UIN IMAM BONJOL PADANG

Math Educa Journal 6(2) (2022): 179-189

MATH EDUCA

Jurnal Matematika dan Pendidikan Matematika

Website: http://ejournal.uinib.ac.id/jurnal/index.php/matheduca

Email: mej.uinibpadang@gmail.com

ESTIMASI PARAMETER REGRESI LOGISTIK DATA PANEL EFEK TETAP UNTUK T=2

[1]Umi Yuliatin, [2]Dedi Rosadi, [3]Ezhari Asfa'ani

[1] Prodi Teknik Instrumentasi Kilang, PEM Akamigas,Indonesia

[2] Prodi Statistika, Universitas Gajah Mada, Indonesia

[3]Prodi Matematika , UIN Imam Bonjol Padang, Indonesia

E-mail: [1] umi.yuliatin@esdm.go.id , [2] dedirosadi@gajahmada.edu , [3] ezhariasfaani@uinib.ac.id

Received: August 2022; Accepted: September 2022 Published: October 2022

Abstract

Logistic regression is a dichotomous classification method that uses several mathematical concepts in the estimating of variables parameters. In the estimation of parameter using the MLE (Maximum Likelihood Estimation) estimation method are obtained by Newton Raphson's numerical method. Unfortunately, this estimation doesn't work in binary panel data with fixed effects for time T=2 because the present of fixed effec α_i *. Thus, Conditional MLE is used to provide consistent estimator of* β *. This estimation shows by sample data* N=1.151 *obtained* $\hat{\beta}_{MLE} = -0.566$ *while the discussion shows the parameter values are at* $\beta = -0.288$.

Keywords: *Panel Data, Fixed Effect, Newton Raphson, MLE, Conditional MLE*

Abstrak

Regresi logistik merupakan salah satu metode klasifikasi dikotomi yang menggunakan beberapa konsep matematis dalam proses estimasi parameter variabel independen yang mempengaruhi. Dalam estimasi parameter variabel independen digunakan metode MLE (*Maximum likelihood Estimation*) yang diperoleh dengan bantuan metode numerik Newton Raphson. Namun dalam susunan data panel dengan efek tetap α_i untuk waktu T=2 estimator MLE memberikan

estimator yang tidak konsisten. Dengan demikian digunakanlah Conditional MLE dalam mengestimasi parameter variabel independen. Dalam paper ini ditunjukkan bahwa estimator Conditional MLE memberikan estimator yang konsisten pada analisis data kategorik respon biner dengan efek tetap α_i . Hal ini ditunjukkan pada simulasi data dengan N=1.151 diperoleh nilai $\hat{\beta}_{MLE} = -0.566$ sedangkan berdasarkan uraian nilai parameter menunjukkan $\beta = -0.288$.

Kata kunci: Data Panel, Efek Tetap, MLE, Conditional MLE

Pendahuluan

Regresi logistik digunakan untuk memodelkan hasil dari variabel dependen

*Corresponding author.

p-ISSN: 2580-6726

e-ISSN: 2598-2133

kategoris. Untuk variabel kategorik tidak tepat menggunakan regresi linier karena nilai respon tidak diukur pada skala rasio dan akan menghasilkan eror yang tidak berdistribusi normal. Selain itu prediksi variabel dependen kategorik hanya dapat mengambil sejumlah nilai dalam rentang tertentu sebagaimana dilakukan pada regresi logistik dengan respon biner.

Regresi logistik biasa dengan respon biner dijalankan pada data *cross section* dengan melibatkan sejumlah sampel (N) dan beberapa variabel independen $x_1, x_2, \ldots, x_k$ (Liski, 2007). Diantaranya penelitian yang menggunakan metode ini adalah Thomas Pentury dkk (Pentury et al., 2016) yang menggunakan regresi logistic untuk menentukan akreditasi SMA di kota Ambon dengan factor jumlah guru sebagai factor yang signifikan, Muinah dkk (Kusnul Kotimah & Pingit Wulandari, 2014) juga menggunakan regresi logistik untuk stratifikasi partisipasi ekonomi perempuan di Jawa Timur, sedangkan Intan (Muflihah, 2017) menggunakan metode ini untuk menentukan factor-faktor yang mempengaruhi *financial distress* .

Di sisi lain sejumlah penelitin yang menggunakan regresi logistik belum memperhatikan perubahan data kategorik yang diamati dari waktu ke waktu (Rosadi, 2006). Sedemikian sehingga masih terdapat informasi yang hilang dalam pengamatan. Pengamatan dari waktu ke waktu ini akan menjadi satu kesatuan bersama variabel independennya menjadi pengamatan data panel.

Data panel atau data longitudinal menjelaskan dua informasi yang berasal dari *cross section* individu dan time series yang merefleksikan perubahan terhadap waktu (Baltagi, 2005). Jika *N* individu diamati dalan *T* waktu, maka banyak observasi yang dimiliki adalah *NT*. Sedangkan pengamatan yang sama juga dapat dilakukan pada data respon biner. Pengamatan yang dilakukan dua kali (T=2) pada sampel yang sama berikut bersama variabel respond dan kovariatnya menarik untuk dibahas. Bagaimanakah estimator MLE (*Maximum Likelihood Estimation*) yang juga dipakai dalam estimasi parameter pada regresi logistik biasa dapat diterapkan pada data panel? Hal tersebut menjadi menarik sebab pada data panel terdapat dua jenis *effect*, yaitu *fixed effect* dan juga *random effect* baik itu berasal dari cross section maupun berasal dari waktu. Dengan demikian estimasi parameter dai variabel independen menjadi menarik untuk dibahas dan ditelusuri.

Dasar Teori

Regresi logistik adalah regresi yang menggunakan dua nilai yang berbeda untuk menyatakan variabel respon y_i, yaitu nilai 0 (gagal) dan nilai 1 (sukses) yang diobservasi melalui variabel latent kontinu y_i^* yaitu (Agresti, 2009) :

$$y_i^* = x_i'\beta + \epsilon_i \qquad \dots(1)$$

Sedangkan,

$$y_i = \begin{cases} 1, & jika \quad y_i^* > 0 \\ 0, & jika \quad y_i^* \leq 0 \end{cases} \qquad \dots..(2)$$

Dengan memandang e_i sebagai variabel random berdistribusi logistik standar, sehingga $y_i = 1$ jika

$$x_i'\beta + \epsilon > 0$$

$$\epsilon > -(x_i'\beta)$$

dan

$$P(y_i = 1|x_i) = f\left(\epsilon < -(x_i'\beta)\right)$$

$$= \int_{-x_i'\beta}^{\infty} \frac{e^{\epsilon}}{1+e^{\epsilon}} d\epsilon$$

$$= \frac{e^{x_i'\beta}}{1+e^{x_i'\beta}} \qquad \dots(3)$$

Dengan cara yang sama diperoleh

$$P(y_i = 0 | x_i) = \frac{1}{1+e^{x_i'\beta}}$$

dengan $x_i = (x_{i1}, x_{i2}, \dots, x_{ip})$ adalah vektor kovariat dan $\beta = (\beta_1, \beta_2, \dots, \beta_p)$ adalah vektor parameter. Dilain pihak suatu transformasi logit $F(x'\beta)$ yang ekuivalen dengan model logistik didefinisikan:

$$\text{Logit}(x_i'\beta) = \ln\left(\frac{F(x_i'\beta)}{1-F(x_i'\beta)}\right)$$

$$= x_{i1}\beta_1 \dots + x_{ip}\beta_p$$

$$= (x_i'\beta)$$

Jelas bahwa logit$(x_i'\beta)$ linier dalam parameter β .

Misalkan suatu sampel terdiri dari N observasi dari pasangan (x_i, y_i) dan model regresi logistik (3) maka penduga likelihood β yang memaksimumkan

$$L(\beta) = \prod_{i=1}^{n} [F(x_i'\beta)]^{y_i}[1 - F(x_i'\beta)]^{1-y_i} \dots(4)$$

Dengan mengambil log dari fungsi loglikelihoodnya maka di estimator MLE untuk β adalah yang memaksimalkan fungsi berikut:

$$\sum_i x_i[y_i - F(x_i'\beta)] = 0 \qquad \dots(5)$$

Iterasi newton Raphson digunakan untuk memperpoleh β yang memaksimumkan (4) sebagaimana dijabarkan (Czepiel, 2012) yang diperoleh dengan menggunakan turunan pertama (matrik gradient) dan turunan kedua dalam bentuk matrik Hessian. Matrik gradient dan Hessian diperoleh dengan kaidah penurunan fungsi vector sebagaimana diulas pada (Kafadar & Schott, 1997). Algoritma newton Raphson pada fungsi non linier juga dibahas (Santiyasa, 2009) untuk keperluan komputasi baik merupakan fungsi dari vector maupun fungsi satu-satu.

Turunan pertama dari fungsi vektor ini sering disebut dengan matrik gradien. β yang memenuhi persamaan tersebut didapatkan secara numerik dengan metode Newton Raphson (Czepiel, 2012). Metode ini berjalan dengan diberikannya *initial value* $\beta^{(0)}$.

$$\beta^{(k+1)} = \beta^{(k)} - \left(\frac{\partial^2 L}{\partial\beta\partial\beta'}\Big|_{\beta=\beta^{(k)}}\right)^{-1} \frac{\partial L}{\partial\beta}\Big|_{\beta=\beta^{(k)}}$$

Iterasi berhenti ketika $\beta^{(k+1)} - \beta^{(k)} \approx 0.$ Matrik Hessian yang merupakan derivatif kedua diberikan:

$$\frac{\partial^2 \ln L}{\partial\beta\partial\beta'} = -\sum_i f(x_i'\beta)x_i x_i' \quad \dots\dots\dots (6)$$

Model Panel Linier

Data panel secara sederhana dapat dibawa ke dalam model linier yang standar yaitu:

$$y_{it} = \mu + x_{it}'\beta + u_{it} \quad \dots (7)$$

dimana i: 1,2,.., N menyatakan indeks individu atau cross section, t=1,… ,T menyatakan waktu. μ menyatakan skalar, β berukuran p x 1 dan x_{it} adalah observasi waktu ke-t dan cross section ke-i dari p variabel independen (Baltagi, 2005).

Dengan α_i adalah efek individu maka u_{it} dalam model satu arah dinyatakan dengan

$$u_{it} = \alpha_i + \epsilon_{it}$$

dimana λ_t menyatakan efek waktu ke-t . Efek tersebut dibedakan menjadi dua, yaitu efek random dimana α_i atau λ_t mengikuti distribusi normal dengan mean 0 dan variansi tertentu, dan efek tetap (fixed effect) salah satu atau keduanya memiliki nilai yang tetap.

Metode yang digunakan untuk memperoleh estimator β yang konsisten adalah dengan transformasi linier (Q_{FE}) (Kafadar & Schott, 1997) untuk model dengan efek tetap. Transformasi ini dibutuhkan untuk mendapatkan estimator yang konsisten untuk β. Transformasi tersebut equivalen dengan mengurangkan setiap observasi pada masing-masing variabel dengan rata-ratanya dan ditujukan untuk mengeliminasi α_i.

Estimator β yang konsisten diperoleh:

$$\hat{\beta}_{QFE} = (X'\Omega X)^{-1} X' \Omega^{-1} Y \quad \dots (8)$$

dengan Ω adalah matrik proyektor yang simetrik dan idempoten. Pada (Astuti, 2010) estimator ini ditulis dalam bentuk notasi penjumlahan matri lainnya.

Sedangkan estimator untuk α_i adalah

$$\hat{\alpha}_i = \bar{y}_i - \bar{x}_i \beta_{QFE,} \qquad i = 1, 2, \dots\dots, N \quad \dots (9)$$

dengan $\bar{x}_i = \dfrac{\sum_{t=1}^{T} x_{it}}{T}$

HASIL PENELITIAN DAN PEMBAHASAN

Estimasi Regresi Logistik Data Panel Respon Biner dengan Efek Tetap

y_{it}^* adalah variabel random laten kontinu fungsi linier dari variabel penjelas x_{it}. Dengan mengasumsikan efek individu α_i fixed (tetap), sehingga :

$$y_{it}^* = x_{it}'\beta + \alpha_i + \epsilon_{it} \quad \dots (10)$$

Selanjutnya y_{it} diobservasi melalui variabel laten y_{it}^* sebagai berikut:

$$y_{it} = \begin{cases} 1 & jika\ y_{it}^* > 0 \\ 0 & jika\ y_{it}^* \leq 0 \end{cases} \quad \dots (11)$$

Dengan ϵ_{it} berdistribusi logistik standar sebagaimana pada persamaan (3) diperoleh

$$P(y_{it} = 1 | x_{it}, \alpha_i) = \frac{e^{x_{it}'\beta + \alpha_i}}{1 + e^{x_{it}'\beta + \alpha_i}} \quad \dots (12)$$

Sedangkan

$$P(y_{it} = 0 | x_{it}, \alpha_i) = \frac{1}{1 + e^{x_{it}'\beta + \alpha_i}} \quad \dots (13)$$

Ketika $T \to \infty$ maka estimator MLE konsisten (Smith & Hsiao, 1988). Sedangkan sering ditemui T kecil dalam data panel, sehingga jumlah observasi terbatas untuk mengestimasi α_i. Apapun estimasi untuk α_i menjadi tidak berarti jika jumlah individu yang terlibat sangat besar, $N \to \infty$. Sehingga jauh lebih penting untuk mengetahui parameter β , sebab hal ini menjadi rumit ketika estimator MLE menjadi tidak konsisten dengan keberadaan α_i dalam model.

Metode Maximum Likelihood

y_{it} adalah variabel random kategorik yang menggambarkan kejadian bernoulli (Subanar, 2012) dengan masing-masing probabilitas sebagaimana disebutkan dalam persamaan (12) dan (13) sehingga diperoleh

$$P(y_{it} \mid \beta, \alpha_i) = F(x_{it}^{'}\beta + \alpha_i)^{y_{it}} \left[1 - F(x_{it}^{'}\beta + \alpha_i)\right]^{1-y_{it}} \quad \ldots (14)$$

Parameter β diestimasi dengan memaksimaumkan fungsi likelihood. Fungsi likelihood dibentuk oleh perkaliaan dari fungsi probabilitas (14) untuk setiap sampel. Parameter β yang memaksimumkan fungsi likelihood disebut dengan estimator MLE (Bain & Engelhardt, 1993).

Fungsi likelihood yang dibentuk oleh (14) adalah

$$L(\beta, \alpha_i) = [F(x_{it}^{'}\beta + \alpha_i)]^{\sum_i \sum_t y_{it}} \times \left[1 - F(x_{it}^{'}\beta + \alpha_i)\right]^{\sum_i \sum_t (1-y_{it})} \quad \ldots (15)$$

β yang memaksimumkan adalah estimator yang juga memaksimumkan fungsi Log-nya. Fungsi loglikelihood (15) menjadi:

$$\ln L = \sum_i \sum_t y_{it} \ln[F(x_{it}'\beta + \alpha_i)] + \sum_i \sum_t (1 - y_{it}) \ln[1 - F(x_{it}'\beta + \alpha_i)] \quad \ldots \quad (16)$$

dengan

$$\ln[F(x_{it}'\beta + \alpha_i)] = (x_{it}'\beta + \alpha_i) - \ln\left[1 + e^{x_{it}'\beta + \alpha_i}\right] \quad \ldots (17)$$

dan

$$\ln[1 - F(x_{it}'\beta + \alpha_i)] = -\ln[1 + e^{x_{it}'\beta + \alpha_i}]$$

Maka persamaan (16) menjadi

$$\ln L = \sum_i \sum_t y_{it}[(x_{it}'\beta + \alpha_i)] - \sum_i \sum_t [\ln(1 + e^{x_{it}'\beta + \alpha_i})] \quad \ldots (18)$$

Turunan pertama terhadap β adalah :

$$\frac{\partial \ln L}{\partial \beta} = \sum_i \sum_t x_{it}[y_{it} - \frac{e^{x_{it}'\beta + \alpha_i}}{1 + e^{x_{it}'\beta + \alpha_i}}] \quad \ldots (19)$$

Turunan pertama terhadap α_i adalah:

$$\frac{\partial \ln L}{\partial \alpha_i} = \sum_t [y_{it} - \frac{e^{x_{it}'\beta + \alpha_i}}{1 + e^{x_{it}'\beta + \alpha_i}}] \quad \ldots (20)$$

METODE PENELITIAN

Data, Intrumen, dan Teknik Pengumpulan Data

Data yang digunakan dalam penelitian ini terdiri dari 1151 perempuan yang sedang berada dalam usia 14-17 tahun (pada wawancara pertama) di Amerika yang diwawancari tiap tahun. Pada analisis data ini digunakan data dengan panjang 5 tahun dan *balance* sehingga keseluruhan observasi yang ada adalah 5755 sebagaimana sumber data (Center for Human Resource Research, 2002).

MLE Pada Data Regresi Logistik Data Panel dengan T=2

Untuk kasus sederhana diperhatikan untuk T=2 dan satu variabel independen, dengan satu variabel penjelas $x_{i1} = 0$ dan $x_{i2} = 1$, maka α_i yang memenuhi (20) menjadi :

$$\hat{\alpha}_i = \begin{cases} \infty \quad , \textit{jika } y_{i1} + y_{i2} = 0 \\ -\infty \ , \textit{jika } y_{i1} + y_{i2} = 0 \\ -\dfrac{\beta}{2} \ , \textit{jika } y_{i1} + y_{i2} = 1 \end{cases} \quad \dots (21)$$

Misalkan

n_0 : *jumlah individu dengan* $y_{i1} + y_{i2} = 0$
n_1 : *jumlah individu dengan* $y_{i1} + y_{i2} = 1$
n_2 : *jumlah individu dengan* $y_{i1} + y_{i2} = 2$

Sehingga β yang menyelesaikan persamaan (19) adalah

$\hat{\beta} = [\ln(\sum_i y_{i2} - n_2) - \ln(n_1 + n_2 - \sum_i y_{i2}]$ ... (22)

Berdasarkan LLN *(Law of Large Number)*:

$$p \lim_{N\to\infty} \frac{1}{N}\left(\sum_i y_{i2} - n_2\right)$$

$$= \frac{1}{N}\sum_{i=1}^{N} \frac{e^{\beta+\alpha_i}}{(1+e^{\alpha_i})(1+e^{\beta+\alpha_i})} \quad \dots (23)$$

dan

$$p \lim_{N\to\infty} \frac{1}{N}\left(n_1 + n_2 - \sum_i y_{i2}\right) =$$

$$= \frac{1}{N}\sum_{i=1}^{N} \frac{e^{\alpha_i}}{(1+e^{\alpha_i})(1+e^{\beta+\alpha_i})} \quad \dots (24)$$

Dengan mensubtitusikan $\hat{\alpha}_i = -\frac{\beta}{2}$ ke persamaan (23) dan (24) maka dari persamaan (24) diperoleh

$$p \lim_{N\to\infty} \hat{\beta} = 2\beta \quad \dots (25)$$

Sehingga $\hat{\beta}$ tidak konsisten dan diperoleh estimator yang baru

$$\beta^* = \frac{\hat{\beta}}{2} \quad \dots (26)$$

Conditional Maximum Likelihood untuk T=2

Conditional Maximum Likelihood Estimation berangkat dari probabilitas bersyarat yang dalam hal ini adalah $\sum_t y_{it}$. Beberapa asumsi yang digunakan adalah y_{it} independen didalam individu i atau antar individu i. Untuk T=2 nilai $\sum_{t=1}^{2} y_{it}$ memberikan nilai yang mungkin muncul adalah $\{0,1,2\}$. Jika diberikan $\sum_{t=1}^{2} y_{it} = 0{,}2$, maka:

$$P[(y_{i1}, y_{i2}) = (0{,}0), (1{,}2) | y_{i1} + y_{i2} = 0{,}2] = 1$$

Untuk $\sum_{t=1}^{2} y_{it} = 1$, maka probabilitas bersyarat :

$$P(y_{i1} = 0, y_{i2} = 1 | y_{i1} + y_{i2} = 1)$$

$$= \frac{P(y_{i1}=0, y_{i2}=1), (y_{i1}+y_{i2}=1)}{P[(y_{i1}+y_{i2}=1)]}$$

$$= \frac{P(y_{i1}=0)}{P(y_{i1}+y_{i2}=1)} \frac{P(y_{i2}=1)}{P(y_{i1}+y_{i2}=1)} \quad \dots (27)$$

Sedangkan

$$P(y_{i1} + y_{i2} = 1 | \beta, \alpha_i) = P(y_{i1} = 0, y_{i2} = 1 | \beta, \alpha_i)$$
$$(P(y_{i1} = 1, y_{i2} = 0 | \beta, \alpha_i)$$

$$= \frac{e^{x_{i2}'\beta+\alpha_i} + e^{x_{i1}'\beta+\alpha_i}}{(1+e^{x_{i1}'\beta+\alpha_i})(1+e^{x_{i2}'\beta+\alpha_i})} \dots (28)$$

Maka persamaan (27) menjadi

$$P(y_{i1}=0,y_{i2}=1|y_{i1}+y_{i2}=1)$$
$$=\frac{e^{x'_{i2}\beta+\alpha_i}}{e^{x'_{i1}\beta+\alpha_i}+e^{x'_{i2}\beta+\alpha_i}}$$
$$=\frac{e^{x'_{i2}\beta}}{e^{x'_{i1}\beta}+e^{x'_{i2}\beta}} \quad \ldots(29)$$

Dengan demikian diperoleh

$$P(y_{i1}=0,y_{i2}=1|y_{i1}+y_{i2}=1)$$
$$=\frac{e^{x'_{i2}\beta}}{e^{x'_{i1}\beta}+e^{x'_{i2}\beta}}$$
$$=F[(x_{i2}-x_{i1})'\beta]\ldots(30)$$

dan

$$\mathrm{P}(y_{i1}=1,y_{i2}=0|y_{i1}+y_{i2}=1)$$
$$=1-F[(x_{i2}-x_{i1})'\beta]\ldots(31)$$

Tampak kedua probabilitas bersama ini tidak bergantung pada α_i.

Misalkan :

$$w_i=\begin{cases}0 & ,\, jika\;\; (y_{i1},y_{i2})=(1,0)\\ 1 & ,\, jika\;\; (y_{i1},y_{i2})=(0,1)\end{cases} \quad \ldots(32)$$

Maka persamaan (30) dan (31) adalah fungsi regresi logistik biasa dengan kovariat $(x_{i2}-x_{i1})$. Dan w_i adalah kejadian Bernoulli dengan fungsi densitas sebagai berikut:

$$P(w_i|\beta)=[F((x_{i2}-x_{i1})'\beta)]^{w_i}\;\;[1-F((x_{i2}-x_{i1})'\beta)]^{1-w_i} \quad \ldots(33)$$

Dengan $\tilde{B}_1=\{i|y_{i1}+y_{i2}=1\}$, maka dengan memaksimumkan fungsi *conditional log-likelihood*

$$\ln L(\beta)=\textstyle\sum_{i\in\tilde{B}_1}\{w_i\ln F[(x_{i2}-x_{i1})'\beta]+(1-w_i)\ln(1-F[(x_{i2}-x_{i1})'\beta])\} \quad \ldots(34)$$

memberikan estimator yang konsisten untuk β, yaitu matrik gradien konvergen ke-0 ketika dievaluasi terhadap β yang benar.

$$\textstyle\sum_{i\in\widetilde{B_1}}(x_{i2}-x_{i1})[y_i-F((x_{i2}-x_{i1})'\beta)]=0 \quad \ldots(35)$$

Sedangkan titik maksimum dicapai dengan koreksi dari matrik Hessian sebagaimana pada persamaan (6) yaitu:

$$J_{\tilde{B}_1}=\frac{\partial^2\ln L}{\partial\beta\partial\beta'}=-\textstyle\sum_i d_i[f((x_{i2}-x_{i1})'_i\beta)]\,(x_{i2}-x_{i1})(x_{i2}-x_{i1})' \quad \ldots(37)$$

dimana d_i didefinisikan sebagai berikut;

$$d_i=\begin{cases}0 & ,\text{jika}\quad y_{i1}+y_{i2}=0\\ 1 & ,\text{jika}\quad y_{i1}+y_{i2}=1\end{cases} \quad \ldots(38)$$

Matrik informasi yang merupakan harga harapan dari matrik Hessian adalah:

$$J=E\left(J_{\tilde{B}_1}\right)$$
$$=-\textstyle\sum_i P_i[f((x_{i2}\,x_{i1})'_i\beta)]\;\;(x_{i2}-x_{i1})(x_{i2}-x_{i1})' \quad \ldots\;(39)$$

Dimana

$$P_i=E(d_i|\alpha_i)$$
$$=\frac{e^{x'_{i1}\beta+\alpha_i}+e^{x'_{i2}\beta+\alpha_i}}{\left(1+e^{x'_{i1}\beta+\alpha_i}\right)\left(1+e^{x'_{i2}\beta+\alpha_i}\right)}$$

Karena tidak ada estimator yang konsisten untuk α_i akibatnya matrik J sulit untuk dievaluasi. Selain itu jika persamaan (30) dan (31) dipandang sebagai regresi logistik biasa maka perolehan β akan dievaluasi berdasarkan

$$J_d=-E\left(\frac{\partial^2\ln L(\beta)}{\partial\beta\partial\beta'}|d\right)$$

$$= -\frac{\partial^2 \ln L(\beta)}{\partial\beta\partial\beta'} \quad \dots(40)$$

Dengan menggunakan *law of iterated expectation* diperoleh:

$$E(J_d) = J \quad \dots(41)$$

Maka menurut *strong Law Large Number*

$$\lim_{N\to\infty}\left(\frac{1}{N}J_d - J\frac{1}{N}\right) \overset{a.s}{\rightarrow} 0$$

Jika $\sum_i \frac{m_i m_i'}{i} < \infty$, dimana m_i mengganti elemen dari matrik $(x_{i2} - x_{i1})$ dengan kuadratnya. Dengan demikian invers dari matrik informasi ini memberikan matrik kovariansi asimtotik untuk *conditional likelihood estimator* $\beta, N \to \infty$. Sehingga diperoleh:

$$\sqrt{N}(\beta - \hat{\beta}) \xrightarrow{d} N\left(0, N\left[\frac{\partial \ln L(\beta)}{\partial\beta\partial\beta'}\right]\right)$$

Ilustrasi Konsistensi Conditional Likelihood Estimatorβ untuk T=2

Sebagaimana didefinisikan pada (32) dan (38), dengan $x_{i1} = 0$ dan $x_{i2} = 1$ untuk setiap $i = 1, \dots, N$ berlaku

$$P(y_{i1}, y_{i2}) = \left(\frac{e^{y_{i2}\beta}}{1+e^{\beta}}\right)^{d_i} \quad \dots(42)$$

Sehingga fungsi likelihood dan loglikelihoodnya dibentuk oleh :

$$L(\beta) = \prod_{i=1}^{N}\left(\frac{e^{y_{i2}\beta}}{1+e^{\beta}}\right)^{d_i}$$

$$\ln L(\beta) = \sum_{i=1}^{N} d_i\left(y_{i2}\beta - \ln(1+e^{\beta}\right) \quad \dots(43)$$

Turunan pertama terhadap β diberikan

$$\frac{\partial \ln L(\beta)}{\partial\beta} = \sum_{i=1}^{N} d_i\left(y_{i2} - \frac{e^{\beta}}{1+e^{\beta}}\right) \quad \dots(44)$$

β yang memenuhi

$$\sum_{i=1}^{N} d_i\left(y_{i2} - \frac{e^{\beta}}{1+e^{\beta}}\right) = 0 \quad \dots(45)$$

diberikan oleh

$$\hat{\beta} = \ln\left(\frac{\sum_{i=1}^{N} y_{i2} - n_2}{n_1 + n_2 - \sum_{i=1}^{N} y_{i2}}\right) \quad \dots(46)$$

dengan $n_0, n_1,$ dan n_2 didefinisikan sebagaimana sebelumnya. Sedangkan turunan kedua diperoleh

$$\frac{\partial^2 \ln L(\beta)}{\partial\beta^2} = -\sum_{i=1}^{N}\frac{e^{\beta}}{(1+e^{\beta})^2} \quad \dots(47)$$

diperoleh $\left.\frac{\partial^2 \ln L(\beta)}{\partial\beta^2}\right|_{\hat{\beta}} < 0$

yang berarti fungsi (43) mencapai maksimum.

Sebagaimana telah dinyatakan dalam (23) dan (24), sehingga menurut Law of Large Number

$$p\lim_{N\to\infty}\hat{\beta} = p\lim_{N\to\infty}\ln\left(\frac{\sum_{i=1}^{N} y_{i2} - n_2}{n_1 + n_2 - \sum_{i=1}^{N} y_{i2}}\right)$$
$$= \beta$$

Sehingga dapat disimpukan bahwa $\hat{\beta}$ adalah estimator yang konsisten untuk β.

Model Efek Tetap Dua Arah

Untuk kasus T kecil, efek waktu yang muncul pada model regresi panel logit tidak mengganggu sebagaimana efek individu α_i. Model efek dua arah ini dimodelkan dengan

$$y_{it}^{*} = x_{it}'\beta + \alpha_i + \lambda_t + \epsilon_{it} \qquad \ldots(48)$$

λ_t adalah efek waktu yang diasumsikan tetap.

Dengan ϵ_{it} berdistribusi logistik standar sehingga diperoleh :

$$P(y_{it} = 1|x_{it}, \alpha_i) = \frac{e^{x_{it}'\beta+\alpha_i+\lambda_t}}{1+e^{x_{it}'\beta+\alpha_i+\lambda_t}},$$

sedangkan

$$P(y_{it} = 0|x_{it}, \alpha_i) = \frac{1}{1+e^{x_{it}'\beta+\alpha_i+\lambda_t}}$$

Untuk $T = 2$ conditional probability diperoleh

$$P(y_{i1} = 0, y_{i2} = 1|y_{i1} + y_{i2} = 1)$$
$$= \frac{e^{x_{i2}'\beta+\lambda_2}}{e^{x_{i1}'\beta+\lambda_1}+e^{x_{i2}'\beta+\lambda_2}} \qquad \ldots(49)$$

Sehingga

$$P(y_{i1} = 0, y_{i2} = 1|y_{i1} + y_{i2} = 1)$$
$$= F[(x_{i2} - x_{i1})'\beta + \lambda_2 - \lambda_1] \qquad \ldots(50)$$

Pandang λ_t sebagai parameter yang akan diestimasi maka dapat dibentuk variabel dummy yang baru dengan nilai

$$z = \begin{cases} 1, & t = 2 \\ 0, & lainnya \end{cases} \qquad \ldots(51)$$

Dengan $\lambda_2 - \lambda_1 = \gamma$, sehingga γ menyatakan perbedaan efek waktu antara ke-2 terhadap efek waktu ke-1, maka persamaan (50) menjadi

$$P(y_{i1} = 0, y_{i2} = 1|y_{i1} + y_{i2} = 1)$$
$$= F[(x_{i2} - x_{i1})'\beta + z_2\gamma] \qquad \ldots(52)$$

Dengan demikian efek waktu merupakan variat baru yang koefisiennya dapat diselesaikan sebagaimana conditional likelihood untuk $T = 2$. Penggunaan variabel dummy digeneralisir untuk $T \geq 2$, yaitu akan terdapat $T - 1$ variabel dummy. Misalkan l menyatakan urutan variabel dummy maka

$$z_{t,l} = \begin{cases} 1, & t = l \\ 0, & lainnya \end{cases} \qquad \ldots(53)$$

dengan $l = 2,3, \ldots, T$

Simulasi Data

Analisis data ini menggunakan software R dengan library **cquad** (Bartolucci & Pigini, 2017) yang dirilis untuk software R edisi 3.0.0 keatas. Variabel respon **pov** merupakan data biner yang bernilai 1 jika perempuan masuk dalam kategori miskin dan 0 untuk yang lain. Variabel lain yang terkait adalah *age* (mur individu saat wawancara pertama), *black* (jika responden berkulit hitam, 0 untuk lainnya) *mother* (jika responden memiliki minimal satu anak, 0 yang lain), *pouse* (1 jika responden tinggal bersama pasangan, 0 yang lain), *school* (1 jika responden mengikuti program pendidikan, 0 yang lain), *hours* (jumlah jam kerja rata-rata perminggu ketika wawancara)

Ketidakkonsistenan parameter β diilustrasikan untuk satu variabel independen untuk $T = 2$, sebut saja variabel independen ini adalah efek tetap dari waktu, yaitu

$$5.year = \begin{cases} 1, & t = 2 \\ 0, & lainnya \end{cases}$$

Dummy variabel ini menyatakan ukuran perubahan waktu waktu ke-5 terhadap waktu ke-1. Untuk 900 responden diperoleh output sebagai berikut:

Tabel 1 . Perbandingan Antara Estimator MLE

dan Conditional MLE

	β	$P > z$	$LR\,Chi^2$
logit	-0.566	<0.001	<0.001
xtlogit	0.226	0.039	0.0386

Dari analisa terdapat 336 sampel yang keluar dari analisis disebabkan respon dari individu tidak berubah terhadap waktu, dengan kata lain terdapat 336 individu dengan kriteria $\sum_t y_{it} = 0,2$.

Untuk $N \to \infty$ estimator likelihood atau (Bartolucci & Pigini, 2017)logit biasa sebagaimana pada persamaan (25) memberikan

$$\beta^* = \frac{-0.566}{2}$$

$$= -0.288$$

SIMPULAN DAN SARAN

Simpulan

Adapun Kesimpulan dari penelitian ini adalah estimator MLE tidak dapat digunakan dalam mengestimasi parameter variabel independen pada data *panel* respon biner sebagaimana pada data respon biner biasa. Hal ini terjadi karena terdapat *cross section effect* α_i yang mengakibatkan estimator MLE tidak konsisten. Untuk T=2 estimator memberikan $\hat{\beta} = 2\beta$.

Simulasi ketidakkonsistenan estimator MLE pada data panel untuk N=5.755 dengan T=2 memberikan nilai $\hat{\beta} = -0.566$, yang seharusnya memberikan nilai $\beta = -0.288$.

Saran

Adapaun saran dari penelitian ini adalah penggunaan Conditional MLE untuk T yang lebih tinggi untuk dapat diperiksa secara komprehensif. Apakah memenuhi kaidah dan syarat akan suatu estimator yang baik bagi parameter variabel independen pada data panel dengan efek tetap.

REFERENSI

Agresti, A. (2009). An introduction to categorical data analysis (2nd edn). In Statistics in Medicine (Vol. 28, Issue 11).

Astuti, A. M. (2010). Fixed Effect Model pada Regresi Data Panel. Beta, 3(2), 134–145.

Bain, L. J., & Engelhardt, M. (1993). Introduction to Probability and Mathematical Statistics. Biometrics, 49(2). https://doi.org/10.2307/2532587

Baltagi, B. (2005). Econometric analysis of panel data. In Vasa.

Bartolucci, F., & Pigini, C. (2017). Cquad: An R and stata package for conditional maximum likelihood estimation of dynamic binary panel data models. Journal of Statistical Software, 78(7). https://doi.org/10.18637/jss.v078.i07

Center for Human Resource Research. (2002). NLSY97 User's Guide ;

Czepiel, S. A. (2012). Maximum Likelihood Estimation of Logistic Regression Models: Theory and Implementation. Class Notes, 1–23. papers3://publication/uuid/4E1E1B7E-9CAC-4570-8949-E96B51D9C91D

Kafadar, K., & Schott, J. R. (1997). Matrix Analysis for Statistics. Journal of the American Statistical Association, 92(440). https://doi.org/10.2307/2965451

Kusnul Kotimah, M., & Pingit Wulandari, S.

(2014). Model Regresi Logistik Biner Stratifikasi Pada Partisipasi Ekonomi Perempuan Di Provinsi Jawa Timur. Jurnal Sains Dan Seni Pomits, 3(1).

Liski, E. P. (2007). An Introduction to Categorical Data Analysis, 2nd Edition by Alan Agresti. International Statistical Review, 75(3). https://doi.org/10.1111/j.1751-5823.2007.00030_6.x

Muflihah, I. Z. (2017). Analisis Financial Distress Perusahaan Manufaktur Di Indonesia dengan Regresi Logistik. Majalah Ekonomi, XXII(2), 254–269.

Pentury, T., Aulele, S. N., & Wattimena, R. (2016). Analisis Regresi Logistik Ordinal. BAREKENG: Jurnal Ilmu Matematika Dan Terapan, 10(1), 55–60. https://doi.org/10.30598/barekengvol10iss1pp55-60

Rosadi, D. (2006). Pengantar Analisis Runtun Waktu. Diktat Kuliah, Program Studi Statistika.

Santiyasa, I. (2009). Algoritma Newton Raphson Dengan Fungsi Non-Linier. Jurnal Ilmu Komputer, 2(1), 8–16.

Smith, R. J., & Hsiao, C. (1988). Analysis of Panel Data. Economica, 55(218). https://doi.org/10.2307/2554479

Subanar. (2012). Statistika Matematika; Probabilitas, Distribusi, dan Asimtotis dalam Statistika. Graha Ilmu.