Model Probit pada Respons Biner Multivariat Menggunakan *Simulated Maximum Likelihood Estimator*

Probit Model on Multivariate Binary Response Using Simulated Maximum Likelihood Estimator

Jaka Nugraha [1)], Suryo Guritno [2)], Sri Haryatmi [2)]
[1] *Jurusan Statistika UII, S3 Matematika UGM*
[2] *Jurusan Matematika UGM*

ABSTRACT

In this paper, we discuss probit model on multivariate binary response. We assume that each of n individuals is observed in T responses. Y_{it} is t^{th} response on i^{th} individual/subject and each response is binary. Each subject has covariate X_i (individual characteristic) and covariate Z_{ijt} (characteristic of alternative j). Response on individual i^{th} can be represented by $\mathbf{Y}_i = (Y_{i1},....,Y_{iT})$, Y_{it} is t^{th} response on i^{th} individual/subject and each response is multinomial. In order to simplify, we choose one of individual characteristics and alternative characteristics. We use simulated maximum likelihood estimator (SMLE) methods to estimate the parameter based on Geweke-Hajivassiliou-Keane (GHK) simulator. We find the first derivative of likelihood function for multivariate binary probit. Then, we expand to multivariate multinomial response. The first derivative is used in the BHHH (Berndt, Hall, Hall, Hausman) iteration to obtain estimators.

Keywords: Random utility model, simulated maximum likelihood estimator, generalized estimating equation, BHHH, GHK simulator, Newton-Raphson

PENDAHULUAN

Pada data panel, pengamatan dilakukan secara berulang terhadap subjek dan variabel yang sama. Pembahasan mengenai pemodelan respon biner pada data panel telah dilakukan oleh banyak peneliti. Model yang sering digunakan adalah model probit dan model logit. Model disusun berdasarkan pendekatan model efek tetap, model efek random dan model dinamik. Metode estimasi paramater yang digunakan adalah metode *maximum likelihood estimator* (MLE), metode momen dan metode *generalized estimating equation* (GEE).

Struktur model yang paling sederhana adalah model independen, yaitu dengan mengasumsikan bahwa antar respon pada subjek yang sama maupun antar subjek adalah independen. Dengan asumsi ini, probabilitas gabungan merupakan perkalian dari probabilitas marginalnya. Liang & Zeger (1986) menyampaikan bahwa analisis logistik maupun probit pada data panel dengan menggunakan pendekatan univariat, yakni mengabaikan adanya korelasi akan menghasilkan estimator parameter yang masih konsinten tetapi jika terdapat korelasi yang besar maka penaksir tersebut menjadi tidak efisien. Prentice (1988) menyampaikan strategi pemodelan menggunakan pendekatan GEE untuk mendapatkan estimator koefisien regresi yang konsisten dan asimtotis normal. Pendekatan GEE tidak menggunakan perhitungan integral rangkap.

Contoyannis *et al.* (2002) menyatakan bahwa pada model efek tetap jika terdapat korelasi antara efek individu terhadap variabel independen maka estimator yang diperoleh menjadi tidak efisien. Dengan menggunakan model efek random, estimator yang didapatkan menjadi lebih efisien. Model lain yang dikembangkan dalam data panel adalah model dinamik. Model dinamik seperti halnya model time series, terdapat pengaruh antar respon secara berurutan (Contoyannis *et al.* 2001). Harris *et al.* (2000) telah melakukan pengujian sifat-sifat estimator model probit pada data panel dan menyimpulkan bahwa estimasi menggunakan MLE masih menghasilkan estimator yang baik meskipun jumlah sampel terbatas.

Model probit pada data panel adalah identik dengan model probit untuk keputusan tunggal, hanya struktur matrik kovariansinya lebih besar. Pengembangan model yang telah dilakukan adalah dengan mengontrol karakteristik individu yang tidak terobservasi dan homogen terhadap perulangan pengukuran. Jika karakteristik individu heterogen, maka akan menjadi masalah dalam estimasi parameter yaitu estimasi parameter menjadi bias (Greene 2003).

Metode SMLE adalah identik dengan MLE, hanya saja proporsi masing-masing pilihan dihitung secara simulasi. Metode simulasi digunakan untuk menghitung integral rangkap. Pada model probit, metode simulasi GHK merupakan metode simulasi yang paling efisien dan bersifat tak bias (Hajivassilou *et al.* 1996). Geweke *et al.* (1997) juga telah menemukan metode perhitungan integral rangkap menggunakan pendekatan simulasi Monte Carlo yang dikenal dengan nama metode GHK mempunyai sifat tak bias dan konsisten.

Nugraha (2000) telah melakukan pengujian sifat-sifat estimator parameter pada regresi logistik bivariat dengan menggunakan metode MLE dan GEE. Metode tersebut menghasilkan estimator yang konsisten. Nugraha *et al.* (2006) menunjukkan bahwa model logistik pada data biner multivariat dengan menggunakan pendekatan GEE menghasilkan penaksir parameter dengan variansi yang lebih kecil dibandingkan dengan pendekatan asumsi independen.

Sering kali pada masing-masing individu diamati beberapa variabel dependen yang berbeda secara bersamaan. Pengamatan seperti ini menghasilkan respon multivariat. Penelitian mengenai pemodelan respon biner multivariat masih sedikit mendapat perhatian dari para peneliti. Sementara itu aplikasi pemodelan respon biner multivariat sangat luas. Berdasarkan pengembangan model respon biner yang telah dilakukan pada data panel, makalah ini membahas penyusunan model pada data respon biner multivariat menggunakan model probit. Estimasi parameter dilakukan dengan menggunakan metode MLE yang didasarkan pada simulasi GHK.

Model utilitas

Diasumsikan bahwa n individu masing-masing diobservasi sebanyak T respon. Y_{it} adalah respon ke-t pada individu/subjek ke-i dan setiap responnya adalah biner. Respon pada individu ke-i, dapat disajikan dalam bentuk $\mathbf{Y}_i = (Y_{i1},....,Y_{iT})$, sehingga Y_{it} = j jika subjek i respon ke-t memilih alternatif j (j=0,1). Diasumsikan bahwa pilihan individu i dalam mengambil keputusan Y_{it} karena mempunyai utilitas maksimum dan dipengaruhi oleh kovariat X_i sebagai karakteristik individu dan kovariat Z_{ijt} sebagai karakteristik alternatif/pilihan j. Untuk menyederhanakan penulisan, diambil satu variabel karakteristik individu dan satu variabel karakteristik pilihan.

Utiliti subjek i memilih alternatif j pada respon ke-t adalah

$U_{ijt} = V_{ijt} + \varepsilon_{ijt}$.

untuk t=1,2,...,T ; i=1,2,...,n ; j=0,1 (1)

dengan

$V_{ijt} = \alpha_{jt} + \beta_{jt}X_i + \gamma_t Z_{ijt}$

U_{ijt} adalah utilitas yang merupakan variabel laten dan V_{ijt} dinamakan representatif utiliti. α_{jt} ,β_{jt} dan γ_t adalah parameter dalam model utilitas. Pada *random utility model* (RUM), diasumsikan bahwa pembuat keputusan (subjek) menentukan pilihan berdasarkan nilai utilitas yang maksimum, sehingga model (1) dapat disajikan dalam bentuk selisih utilitas,

$$U_{it} = U_{i1t} - U_{i0t} = (V_{i1t} - V_{i0t}) + (\varepsilon_{i1t} - \varepsilon_{i0t}) = V_{it} + \varepsilon_{it} \quad (2)$$

dengan

$V_{it} = (\alpha_{1t} - \alpha_{0t}) + (\beta_{1t} - \beta_{0t})X_i + \gamma_t(Z_{i1t} - Z_{i0t})$

dan $\varepsilon_{it} = \varepsilon_{i1t} - \varepsilon_{i0t}$.

Selanjutnya dapat disusun hubungan antara Y_{it} dan variabel laten U_{it}, yaitu

$y_{it} = 1 <=> U_{i1t} > U_{i0t} <=> U_{it} > 0 <=> -V_{it} < \varepsilon_{it}$

dan

$y_{it} = 0 <=> U_{i1t} < U_{i0t} <=> U_{it} < 0 <=> -V_{it} > \varepsilon_{it}$

Probabilitas subjek i memilih (y_{i1} = 1,...., y_{iT} = 1) adalah

$$\begin{aligned} P(y_{i1} = 1,..., y_{iT} = 1) &= P(0 < U_{i1},..., 0 < U_{iT}) \\ &= P(-V_{i1} < \varepsilon_{i1},..., -V_{iT} < \varepsilon_{iT}) \\ &= \int_{\varepsilon i} I(-V_{it} < \varepsilon_{it}).f(\varepsilon_i)d\varepsilon_i \forall t \end{aligned} \quad (3)$$

dengan $\boldsymbol{\varepsilon}'_i = (\varepsilon_{i1},..., \varepsilon_{iT})$. Nilai probabilitas ini merupakan hitungan integral rangkap T dan tergantung pada parameter $\theta = (\alpha,\beta,\gamma)$ maupun distribusi $\boldsymbol{\varepsilon}$. Dalam hal ini akan digunakan model probit. Metode estimasi parameter yang digunakan pada model probit adalah metode MLE.

Model probit diturunkan dari asumsi bahwa vektor $\varepsilon_i^{'}$ berdistribusi normal multivariat dengan mean nol dan matrik kovariansi $\boldsymbol{\Sigma}$. Fungsi densitas untuk ε_i adalah

$$f(\varepsilon_i) = \phi(\varepsilon_i) = \frac{1}{(2\pi)^{T/2} |\Sigma|^{1/2}} \exp[-\frac{1}{2}\varepsilon_i^{'}\Sigma^{-1}\varepsilon_i] \quad (4)$$

Probabilitas marginal (untuk suatu t dan i) adalah

$\pi_{it} = P(y_{it}=1|X_i,Z_i) = P(-V_{it} < \varepsilon_{it}) = 1- \Phi(-V_{it})$ (5)

sehingga $P(Y_{it} = y_{it}) = \pi_{it}^{y_{it}}(1-\pi_{it})^{1-y_{it}}$

dari sifat simetris distribusi normal maka persamaan (5) dapat juga dinyatakan sebagai

$$\begin{aligned}\pi_{it} &= P(y_{it}=1|X_i,Z_i)\\ &= P(-V_{it} < \varepsilon_{it}) = P(\varepsilon_{it} < V_{it}) = \Phi(V_{it})\end{aligned} \tag{6}$$

dengan

$$\Phi(V_{it}) = \int_{-\infty}^{V_{it}} \frac{1}{(2\pi\sigma_t^2)^{1/2}} \exp[-\frac{1}{2\sigma_t^2}\varepsilon_{it}^2] d\varepsilon_{it}\ .$$

Fungsi likelihood dari sampel random berukuran n adalah

$$L(\theta) = \prod_{i=1}^{n} L_i(Y_i \mid X_i, Z_i, \theta) \tag{7}$$

dengan $\mathbf{Y}'_i = (Y_{i1},...,Y_{iT})$ merupakan vektor observasi (respon) biner.

HASIL DAN PEMBAHASAN

Model probit independen

Probabilitas marginal untuk respon Y_{it} adalah

$$\begin{aligned}P(Y_{it} = y_{it}) &= \pi_{it}^{y_{it}}(1-\pi_{it})^{1-y_{it}}\\ &= (\Phi(V_{it}))^{yit}(1-\Phi(V_{it}))^{1-yit}\end{aligned}$$

dan dari persamaan (5) dipunyai

$$1-\pi_{it} = \Phi(-V_{it})$$

maka probabilitas marginalnya dapat dinyatakan sebagai

$$P(Y_{it} = y_{it}) = \Phi[(2y_{it}-1)V_{it}] \tag{8}$$

Jika diasumsikan Y_{it} saling independen untuk setiap t dan i maka

$$\begin{aligned}P(Y_{i1} = y_{i1},...,Y_{iT} = y_{iT}) &= \prod_{t=1}^{T}\pi_{it}^{y_{it}}(1-\pi_{it})^{1-y_{it}}\\ &= \prod_{t=1}^{T}\Phi\left((2y_{it}-1)V_{it}\right)\end{aligned} \tag{9}$$

Fungsi likelihoodnya adalah

$$\begin{aligned}L(\theta) &= \prod_{i=1}^{n} L_i(Y_i \mid X_i, Z_i;\theta)\\ &= \prod_{i=1}^{n}\prod_{t=1}^{T}\Phi\left((2y_{it}-1)V_{it}\right)\end{aligned} \tag{10}$$

dan fungsi log-likelihoodnya adalah

$$LL(\theta) = \sum_{i=1}^{n}\sum_{t=1}^{T}\ln\left(\Phi\left((2y_{it}-1)V_{it}\right)\right) \tag{11}$$

Derivatif pertama LL(θ) terhadap parameter $\theta_t = (\alpha_{1t}, \alpha_{0t}, \beta_{0t}, \beta_{1t}, \gamma_t)'$ adalah

$$\begin{aligned}\frac{\partial LL(\theta)}{\partial\alpha_{0t}} &= -\sum_{i=1}^{n}\frac{(2y_{it}-1)\phi\left((2y_{it}-1)V_{it}\right)}{\Phi\left((2y_{it}-1)V_{it}\right)};\\ \frac{\partial LL(\theta)}{\partial\alpha_{1t}} &= \sum_{i=1}^{n}\frac{(2y_{it}-1)\phi\left((2y_{it}-1)V_{it}\right)}{\Phi\left((2y_{it}-1)V_{it}\right)}.\\ \frac{\partial LL(\theta)}{\partial\beta_{0t}} &= \sum_{i=1}^{n}\frac{-(2y_{it}-1)X_i\phi\left((2y_{it}-1)V_{it}\right)}{\Phi\left((2y_{it}-1)V_{it}\right)};\\ \frac{\partial LL(\theta)}{\partial\beta_{1t}} &= \sum_{i=1}^{n}\frac{(2y_{it}-1)X_i\phi\left((2y_{it}-1)V_{it}\right)}{\Phi\left((2y_{it}-1)V_{it}\right)}.\\ \frac{\partial LL(\theta)}{\partial\gamma_{t}} &= \sum_{i=1}^{n}\frac{(2y_{it}-1)Z_{it}\phi\left((2y_{it}-1)V_{it}\right)}{\Phi\left((2y_{it}-1)V_{it}\right)}\end{aligned} \tag{12}$$

dengan $V_{it} = (\alpha_{1t}-\alpha_{0t}) + (\beta_{1t}-\beta_{0t})X_i + \gamma_t Z_{ijt}$, untuk t=1,...,T.

Jika $\frac{\partial LL(\theta)}{\partial\alpha_{0t}} = \frac{\partial LL(\theta)}{\partial\alpha_{1t}} = 0$ maka α_{0t} dan α_{1t} tidak teridentifikasi. Demikian juga, jika $\frac{\partial LL(\theta)}{\partial\beta_{0t}} = \frac{\partial LL(\theta)}{\partial\beta_{1t}} = 0$ maka β_{0t} dan β_{1t} tidak teridentifikasi. Oleh karena itu salah satu diberi nilai tertentu (misal α_{0t}=1 dan β_{0t} =1). Jika $\theta_t = (\alpha_{1t}, \beta_{1t}, \gamma_t)'$ dan α_{0t}=1 dan β_{0t} =1 maka dari persamaan (12) diperoleh persamaan penaksir

$$\begin{aligned}\frac{\partial LL}{\partial\theta_t} &= \begin{pmatrix}\frac{\partial LL}{\partial\alpha_{1t}}\\ \frac{\partial LL}{\partial\beta_{1t}}\\ \frac{\partial LL}{\partial\gamma_t}\end{pmatrix}\\ &= \sum_{i=1}^{n}\frac{(2y_{it}-1)\phi\left((2y_{it}-1)V_{it}\right)}{\Phi\left((2y_{it}-1)V_{it}\right)}\begin{pmatrix}1\\ X_i\\ Z_{it}\end{pmatrix} = 0\end{aligned} \tag{13}$$

dengan $V_{it} = (\alpha_{1t}-1)+(\beta_{1t}-1)X_i+\gamma_t Z_{ijt}$, untuk t=1,...,T. MLE untuk θ_t merupakan penyelesaian persamaan penaksir ini.

Derivatif ke dua fungsi log-likelihood (12) adalah

$$\frac{\partial^2 LL}{\partial\theta'_t\,\partial\theta_t} = -\sum_{i=1}^{n}\frac{(2y_{it}-1)^2V_{it}\phi\left((2y_{it}-1)V_{it}\right)}{\Phi\left((2y_{it}-1)V_{it}\right)}\begin{pmatrix}(2y_{it}-1) & X_i & Z_{it}\\ X_i & (2y_{it}-1)X_i^2 & X_iZ_{it}\\ Z_{it} & X_iZ_{it} & (2y_{it}-1)Z_{it}^2\end{pmatrix} -$$

$$\sum_{i=1}^{n}\frac{\left((2y_{it}-1)\phi\left((2y_{it}-1)V_{it}\right)\right)^2}{\Phi\left((2y_{it}-1)V_{it}\right)^2}\begin{pmatrix}1 & X_i & Z_{it}\\ X_i & 1 & X_iZ_{it}\\ Z_{it} & X_iZ_{it} & 1\end{pmatrix} \tag{14}$$

Persamaan penaksir (13) dapat diselesaikan dengan menggunakan iterasi Newton-Raphson. Jika $\theta_t = (\alpha_{1t}, \beta_{1t}, \gamma_t)'$ dan untuk t tertentu maka persamaan iterasi ke-(k+1) adalah

$$\theta_t^{(k+1)} = \theta_t^{(k)} - (H_t^{(k)})^{-1}g_t^{(k)} \tag{15}$$

dengan

$$g_t^{(k)} = \sum_{i=1}^{n} \frac{(2y_{it}-1)\phi\left((2y_{it}-1)V_{it}^{(k)}\right)}{\Phi\left((2y_{it}-1)V_{it}^{(k)}\right)} \begin{pmatrix} 1 \\ X_i \\ Z_i \end{pmatrix}$$

$$H_t^{(k)} = -\sum_{i=1}^{n} \frac{(2y_{it}-1)^2 V_{it}^{(k)}\phi\left((2y_{it}-1)V_{it}^{(k)}\right)}{\Phi\left((2y_{it}-1)V_{it}^{(k)}\right)} \begin{pmatrix} (2y_{it}-1) & X_i & Z_{it} \\ X_i & (2y_{it}-1)X_i^2 & X_iZ_{it} \\ Z_{it} & X_iZ_{it} & (2y_{it}-1)Z_{it}^2 \end{pmatrix} -$$

$$\sum_{i=1}^{n} \frac{\left((2y_{it}-1)\phi\left((2y_{it}-1)V_{it}^{(k)}\right)\right)^2}{\Phi\left((2y_{it}-1)V_{it}^{(k)}\right)^2} \begin{pmatrix} 1 & X_i & Z_{it} \\ X_i & 1 & X_iZ_{it} \\ Z_{it} & X_iZ_{it} & 1 \end{pmatrix}$$

Untuk melakukan pengujian terhadap estimator dapat digunakan sifat Normal asimtotis pada penaksir MLE**,**

$$\hat{\theta}_t \xrightarrow{a} N[\theta_t, \{I(\theta_t)\}^{-1}] \qquad (16)$$

dengan

$$I(\theta_t) = -E\left[\frac{\partial^2 \log LL_t(\theta_t; X_{it})}{\partial\theta_t\,\partial\theta_t'}\right] = -E[H_t]$$

Secara umum jika terdapat lebih dari satu variabel karakteristik individu (X_{i1},..,X_{iM}) dan variabel karakteristik pilihan (Z_{ij1t},...,Z_{ijkt}) maka

$V_{ijt} = \alpha_{jt} + \beta_{1jt}X_{i1} + \ldots + \beta_{Mjt}X_{iM} + \gamma_{1t}Z_{ij1t}, \ldots, + \gamma_{Kt}Z_{ijKt}$

untuk i=1,..n ; j=0,1 dan t=1,..,T. Agar parameternya teridentifikasi maka ditentukan nilai α_{0t}=1 dan β_{m0t} = 1 untuk semua m dan t. Parameter yang diestimasi adalah $\theta = (\alpha_{1t}, \beta_{11t}, \ldots, \beta_{M1t}, \gamma_{1t}, \ldots, \gamma_{Kt})'$. MLE merupakan penyelesaian dari persamaan

$$\sum_{i=1}^{n} \frac{(2y_{it}-1)\phi\left((2y_{it}-1)V_{it}\right)}{\Phi\left((2y_{it}-1)V_{it}\right)} \begin{pmatrix} 1 \\ X'_i \\ Z'_{it} \end{pmatrix} = 0 \qquad (17)$$

dengan $\mathbf{X}_i = [X_{i1} \ldots, X_{iM}]$, $\mathbf{Z}_{ijt} = [Z_{ij1t}, \ldots, Z_{ijKt}]$, $\mathbf{V}_{it} = V_{i1t} - V_{i0t}$, $Z_{it} = Z_{i1t} - Z_{i0t}$ untuk setiap t.

Model probit biner multivariat

Vektor $\boldsymbol{\varepsilon}_i = (\varepsilon_{i1}, \ldots, \varepsilon_{iT})'$ berdistribusi normal multivariat dengan mean nol dan matrik kovariansi $\boldsymbol{\Sigma}$ dan masing-masing ε_{it} berdistribusi normal standard.

$\varepsilon_i' \sim$ MN(0, Σ) dan ε_{it} ~N(0,1) untuk t=1,...,T.

$$\Sigma = \begin{pmatrix} 1 & \sigma_{12} & \ldots & \sigma_{1T} \\ \sigma_{21} & 1 & \ldots & \sigma_{2T} \\ \ldots & \ldots & \ldots & \ldots \\ \sigma_{T1} & \sigma_{T1} & \ldots & 1 \end{pmatrix}$$

Karena Σ merupakan matrik simetri, maka $\sigma_{tt'} = \sigma_{t't}$ untuk t,t'=1,...,T. y_{it} = 0 jika responden i memilih alternatif pertama dan y_{it} = 1, jika responden i memilih alternatif ke dua. Probabilitas marginalnya adalah

$$P(Y_{it} = y_{it}) = \Phi[(2y_{it}-1)V_{it}]$$

Probabilitas gabungannya adalah

$$P(Y_{i1} = y_{i1}, \ldots, Y_{iT} = y_{iT}) = P\left(\varepsilon_{i1} < (2y_{i1}-1)V_{i1}, \ldots, \varepsilon_{iT} < (2y_{iT}-1)V_{iT}\right)$$

$$= \int_{-\infty}^{w_{iT}} \ldots \int_{-\infty}^{w_{il}} \ldots \int_{-\infty}^{w_{i1}} \phi_T(\varepsilon_i; \theta; \Sigma)\,d\varepsilon_i = \int_{D(Y_i)} \phi_T(\varepsilon_i; \theta; \Sigma)\,d\varepsilon_i$$

$$= \Phi_T(w_i; 0; \Sigma)) \qquad (17)$$

dengan $w_{it} = (2y_{it}-1)V_{it}$ dan D($\mathbf{Y}_i$) = [-∞,w_{i1}]…[-∞,w_{il}]…[-∞,w_{iT}]. ϕ_T menyatakan fungsi densitas normal multivariat dari respon sebanyak T. Fungsi log likelihoodnya adalah

$$LL(\theta; \Sigma) = \sum_{i=1}^{n} \ln \Phi_T(w_i; 0; \Sigma) \qquad (19)$$

$\Phi_T(w_i; 0; \Sigma)$ dihitung menggunakan simulator GHK dengan faktor Cholesky C, sehingga parameter yang diestimasi adalah $\boldsymbol{\omega}$ = (θ, c). c adalah elemen-elemen matrik $\boldsymbol{C}$.

$$\boldsymbol{C} = \begin{pmatrix} c_{11} & 0 & 0 & 0 & 0 & \ldots\ldots & 0 \\ c_{21} & c_{22} & 0 & 0 & 0 & \ldots\ldots. & 0 \\ c_{31} & c_{32} & c_{33} & 0 & 0 & \ldots\ldots. & 0 \\ \ldots. & \ldots & \ldots & \ldots & \ldots & \ldots\ldots\ldots. & \\ c_{T1} & c_{T2} & c_{T3} & \ldots & \ldots & \ldots\ldots.. & c_{TT} \end{pmatrix}$$

sehingga persamaan utiliti (1) menjadi

$U_{it} = V_{it} + \sum_{l=1}^{t} c_{tl}\eta_{li}$ untuk t=1,...,T dan η_i ~ N(0,**I**)

Dengan menggunakan algoritma pada simulasi GHK (Train, 2003) diperoleh

$$\breve{p}_i^r = \prod_{t=1}^{T} \Phi_{it}^r = \prod_{t=1}^{T} \Phi\left(\frac{-((2y_{it}-1)V_{it} + \sum_{k=1}^{t-1} c_{tk}\eta_k^r)}{c_{tt}}\right) \qquad (20)$$

indek r menyatakan pengambilan ke-r dalam simulasi,

$$\breve{p}_i = \frac{1}{R}\sum_{r=1}^{R} \breve{p}_i^r$$

Fungsi log-likelihood menjadi

$$\operatorname{sim}\log L(\omega) = \sum_{i=1}^{n} \log\left(\frac{1}{R}\sum_{r=1}^{R} \breve{p}_i^r(\omega)\right)$$

$$= \sum_{i=1}^{n} \log\left(\frac{1}{R}\sum_{r=1}^{R}\prod_{t=1}^{T} \Phi_{it}^r\right) \qquad (21)$$

Estimator ω yang dihitung menggunakan metode Newton-Raphson memerlukan derivatif pertama dan derivatif ke dua dari fungsi log-likelihood (21). Untuk menghidari derivatif kedua dari fungsi log likelihood, dapat digunakan metode iterasi BHHH (Chong & Zak 1996). Metode BHHH hanya memerlukan derivatif pertama.

Selanjutnya akan dihitung derivatif pertama fungsi log likelihood dengan menggunakan notasi

$$a_{li,r} = \begin{cases} -\left[\sum_{h=1}^{l-1} \frac{c_{lh}}{c_{ll}} \eta_{hi}^{r} + \frac{(2y_{il}-1)V_{il}}{c_{ll}} \right] & l > 1 \\ \frac{(2y_{il}-1)V_{i1}}{c_{11i}} & l = 1 \end{cases} \tag{22}$$

Persamaan (20) menjadi

$$\breve{p}_i^r = \Phi\left(a_{1i,r}\right).\Phi\left(a_{2i,r}\right)....\Phi\left(a_{Ti,r}\right) = \Phi_{1i}^r.\Phi_{2i}^r....\Phi_{Ti}^r = \prod_{l=1}^{T} \Phi_{li}^r$$

Derivatif pertama fungsi likelihood (21) adalah

$$\frac{\partial}{\partial\omega} \text{sim} \log L(\omega) = \sum_{i=1}^{n} \frac{1}{\frac{1}{R}\sum_{r=1}^{R} \breve{p}_i^r(\omega)} \frac{1}{R} \sum_{r=1}^{R} \left(\left(\breve{p}_i^r(\omega) \right) \sum_{l=1}^{T} \frac{\phi_{li}^r}{\Phi_{li}^r} . \frac{a_{li,r}}{\partial\omega} \right) \tag{23}$$

dengan

$$\frac{\partial a_{li,r}}{\partial\omega} = \begin{cases} -\left[\sum_{h=1}^{l-1} \frac{c_{lh}}{c_{ll}} u_{hi,r} . \frac{\phi(a_{hi,r})}{\phi(\eta_{hi,r})} . \frac{\partial a_{hi,r}}{\partial\omega} + \frac{(2y_{il}-1)}{c_{ll}} \frac{\partial V_{il}}{\partial\omega} \right], & l > 1 \\ -\frac{(2y_{il}-1)}{c_{11}} \frac{\partial V_{il}}{\partial\omega}, & l = 1 \end{cases}$$

$$\frac{\partial a_{li,r}}{\partial c_{jk}} = \begin{cases} -\sum_{h=1}^{l-1} u_{hi,r} \frac{c_{lh}}{c_{ll}} \frac{\phi(a_{hi,r})}{\phi(\eta_{hi,r})} \frac{\partial a_{hi,r}}{\partial c_{jk}} & \text{untuk} \quad k \le j < l \\ -\frac{\eta_{ki,r}}{c_{ll}} & \text{untuk} \quad k < j = l \\ -\frac{a_{li,r}}{c_{ll}} & \text{untuk} \quad k = j = l \end{cases}$$

Dengan menggunakan langkah-langkah penyusunan model dan estimasi parameter pada model probit biner multivariat, dapat dikembangkan untuk model probit multinomial multivariat

KESIMPULAN

Model probit dapat diimplementasikan dalam pemodelan data biner multivariat yang didasarkan pada *Random Utility Model*. Metode MLE dapat digunakan untuk mengestimasi parameter. Jika diasumsikan Y_{it} saling independen untuk setiap *t* dan *i*, MLE merupakan penyelesaian dari persamaan

$$\sum_{i=1}^{n} \frac{(2y_{it}-1)\phi\left((2y_{it}-1)V_{it}\right)}{\Phi\left((2y_{it}-1)V_{it}\right)} \begin{pmatrix} 1 \\ X'_i \\ Z'_{it} \end{pmatrix} = 0$$

dengan $\mathbf{X}_i = [X_{i1} \; .... \; ,X_{iM}]$, $\mathbf{Z}_{ijt} = [Z_{ij1t},...,Z_{ijKt}]$, $V_{it} = V_{i1t} - V_{i0t}$, $Z_{it} = Z_{i1t} - Z_{i0t}$ untuk setiap t.

Jika diasumsikan Y_{it} saling berkorelasi, maka fungsi likelihoodnya akan melibatkan hitungan integral rangkap. Nilai probabilitas yang merupakan hitungan integral rangkap dalam fungsi likelihood dapat diselesaikan menggunakan simulator GHK. Untuk menghindari penggunaan derivatif kedua dari fungsi log-likelihood, persamaan penaksir yang diperoleh dari MLE dapat diselesaikan menggunakan iterasi BHHH.

DAFTAR PUSTAKA

Chong EKP & Zak SL. 1996. *An Introduction to Optimization*. John Wiley & Sons, Inc.

Contoyannis P, Andrew MJ & Rice N. 2001. Dynamics of Health in British Household: Simulation-Based Inference in Panel Probit Model. *Working Paper* Department of Economics and Related Studies, University of York.

Contoyannis P, Andrew MJ & Gonzales RL. 2002. Using Simulation-Based Inference With Panel Data In Health Economics. *Working Paper* Department of Economics and Related Studies, University of York.

Geweke JF, Keane MP & Runkle DE. 1997. Statistical Inference in The Multinomial Multiperiode Probit Model. *Journal of Econometrics* **80**: 125-165.

Greene W. 2003. *Econometrics Analysis*. 5 Editions. Prentice Hall.

Hajivassiliou V, McFadden D & Ruud P. 1996. Simulation of Multivariate Normal Rectangle Probabilities and Their derivatives: Theoretical and Computational Results. *Journal of Econometrics* **72**: 85–134.

Harris MN, Macquarie LR & Siouclis AJ. 2000. Comparison of alternative Estimators for Binary Panel Probit Models. *Melbourne Institute* Working Paper no 3/00.

Liang KY & Zeger SL. 1986. Longitudinal Data Analysis Using Generalized Linear Models. *Biometrika* **73**: 13-22.

Nugraha J. 2000. *Penaksiran parameter pada regresi logistik bivariat*. Makalah Seminar Nasional Matematika di UGM.

Nugraha J, Guritno S & Haryatmi S. 2006. *Model Regresi Logistik untuk Respons Biner Multivariat dengan Generalized Estimating Equation.* Makalah Seminar Nasional Matematika dan Pendidikan Matematika di UNY.

Prentice. 1988. Correlated Binary Regression with Covariates Specific to Each Binary Observation. *Biometrics* **44**: 1043-1048.

Train K. 2003. *Discrete Choice Methods with Simulation*, UK Press, Cambridge.