MODEL REGRESI LOGISTIK KEKAR PADA KINERJA PEREKONOMIAN MENGGUNAKAN ESTIMASI BIANCO-YOHAI DAN BIANCO-YOHAI TERBOBOTI

*Robust Logistic Regression Model on Economic Performance Using Bianco-Yohai
and Weighted Biaco-Yohai Estimation*

Lia Yuliana¹, Farid Ridho², Rudi Salam³

Politeknik Statistika STIS^{1,2,3}

Jalan Otista Raya No. 64C Jakarta Timur 13330

E-mail: lia@stis.ac.id

ABSTRAK

Metode pendugaan kemungkinan maksimum atau *maximum likelihood estimator* (MLE) biasanya digunakan untuk menduga parameter model regresi logistik karena efisiensinya di bawah model parametrik. Namun, bukti menunjukkan bahwa metode ini sangat rentan terhadap amatan terluar atau pencilan. Metode regresi logistik kekar digunakan untuk memperbaiki masalah ini. Tulisan ini membahas kinerja pendugaan kemungkinan maksimum dan dua penduga kekar yaitu Bianco dan Yohai (BY) dan BY terboboti atau *weighted BY* (WBY). Sebagai terapan akan digunakan data kinerja perekonomian provinsi di Indonesia sebagai peubah respon. Sedangkan untuk peubah penjelas digunakan persentase penduduk yang mengalami keluhan kesehatan dan angka partisipasi murni tingkat SLTP. Dengan melihat statistik *chi-square* dan kedekatan nilai dugaan kekar dengan MLE-Clean, penduga kekar BY adalah lebih baik dari WBY. Hasil dari pendugaan BY didapatkan bahwa bahwa peubah yang memengaruhi kinerja perekonomian adalah persentase penduduk yang mengalami keluhan kesehatan.

Kata kunci: kinerja perekonomian, penduga kekar, regresi logistik

ABSTRACT

The method of estimating the maximum likelihood or maximum likelihood estimator (MLE) is usually used to estimate the parameters of the logistic regression model because the efficiency is below the parametric model. However, the evidence shows that this method is very vulnerable to outward observation or outliers. The robust logistic regression method was used to correct this problem. This paper discusses the performance of the maximum likelihood estimation and two sturdy estimators, namely Bianco and Yohai (BY) and weighted BY (WBY). As an application, provincial economic performance data in Indonesia will be used as a response variable. Meanwhile, the explanatory variables used are the percentage of the population experiencing health complaints and the net enrollment rate at the junior secondary level. By looking at the chi-square statistic and the approximate value of the firmness estimate with MLE-Clean, the estimator of BY is better than WBY. The result of BY estimation shows that the variable affecting economic performance is the percentage of the population experiencing health complaints.

Keywords: economic performance, robust estimator, logistic regression

PENDAHULUAN

Model regresi logistik telah banyak diterapkan pada bidang kesehatan dan ilmu-ilmu sosial. Model ini digunakan untuk menggambarkan pengaruh dari peubah penjelas terhadap peubah respon biner. Model regresi logistik mengasumsikan peubah respon mengikuti sebaran Bernoulli yang saling bebas yang bernilai 1 (untuk menyatakan kejadian sukses) dan 0 (untuk menyatakan kejadian tidak sukses). Jika terdapat suatu peubah respon biner Y dan suatu vektor peubah penjelas $\mathbf{x} = (x_1, \dots, x_p)^T$, maka bentuk dari model regresi logistik biner adalah (Ahmad, Ramli, Midi. 2010):

$$P(Y = 1 | X = x_i) = F(x_i^T \boldsymbol{\beta}) = \frac{\exp(x_i^T \boldsymbol{\beta})}{1 + \exp(x_i^T \boldsymbol{\beta})}, \quad i = 1, 2, \dots, n \quad (1)$$

dimana $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_p)^T$ adalah vektor parameter dan F adalah fungsi sebaran kumulatif yang sangat naik. Untuk menduga parameter $\boldsymbol{\beta}$ biasanya digunakan metode MLE karena efisiensinya pada model parametrik. Dugaan dari MLE dapat diperoleh dengan memaksimumkan persamaan berikut (Ahmad, Ramli, Midi. 2010)

$$\hat{\boldsymbol{\beta}}_{MLE} = \arg \max_{\boldsymbol{\beta}} \sum_{i=1}^n l(Y_i, x_i; \boldsymbol{\beta}) \quad (2)$$

di mana $l(Y_i, x_i; \boldsymbol{\beta}) = Y_i \ln F(x_i^T \boldsymbol{\beta}) + (1 - Y_i) \ln [1 - F(x_i^T \boldsymbol{\beta})]$

yang nantinya merupakan prosedur yang efisien secara asimtotik dalam menduga $\boldsymbol{\beta}$. Cara lain untuk menduga MLE adalah dengan meminimumkan *deviance*, $D_i(\boldsymbol{\beta})$, (Ahmad, Ramli, Midi. 2010) yaitu

$$\hat{\boldsymbol{\beta}}_{MLE} = \arg \min_{\boldsymbol{\beta}} \sum_{i=1}^n D_i(\boldsymbol{\beta}) \quad (3)$$

di mana $D_i(\boldsymbol{\beta}) = Y_i [-\ln F(x_i^T \boldsymbol{\beta})] + (1 - Y_i) \{-\ln [1 - F(x_i^T \boldsymbol{\beta})]\}$

Dengan melakukan penurunan (2) terhadap $\boldsymbol{\beta}$ akan dihasilkan persamaan skor kemungkinan (Ahmad, Ramli, Midi. 2010)

$$\sum_{i=1}^n [Y_i - F(x_i^T \boldsymbol{\beta})] x_i = 0 \quad (4)$$

Solusi dari persamaan tersebut dipecahkan secara iteratif menggunakan metode Newton-Raphson atau *Scoring* Fisher. Namun, hasil dugaan dari MLE sangat sensitif atau mudah dipengaruhi oleh amatan terluar dan karena itu dikatakan tidak kekar (Bianco, Martinez. 2009; Ahmad, Ramli, Midi. 2010). Pregibon (1981, 1982) telah memberikan contoh mengenai sifat sensitif dari MLE terhadap pencilan. Oleh karena itu, untuk mengatasi permasalahan ini diperlukan beberapa pilihan metode pendugaan parameter yang kekar yang sangat sedikit dipengaruhi oleh keberadaan pencilan.

Secara umum, terdapat dua pendekatan utama untuk membuat MLE menjadi kekar pada regresi logistik. Pertama adalah pendekatan yang menggunakan pembobotan pada fungsi skor kemungkinan pada (4), yang disebut kelas Mallows (Mallows, 1975; Hampel, et al., 1986; Stefanski, et al., 1986; Kunsch et al., 1989). Pendekatan kekar yang kedua adalah seperti yang

diusulkan oleh Pregibon (1982), Copas (1988), Bianco dan Yohai (1996), dan Croux dan Haesbroeck (2003). Pendekatan ini bekerja menggunakan fungsi objektif pada (3). Namun tulisan ini bertujuan menerapkan dan membandingkan metode pendugaan regresi logistik kekar dari pendekatan fungsi objektif, termasuk membandingkan dengan MLE. Sebagai terapan, akan digunakan data kinerja perekonomian provinsi di Indonesia sebagai peubah respon.

METODE

Pada pendekatan kekar yang menggunakan fungsi objektif, dimulai dari Pregibon (1982) yang bekerja secara langsung dengan fungsi objektif pada (3). Pregibon (1982) mengganti fungsi deviance pada (4) dengan suatu penduga kekar yang didefinisikan dengan

$$\hat{\beta} = \arg \min_{\beta} \sum_{i=1}^n \lambda \left[D_i(x_i^T \beta, y_i) \right] \quad (5)$$

dimana λ adalah fungsi Huber yang sangat naik. Penduga ini dibuat untuk memberikan pembobot yang kecil untuk amatan yang dianggap tidak baik oleh model. Namun, penduga ini tidak menurunkan penimbang pada pengamatan yang berpengaruh pada ruang rancangan dan tidak konsisten. Oleh karena itu, Bianco dan Yohai (1996) meningkatkan metode ini dan hasilnya adalah konsisten dan lebih kekar dibandingkan penduga Pregibon dengan mendefinisikan

$$\beta = \arg \min_{\beta} \sum_{i=1}^n \left\{ \rho \left[D(x_i^T \beta, y_i) + G(F(x_i^T \beta)) + G(1 - F(x_i^T \beta)) \right] \right\} \quad (6)$$

Nilai ρ yang dipilih oleh Bianco dan Yohai (1996) adalah tertutup, bisa diturunkan, dan merupakan fungsi tidak turun yang didefinisikan

$$\rho(x) = \begin{cases} x - (x^2/2k) & , \text{ jika } x \leq k \\ k/2 & , \text{ selainnya} \end{cases} \quad (7)$$

dimana k adalah bilangan positif, $G(x) = \int_0^x \psi(-\ln u) du$ dan $\psi(x) = \rho'(x)$, tetapi ditekankan juga bahwa dimungkinkan untuk menggunakan nilai ρ yang lain. Croux dan Haesbroeck (2003) memperluas penduga BY dengan memasukkan pembobot untuk mengurangi pengaruh dari pengamatan terluar pada ruang kovariat. BY terboboti (WBY) dinamakan juga penduga Croux dan Haesbroeck (CH) yang didefinisikan sebagai berikut

$$\beta = \arg \min_{\beta} \sum_{i=1}^n w(x_i) \left\{ \rho \left[D(x_i^T \beta, y_i) + G(F(x_i^T \beta)) + G(1 - F(x_i^T \beta)) \right] \right\} \quad (8)$$

dimana pembobot $w(x_i)$ supaya menjadi fungsi menurun dari jarak Mahalanobis kekar, dihitung menggunakan penduga *Minimum Covariance Determinant* (MCD) yaitu

$$w(x_i) = \begin{cases} 1 & \text{jika } RD_i^2 \leq \chi_{p,0.975}^2 \\ 0 & \text{selainnya} \end{cases} \quad (9)$$

Penduga WBY ini tetap konsisten karena pembobotan hanya dikenakan pada peubah prediktor. Tetapi, prosedur pembobotan yang digunakan juga mengurangi pembobot dari titik *leverage* yang baik, yang seharusnya tidak perlu, sehingga bisa menyebabkan hilangnya efisiensi.

Sebagai terapan, akan digunakan data kinerja perekonomian provinsi di Indonesia tahun 2018 sebagai peubah respon dimana kode 0 artinya melambat dan kode 1 artinya lebih cepat. Sedangkan peubah predictor yang digunakan adalah persentase penduduk yang mengalami keluhan kesehatan (x1) dan angka partisipasi murni (APM) SLTP (x2). Amatan pada penelitian ini adalah seluruh provinsi yang ada di Indonesia. Data yang digunakan adalah data sekunder yang berasal dari publikasi Badan Pusat Statistik (BPS) tahun 2018.

Tabel 1. Daftar peubah respon dan peubah penjelas

Peubah	Keterangan
Peubah respon	
- Kinerja perekonomian	0 : Melambat 1 : Lebih cepat
Peubah penjelas:	
- Persentase penduduk yang mengalami keluhan kesehatan	Kontinu
- Angka partisipasi murni (APM) SLTP	Kontinu

Sumber: Diolah

Pada penelitian sebelumnya yang dilakukan oleh Ahmad, Ramli, Midi. (2010) dan Croux, Haesbroeck. (2003), dengan menggunakan data vaso dan data leukemia untuk membandingkan kedua metode pendugaan. Hasil dengan menggunakan data vaso, diperoleh bahwa BY lebih baik dari WBY, sedangkan pada data leukemia hasilnya adalah WBY lebih baik dari BYF.

Sampai dengan tahun 2018, Indonesia merupakan negara berkembang dengan tingkat pertumbuhan ekonomi yang stabil. Perekonomian Indonesia didorong oleh beberapa faktor seperti konsumsi masyarakat, investasi dan ekspor. Tetapi indikator paling mendasar dari pertumbuhan ekonomi adalah pendidikan dan kesehatan. Pendidikan dan kesehatan merupakan dua hal yang harus menjadi prioritas pemerintah karena keduanya berkaitan dengan kualitas Sumber Daya Manusia (SDM). Jika suatu negara tidak memiliki pendidikan yang baik maka akan menjadi negara terbelakang. Sama halnya dengan kesehatan, jika kondisi kesehatan masyarakatnya kurang baik maka negara tersebut akan tertinggal.

Siddique, H M A. Mohey-ud-din, G. dan Kiani, A. (2018) menyebutkan bahwa pendidikan tingkat menengah dan tinggi berkontribusi pada peningkatan pertumbuhan ekonomi. Dampak harapan hidup, kematian bayi, partisipasi pendidikan menengah dan tinggi terhadap pertumbuhan ekonomi lebih kuat di negara-negara berpenghasilan menengah atas dibandingkan dengan negara-negara berpenghasilan menengah ke bawah. Selain itu, Bills dan Klenow (2000), Self dan Grabowski (2004) menyebutkan bahwa pendidikan berkorelasi positif dengan tingkat pertumbuhan per kapita Produk Domestik Bruto di seluruh negara.

Pengaruh kesehatan terhadap pertumbuhan ekonomi ditunjukkan oleh beberapa penelitian seperti Aka, B. F., & Dumon, J. C. (2008) menyebutkan bahwa dalam jangka panjang proses pertumbuhan ekonomi dijelaskan oleh kesehatan masa lalu dan tingkat pendidikan. Penelitian yang dilakukan Barro, R.J. (2013) menggunakan data antar negara, terdapat hubungan parsial antara pertumbuhan ekonomi dan harapan hidup dimana hubungan antara status kesehatan dan pertumbuhan ekonomi adalah positif.

HASIL DAN PEMBAHASAN

Sebelum melakukan pendugaan menggunakan metode kekar, perlu dilakukan eksplorasi terhadap data, apakah memang data memuat pencilan sehingga memang perlu digunakan metode regresi logistik kekar. Pendeteksian pencilan pada regresi logistik bisa menggunakan plot dari sisaan Pearson. Gambar 1 adalah hasil plot antara nomor amatan dan sisaan Pearson dari data kinerja perekonomian. Dari Gambar 1 terlihat bahwa diduga terdapat pencilan pada data yang ada (titik dengan lingkaran merah). Dengan kondisi data seperti ini, maka metode regresi logistik kekar bisa digunakan.

Gambar 1. Plot antara nomor amatan dan sisaan Pearson

Setelah laik untuk digunakan metode kekar, akan diilustrasikan perilaku dari kedua metode pendugaan parameter regresi logistik kekar yang digunakan dalam penelitian yaitu pendugaan BY dan WBY. Hasil dari pendugaan koefisien dan galat baku akan disajikan untuk data asli dan data modifikasi. Data modifikasi adalah data asli dengan pengamatan pencilan yang sudah dihilangkan. Penduga yang baik adalah penduga yang dugaan parameternya dekat dengan penduga MLE dari data modifikasi (data tanpa pencilan). Kordzakhia et al (2001) menyarankan kriteria lain untuk mengevaluasi kinerja dari berbagai penduga kekar. Kriteria tersebut adalah statistik chi-square berdasarkan transformasi arcsin χ^2_{arc} (Dobson, 1990) yang didefinisikan sebagai

$$\chi^2_{arc} = \sum_{i=1}^n 4 \left[\arcsin \sqrt{y_i} - \arcsin \sqrt{\pi_i} \right]^2 \text{ dimana } \sqrt{\pi_i} \text{ adalah peluang fit untuk } i = 1, 2, \dots, n. \text{ Semakin}$$

kecil nilai χ^2_{arc} semakin baik kebaikan suai.

Tabel 2. Parameter duga, galat baku, dan ukuran kebaikan suai kinerja ekonomi provinsi untuk penduga MLE, MLE-Clean, BY dan WBY

Metode Pendugaan	Intersep		Keluhan kesehatan		Apm SLTP		χ^2_{arc}
	Koef.	g.b.	Koef.	g.b.	Koef.	g.b.	
MLE	-3.76	4.80	-0.18	0.10	0.12	0.07	71.37
MLE-Clean	-7.36	5.60	-0.33	0.14	0.22	0.10	68.60
BY	-4.26	4.90	-0.20	0.12	0.13	0.08	70.66
WBY	-8.90	21.07	-0.14	0.10	0.17	0.26	72.45

Sumber: Diolah

Tabel 1 menampilkan parameter duga dan galat baku duga untuk berbagai prosedur termasuk MLE-Clean yang adalah duga MLE setelah pengamatan (#19 = NTT) dikeluarkan. Dari

hasil pengolahan diketahui juga bahwa peubah yang signifikan memengaruhi kinerja perekonomian adalah persentase penduduk yang mengalami keluhan kesehatan. Dari hasil ini pula maka penduga yang baik adalah BY karena mempunyai dugaan parameter dekat dengan MLE-Clean dan memberikan chi-square yang paling kecil.

Tabel 3. Estimasi koefisien regresi logistik.

	Koefisien	Std. Error	p-Value
Konstanta	-4,26	4,90	0,39
X1	-0,20	0,12	0,09
X2	0,13	0,09	0,12

Sumber: Diolah

Berdasarkan hasil uji parsial, dapat dilihat bahwa nilai p-value (0,000) dari X1 lebih kecil dari tingkat signifikansi (α) = (0,1), sehingga keputusan tolak H_0 . Dengan demikian dapat disimpulkan bahwa variabel persentase penduduk yang mengalami keluhan kesehatan memiliki pengaruh signifikan terhadap kinerja perekonomian provinsi di Indonesia tahun 2018. Hal ini sejalan dengan penelitian Bloom, D. E., Canning, D., & Sevilla, J. (2004). yang menyebutkan bahwa kesehatan yang baik memiliki efek positif, cukup besar, dan secara statistik signifikan pada output agregat.

Sedangkan nilai p-value (0,000) dari X2 lebih besar dari tingkat signifikansi (α) = (0,1), sehingga keputusan tidak dapat menolak H_0 . Dengan demikian dapat disimpulkan bahwa angka partisipasi murni (APM) SLTP tidak berpengaruh signifikan terhadap kinerja perekonomian provinsi di Indonesia tahun 2018. Hal ini sejalan dengan penelitian Arham, M.A. dan Dai, S I S. (2019) yang menyebutkan bahwa kinerja pendidikan setiap provinsi di Indonesia yang diukur dari APM SLTP dan tamatan SLTA dapat meningkatkan pertumbuhan ekonomi dengan asumsi bahwa keterampilan mereka cukup memadai untuk meningkatkan produktivitas.

Berdasarkan tabel uji parsial, nilai odds ratio persentase penduduk yang mengalami keluhan kesehatan adalah 0,819. Hal tersebut menunjukkan bahwa setiap peningkatan 1 persen penduduk yang mengalami keluhan kesehatan, memiliki kecenderungan bahwa kinerja perekonomian provinsi di Indonesia tahun 2018 melambat sebesar 0,819 kali dengan asumsi variabel lainnya konstan.

KESIMPULAN

Metode regresi logistik kekar dapat digunakan untuk mengatasi masalah jika terjadi amatan yang mengalami pencilan. Penelitian ini membandingkan antara metode pendugaan kemungkinan maksimum atau *maximum likelihood estimator* (MLE) dan dua penduga kekar yaitu Bianco dan Yohai (BY) dan BY terboboti atau *weighted BY* (WBY).

Penerapan metode ini dilakukan pada data kinerja perekonomian provinsi di Indonesia sebagai peubah respon dan dua peubah penjelas yaitu persentase penduduk yang mengalami keluhan kesehatan dan angka partisipasi murni tingkat SLTP. Dengan melihat statistik *chi-square* dan kedekatan nilai dugaan kekar dengan MLE-Clean, penduga kekar BY adalah lebih baik dari WBY. Hasil dari pendugaan BY didapatkan bahwa bahwa peubah yang memengaruhi kinerja perekonomian adalah persentase penduduk yang mengalami keluhan kesehatan.

Berdasarkan hasil uji parsial, menunjukkan bahwa variabel persentase penduduk yang mengalami keluhan kesehatan memiliki pengaruh signifikan terhadap kinerja perekonomian provinsi di Indonesia tahun 2018. Sedangkan angka partisipasi murni (APM) SLTP tidak berpengaruh signifikan terhadap kinerja perekonomian provinsi di Indonesia tahun 2018. Setiap peningkatan 1 persen penduduk yang mengalami keluhan kesehatan, memiliki kecenderungan bahwa kinerja perekonomian provinsi di Indonesia tahun 2018 melambat sebesar 0,819 kali dengan asumsi variabel lainnya konstan.

DAFTAR PUSTAKA

Ahmad, S., Ramli, NM., Midi, H. (2010). Robust Estimators in Logistic Regression: A Comparative Simulation Study, *Journal of Modern Applied Statistical Methods*: Vol. 9: Iss. 2, Article 18: 502-511.

- Aka, B. F., & Dumon, J. C. (2008). Health, Education and Economic Growth: Testing for Long- Run Relationships and Causal Links. *Applied Econometrics and International Development Development*, 8(2), 101-113.
- Arham, M.A. dan Dai, S I S. (2019). Analisis Pembiayaan, Kinerja Pendidikan, dan Pertumbuhan Ekonomi di Indonesia. *Jurnal Economia*. Vo. 15, No. 2, Oktober 2019, 292-305. Website: <https://journal.uny.ac.id/index.php/economia>.
- Barro, R. J. (2013). Health and Economic Growth. *Annals of Economics and Finance*, 14(2), 305-342. [http://doi.org/10.1016/S1574-0684\(05\)01006-3](http://doi.org/10.1016/S1574-0684(05)01006-3).
- Bianco, A.M., Martinez, E. (2009). Robust testing in the logistik regression model. *Computational Statistics & Data Analysis*; 53:4095–4105.
- Bils, M. and Klenow, P. J. (2000), Does Schooling Cause Growth?. *The American Economic Review*. 90(5), pp. 1160-1183.
- Bloom, D. E., Canning, D., & Sevilla, J. (2004). The effect of health on economic growth: A production function approach. *World Development*, 32(1), 1-13. <http://doi.org/10.1016/j.worlddev.2003.07.002>
- Dobson, A.J. (1990). An Introduction to Generalized Linear Models. Chapman & Hall, London.
- Kordzakhia, N., Mishra, G.D., Reiersolmoen, L. (2001). Robust estimation in the logistik regression model. *Journal of Statistical Planning and Inference*, 98, 211-223.
- Pregibon, D. (1981). Logistic regression diagnostics. *Annals of Statistics*, 9, 705-724.
- Pregibon, D. (1982). Resistant fits for some commonly used logistik models with medical applications. *Biometrika*, 73, 413-425.
- Self, S. dan Grabowski, R. (2004), „Does education at all levels cause growth? India, a case study. *Economics of Education Review*, 23, pp. 47-55.
- Siddique, H M A. Mohey-ud-din, G. dan Kiani, A. (2018). Health, Education and Economic Growth Nexus: Evidence from Middle Income Countries *Global Social Sciences Review (GSSR)* URL: [http://dx.doi.org/10.31703/gssr.2018\(III-IV\).05](http://dx.doi.org/10.31703/gssr.2018(III-IV).05) p-ISSN 2520-0348, e-ISSN 2616-793X DOI: 10.31703/gssr.2018(III-IV).05 Vol. III, No. IV (Fall 2018) Page: 68 – 86.