post Hoc Untuk R^2	69
15	Uji Post Hoc Untuk RMSE	69
16	Peubah Respon dan Peubah Bebas Penelitian.	75
17	Statistik Deskriptif Variabel Penelitian	76
18	Hasil Pengujian Asumsi Klasik	76
19	Evaluasi Kinerja Model Dengan Berbagai Kriteria	77
20	Hipotesis Pengaruh Peubah Bebas	79

DAFTAR GAMBAR

1	Kerangka Pengembangan Model	4
2	Peta Konsep Penelitian	7
3	Hubungan Antara Penulis dengan VosViewer	26
4	Visualisasi Jaringan Dengan VosViewer	27
5	Alur Kerja Pemodelan Regresi Beta dan Perbandingan Model Menggunakan Pendekatan GLM, GLMM, dan GEE	34
6	RMSE dan Pseudo R^2 untuk GLM, GLMM, dan GEE	35
7	Uji Pos Hoc R^2 Pada Skenario Kompleks	70
8	Selisih RMSE GLMM-GEE-Tree Terhadap Model Lainnya Pada Skenario Komplek	70
9	Flowchart Kajian Empiris Skenario Berbagai Model	75
10	Visualisasi Struktur Pohon GLMM-GEE-Tree	77



DAFTAR LAMPIRAN

1	Statistik Deskriptif R^2 dan RMSE Berdasarkan Model dan Skenario	99
2	Uji Post Hoc R^2 Berdasarkan Kombinasi Berbagai Skenario	100
3	Uji Post Hoc RMSE Berdasarkan Kombinasi Berbagai Skenario	102
4	Sintaks Software R Simulasi Model GLMM-GEE	104
5	Sintaks Software R Simulasi Model GLMM-GEE-Tree	112

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.