



RESEARCH ARTICLE

Komparasi Pemodelan Logit, Probit dan Clog-Log Pada Regresi Beta (Studi Kasus: Pengaruh IPG dan IDG Terhadap IKG di Indonesia Tahun 2020)

Suryadiningrat¹, Pardomuan Robinson Sihombing^{2*)}

Published online: 30 November 2022

Abstract: Penelitian ini bertujuan membandingkan pemodelan Indeks Ketimpangan Gender (IKG) menggunakan regresi beta dengan link function logit, probit dan clog-log. Adapun variabel independen yang digunakan adalah Indeks Pembangunan Gender (IPG) dan Indeks Pemberdayaan Gender (IDG). Data bersumber dari Badan Pusat Statistik. Ketiga pemodelan memberikan hasil yang sama baik untuk uji simultan dan parsial dalam pemodelan kasus pemodelan IKG. Model clog-log memberikan hasil terbaik dengan kriteria nilai error terkecil (AIC dan BIC) serta nilai koefisien determinasi terbesar. IPG berpengaruh signifikan negatif terhadap IKG sedangkan IDG belum berpengaruh signifikan negatif terhadap IKG. Berdasarkan hasil ini diharapkan para pemangku kebijakan dapat memberikan kebijakan yang komprehensif dan tepat sasaran dalam mengurangi ketimpangan gender di Indonesia.

Keywords: : clog-log, IDG, IKG, IPG, logit, probit

INTRODUCTION

Salah satu pemodelan statistik yang digunakan dalam menganalisis pengaruh antar variabel adalah metode regresi. Menurut Gujarati (2004), pemodelan regresi khususnya pada regresi Gaussian/ regresi klasik pada umumnya mengasumsikan bahwa data yang digunakan berdistribusi secara normal atau sering dikenal dengan *normal based regression*. Selain itu model regresi klasik juga memiliki asumsi non-heteroskedastis pada varians data dan non-autokorelasi pada kesalahan antar waktu. Pengasumsian ini terkadang tidak dapat terpenuhi, misalnya, dalam beberapa kasus data yang merupakan bilangan cacah, berbentuk kategori atau berupa proporsi/ rasio/ *rate*. Dalam hal ini, data-data tersebut mengikuti distribusi tertentu yang dikenal dengan distribusi keluarga eksponensial. Menurut Agresti (2002), pemodelan pada data yang berdistribusi keluarga eksponensial adalah dengan model *Generalized Linear Model* (GLM). Adapun salah satu pemodelan dalam GLM adalah regresi beta, dimana data dengan variabel dependen yang digunakan berupa data rasio atau proporsi. Data rasio/ proporsi ini nilainya berada dalam interval 0 sampai 1. Menurut Johnson & Kotz (1995),

regresi beta menggunakan pendekatan distribusi beta, dimana distribusi ini sangat fleksibel dalam berbagai fenomena ketidakpastian. Lebih lanjut menurut Swearingen (2011), pemodelan dengan regresi beta akan memberikan estimator parameter yang akurat dan efisien dibandingkan dengan metode *least squares* biasa ketika variabel respon yang diamati tidak simetris dalam distribusi atau masalah heteroskedastisitas.

Di dalam pemodelan GLM terdapat tiga fungsi penghubung yaitu model logit, probit dan clog-log. Perbandingan pemodelan ini telah banyak dilakukan oleh peneliti-peneliti lainnya seperti Ratnasari & Putri (2015) melakukan perbandingan ketiga model fungsi penghubung pada studi kasus Partisipasi Perempuan dalam Pembangunan Ekonomi di Kalimantan Selatan. Hasil yang didapat ketiga model memberikan hasil yang tidak signifikan berbeda, akan tetapi pemodelan probit memiliki nilai AIC terkecil R Square terbesar. Lebih lanjut Marizal et al (2015) mengaplikasikan ketiga model pada Pemodelan Kawasan Panen Kelapa Sawit Malaysia dengan hasil bahwa model clog-log menghasilkan nilai AIC terkecil.

Salah satu pemodelan yang dapat diaplikasikan dalam model regresi beta adalah data Indeks Ketimpangan Gender (IKG). IKG merupakan salah satu permasalahan yang menjadi perhatian di samping permasalahan kemiskinan, ketimpangan pendapatan, pertumbuhan ekonomi, pengangguran dan pembangunan manusia. Data IKG berkisar diantara 0 dan 1. Penelitian tentang IKG telah dilakukan diantaranya Aktaria dan Handoko (2012) dengan mengaitkan ketimpangan gender dalam pertumbuhan ekonomi. Hasil analisis penelitian menunjukkan pengaruh

BPS-Statistics Indonesia

*) *corresponding author*

Pardomuan Robinson Sihombing

Email: robinson@bps.go.id

negatif dan signifikan antara ketimpangan gender dengan pertumbuhan ekonomi. Lebih lanjut Marsono (2021) menganalisis model spasial pada indeks ketimpangan gender di Indonesia. Hasil yang didapat IPG berpengaruh terhadap IKG sedangkan IDG belum berpengaruh signifikan terhadap IKG.

Berdasarkan latar belakang di atas belum terdapat kajian performa *link function* logit, probit dan clog-log pada regresi beta, maka penulis tertarik untuk membandingkan ketiga metode tersebut dalam studi kasus pengaruh IPG dan IDG terhadap IKG. Kriteria untuk memilih model terbaik didasarkan pada nilai AIC dan BIC terkecil dan r-square terbesar.

METHODS

Penelitian ini menggunakan data yang dipublikasikan dari Badan Pusat Statistik (2021). Referensi waktu penelitian adalah 2020. Variabel penelitian yang digunakan dapat dilihat pada Tabel 1.

Tabel1.
Variabel penelitian

Variabel	Satuan
Indeks Ketimpangan Gender (IKG)	Poin
Indeks Pembangunan Gender (IPG)	Poin
Indeks Pemberdayaan Gender (IDG)	Poin

Model Regresi Beta

Regresi beta model digunakan jika data mengikuti distribusi beta, dimana nilainya antara nol hingga satu. Menurut Walpole (2012), fungsi distribusi beta dapat ditulis sebagai berikut:

$$f(y; a, b) = \frac{\Gamma(a + b)}{\Gamma(a)\Gamma(b)} y^{a-1} (1 - y)^{b-1}$$

dengan $0 < y < 1$; $a > 0$, $b > 0$, dan $(.)$ adalah fungsi gamma. Persamaan dalam regresi beta adalah:

Tabel 2. Uji Kebaikan Model dan Hipotesis

Uji Goodness of Fit	Hipotesis Null	Hipotesis Alternatif	Tolak Ho
Uji Simultan/ (Uji X ²)	Model Tidak Sesuai/ Semua variabel tidak berpengaruh	Model sesuai/ minimal 1 variabel berpengaruh signifikan	P Value < 0.05
Uji Parsial/ Uji Z	Variabel independent ke-i tidak berpengaruh	Variabel independent ke-i berpengaruh	P Value < 0.05

RESULTS AND DISCUSSIONS

Sebelum membahas lebih lanjut pemodelan dalam analisis regresi, Tabel 3 menyajikan statistik deskriptif untuk masing-masing variabel penelitian. Rata-rata IKG sebesar 0.41poin, dengan nilai terendah sebesar 0.07 poin pada Provinsi DI Yogyakarta, dengan nilai tertinggi sebesar 0.53 poin pada Provinsi NTB. Rata-rata IPG 90.23, dengan nilai terendah sebesar 79.59 poin pada Provinsi Papua,

Dalam pemodelan regresi beta terdapat tiga pendekatan *link function* (Hardin & Hilbe, 2007) yaitu:

a. Model Regresi Beta Logistik

$$g(\mu) = \text{logit}(\mu) = \ln \left[\frac{\mu}{1-\mu} \right] = \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p \quad 2$$

$$\text{Dengan } \mu = \frac{e^{(x^T \beta)}}{1 + e^{(x^T \beta)}}$$

b. Model Regresi Beta Probit/ Normit

$$g(\mu) = \Phi(\mathbf{Z}) = \Phi(\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p) \quad 3$$

dengan: $\Phi = \text{probability density function (PDF)} = \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}}$

c. Model Regresi Complementary Log-Log (Interval Censored Survival)

$$g(\mu) = 1 - \frac{1}{\exp(\exp[\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p])} = \frac{1 - \exp(-\exp[\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p])}{1 - \exp(-\exp[\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p])} \quad 4$$

Kriteria untuk memilih model terbaik

Pada penelitian ini, pemilihan model didasarkan pada AIC kriteria (Akaike, 1974) dan BIC. (Gideon Schwarz, 1978) Rumus yang digunakan adalah:

$$\begin{aligned} AIC &= -2 L(\hat{\theta}) + 2p & 5 & 5 \\ BIC &= -2 L(\hat{\theta}) + p \ln(n) & 6 & 6 \end{aligned}$$

$L(\hat{\theta})$ di sini adalah nilai kemungkinan likelihood, dan p adalah jumlah parameter yang akan diperkirakan, termasuk konstanta. Model terbaik adalah model yang memiliki nilai AIC dan BIC terkecil $L(\hat{\theta})$ (Widarjono, 2007).

Selanjutnya dilakukan pengujian kebaikan model. Adapun pengujian kebaikan model menurut Gujarati (2004) dapat dilihat pada Tabel 2.

dengan nilai tertinggi sebesar 94.80 poin pada provinsi DI Yogyakarta. Rata-rata IDG 69.48 poin, dengan nilai terendah sebesar 51.96 poin pada provinsi NTB, dengan nilai tertinggi sebesar 82.41 poin pada Provinsi Kalimantan Tengah.

Tabel3. Analisis deskriptif variabel penelitian

Variabel	Rata-rata	Std	Min	Maks
IKG	0,41	0,10	0,07	0,53
IPG	90,23	3,36	79,59	94,80
IDG	69,48	6,85	51,96	82,41

Selanjutnya dalam pemodelan regresi diharapkan tidak terjadi korelasi yang tinggi antar variabel independen. Hasil ini ditunjukkan oleh nilai Variant Inflation Factor (VIF) dan nilai toleransi (1/VIF). Model yang baik tidak mengandung nilai korelasi yang tinggi antara variabel independen dengan nilai referensi VIF 10 dan toleransi 0,1. Pada Tabel 4, semua nilai VIF variabel independen kurang dari sepuluh, dan nilai toleransi > 0,1 sehingga model yang digunakan tidak mengalami multikolinearitas.

Tabel4. Nilai VIF variabel independen

Variabel	VIF	1/VIF
IPG	1,09	0,9194
IDG	1,09	0,9194

Tabel5. Uji Asumsi Klasik

Ujian	Nilai Uji	Nilai Probabilita	Kesimpulan
Normality Test	7,37	0,025	Data Tidak Normal
Uji Breusch Pagan	5,70	0,017	Heterokedastis
Ramsey Test	5,62	0,003	Pola tidak linear

Tabel 6. Perbandingan model regresi normal dan beta

Variabel	OLS	logit	probit	cloglog
IPG	-0,014**	-0,068**	-0,043**	-,048**
IDG	-0,001	-0,0072	-0,004	-0,005
C	1,845***	6,316***	3,960***	4,153**
scale		3,230***	3,233**	3,209**
Chi/F	6,11	11,331	11,452	10,637
p	0,005	,003	,003	,004
AIC	-65,680	-55,945	-56,066	-55,251
BIC	-61,101	-49,840	-49,961	-49,145
R ²	0,2828	0,271	0,273	0,283

Catatan: * p<0,05; ** p<0,01

Pada Tabel 5 dapat dilihat pengujian asumsi klasik untuk memastikan bahwa penggunaan regresi berbasis normal/ OLS tidak efisien dan tidak memenuhi asumsi yang digunakan. Terlihat bahwa baik untuk uji normalitas, heterokedastisitas dan linearitas nilai probabilita kurang dari alpha, dalam hal ini kurang dari 5 persen.

Selanjutnya pada pada Tabel 6 menunjukkan perbandingan antara model regresi menggunakan model regresi berdasarkan distribusi normal dan regresi beta dengan 3 alternatif *link function*. Terlihat bahwa jika dilihat dari uji simultan dan banyaknya variabel yang signifikan (uji parsial) ketiga model menghasilkan performa yang sama baiknya. Di sisi lain dilihat dari nilai AIC dna BIC terkecil dan *r square* terbesar maka model regresi beta dengan *link function* clog-log memberikan performa terbaik. Hal ini dikarenakan model clog-log mengakomodir adanya data yang *rare cases*. Hal ini juga senada dengan penelitian Marizal et al (2015) yang menyatakan model clog-log memberikan performa terbaik. Selanjutnya yang akan diintrepetasikan adalah model clog-log sebagai model terbaik.

HASIL DISKUSI

Nilai koefisien determinasi sebesar 0,283 artinya variabel IKG dapat dijelaskan oleh IPG dan IDG sebesar 28,3

persen sisanya oleh variabel lain di luar. IPG berpengaruh signifikan negatif terhadap IKG, artinya kenaikan IPG akan menurunkan IKG. Sementara IDG belum berpengaruh negatif terhadap IKG. Hal ini senada dengan penelitian Marsono (2021) yang menyatakan IPG berpengaruh terhadap IKG sedangkan IDG belum berpengaruh signifikan terhadap IKG.

CONCLUSION AND RECOMMENDATION

Dari pembahasan di atas dapat disimpulkan bahwa model regresi beta lebih baik dibandingkan model dengan distribusi normal. Jika dibandingkan ketiga *link function* maka model clog-log memberikan performa terbaik, hal ini dicerminkan oleh AIC dan BIC yang lebih kecil dan *r square* terbesar. Hasil penelitian menunjukkan IPG berpengaruh signifikan negatif terhadap IKG sedangkan IDG belum berpengaruh signifikan negatif terhadap IKG. Berdasarkan hasil ini diharapkan dapat diberikan kebijakan yang komprehensif dan tepat sasaran dalam mengurangi ketimpangan gender di Indonesia.

Untuk penelitian lebih lanjut, dimungkinkan untuk menambahkan variabel lain yang berpotensi mempengaruhi IKG seperti rasio gini, kemiskinan, pertumbuhan ekonomi, pengangguran dan lainnya. Di sisi

lain, penelitian selanjutnya dapat menggunakan aplikasi model regresi lainnya, seperti model regresi beta dengan data panel atau menggunakan model regresi fractional.

REFERENCES

- Agresti, A. (2002). *Categorical Data Analysis*. New York. Inc. John Wiley and Sons.
- Akaike, H. (1974). A New Look at the Statistical Model Identification. *IEEE Transactions on Automatic Control*, 19(6), 716–723. <https://doi.org/10.1109/TAC.1974.1100705>
- Aktaria, E., & Handoko, B. S. (2012). Ketimpangan Gender Dalam Pertumbuhan Ekonomi. *Jurnal Ekonomi Pembangunan: Kajian Masalah Ekonomi Dan Pembangunan*, 13(2), 194. <https://doi.org/10.23917/jep.v13i2.168>
- BPS. (2021). *Kajian Penghitungan Indeks Ketimpangan Gender 2021*.
- Gideon Schwarz. (1978). Estimating The Dimension of a Model. *The Annals of Statistics*, 6(2), 461–464.
- Gujarati, D. (2004). *Basic Econometrics BY Gujarati* (pp. 1–1002). McGraw-Hill Inc.
- Hardin, J. ., & Hilbe, J. M. (2007). *Generalized Linear Models and Extensions*. Stata Corp.
- Johnson, N., Kotz, S., & Balakrishnan, N. (1995). *Continuous Univariate Distributions*. Wiley.
- Marizal, M., Nazam, N. B., Yendra, R., Desvina, A. P., Matematika, J., Islam, U., Sultan, N., Kasim, S., & Matematika, J. (2015). *Pemodelan Kawasan Panen Kelapa Sawit Malaysia Menggunakan Model Logistik*. November, 435–440.
- Marsono, M. (2021). Deteksi Spasial Pada Model Indeks Ketimpangan Gender Indonesia. *BUANA GENDER: Jurnal Studi Gender Dan Anak*, 6(1). <https://doi.org/10.22515/bg.v6i1.3482>
- Ratnasari, V., & Putri, R. C. (2015). Pemodelan Logit, Probit dan Complementary Log-Log pada Studi Kasus Partisipasi Perempuan dalam Pembangunan Ekonomi di Kalimantan Selatan. *Jurnal Sains Dan Seni ITS*, 4(2), 181–186. <https://media.neliti.com/media/publications/15676-ID-pemodelan-logit-probit-dan-complementary-log-log-pada-studi-kasus-partisipasi-pe.pdf>
- Swearingen, C. J., Castro, M. S. M., & Bursac, Z. (2011). Modeling percentage outcomes: the %beta_regression macro. *SAS Global Forum 2011*, 1–12.
- Walpole, R. E. (2012). *Probability & Statistics for Engineers & Scientists*. Pearson.
- Widarjono, A. (2007). *Ekonometrika: Teori dan Aplikasi untuk Ekonomi dan Bisnis*. Ekonosia Fakultas Ekonomi Universitas Islam Indonesia.