MODEL LOGIT (LOGISTIC REGRESSION) SEBAGAI MODEL PROBABILITAS LINIER ALTERNATIF, DALAM CONTOH APLIKASI TEKNIK INDUSTRI

Sugeng Santoso[1]

ABSTRACT

Regression model in which the regressand evokes a yes or no or present or absent response are known as dichotomous, or dummy, dependent variable regression models. They are applicable in a wide variety of fields and are used extensively in survey or census-type data. Among the methods that are used to estimated such models, this paper considered three-LPM,logit and probit. The LPM is the simplest of the three models to use but has several limitations. Logit and probit models guarantee that estimated probabilities lie in the 0-1 range and that they are nonlinearly to the explanatory variables. Logit model is slightly less involved because by taking the logarithm of odd ratio what appears to be a highly nonlinier model become a linier (in the parameter) model can be estimated within the standard OLS framework. Linear regression models an extremely flexible tool that is capable of handling many interesting problems encountered in empirical studies.

Key Words : Regression dummy variable, logit, flexible tool.

PENDAHULUAN

Dalam analisis regresi seringkali terjadi bahwa variabel dependen dipengaruhi variabel kualitatitaif maupun variabel kualitatif. Sedangkan variabel dependen sendiri dapat berupa variabel kualitatif. Model Regresi atas variebel dependen dummy membahas model regresi dimana variabel dependen (tak bebas) bersifat dikotomi, mengambil nilai 1 atau 0. Ada beberapa contoh dalam aplikasi Teknik Industri; misalkan kita ingin mempelajari partisipasi tenaga kerja laki-laki dewasa sebagai fungsi tingkat pengangguran, tingkat upah rata-rata pendapatan keluarga, pendidikan dan seterusnya. Seseorang bisa termasuk dalam tenaga kerja atau tidak. Jadi variabel dependen, partisipasi tenaga kerja hanya dapat mengambil dua nilai: 1 jika seseorang termasuk tenaga kerja dan 0 jika tidak. Khusus tentang model probabilitas Logit dijumpai dalam Teknik Industri sebagai model logistik atau lengkapnya fungsi distribusi kumulatif logistik. Penaksiran terhadap model probabilitas linier (LPM) menghadapi beberapa masalah khusus antara lain: (1) ketidak-normalan disturbansi; (2) varians heteroskedasitas disturbansi; (3) tidak dipenuhinya $0 \leq E(Y_i / X_i) \leq 1$. Sedangkan masalah diatas dapat diatasi dengan model logit, probit, namun

[1] *Staff Pada BPPT Indonesia*

kedua model ini adalah non-linier, sehingga perlu ditransformasikan kedalam bentuk linier dalam parameter lebih dahulu. Tulisan ini berusaha untuk bisa memberikan gambaran tentang regresi atas variabel dependen dummy, macam model probabilitas linier, penggunaan model logit dan interpretasinya.

TINJAUAN PUSTAKA

Model-model Probabilitas Linier sering mengacu pada analisis Logit, yang merupakan kombinasi Regresi Berganda dan *Multiple Discriminant Analysis* (MDA). Teknik ini serupa dengan analisis regresi berganda dalam satu atau lebih variabel independen yang digunakan untuk memperkirakan variabel dependen tunggal (Hair et al, 1992).

Perbedaan analisis ini dengan regresi berganda adalah variebel dependen bersifat nonmetrik. Karena itu terdapat perbedaan dalam metode estimasi dan asumsi tentang tipe distribusi pokok. Perbedaan dengan Analisis Diskriminan karena model probabilitas linier menampung semua jenis variabel independen (metrik dan non-metrik) dan tidak diperlukan asumsi normalitas multivariate.

Model-model Probabilitas Linier adalah adalah model yang meregresi variabel dependen dummy yang terdiri dari : (1) LPM (Linier Probability Model); (2) Model Logit; (3) Model Probit; (4); Model Tobit (Gujarati,1995).

Untuk memahami masing-masing model akan diuraikan sebagai berikut:

(1) LPM (Linier Probability Model)

Misalkan ada model sederhana: $Y_i = \beta_1 + \beta_2 X_i + u_i$ ....(1)

dimana: X_i = pendapatan keluarga

Y = 1, jika keluarga memiliki mobil

Y = 0, jika keluarga tidak memiliki mobil

Persamaan (1) yang mengekspresikan dikotomi Y_i sebagai fungsi linier variabel eksplanatori X_i, dinamakan Model Probabilitas Linier (LPM).

$E(Y_i/ Xi)$ adalah ekspektasi kondisi Y_i dengan X_i given, dapat diinterpretasikan sebagai probabilitas kondisi yang akan terjadi dengan diketahui X_i , ditulis $Pr(Y = 1/ X_i)$.

Diasumsikan $E(u_i) = 0$; (agar estimator unbiased), didapatkan:

$E(Y_i/ X_i) = \beta_1 + \beta_2 X_i$ .... (2)

P_i = probabilitas bila $Y_i = 1$ (jika peristiswa terjadi) dan $1-P_i$ = probabilitas jika $Y_i = 0$ (jika peristiwa tidak terjadi).

Y_i	**Probabilitas**
0	$1- P_i$
1	P_i
Total	1

Dengan definisi ekspektasi matematika:

$E(Y) = 0(1- P_i) + 1 (P_i) = P_i$ ....(3)

Pers (2) dan (3): $E(Y_i/X_i) = \beta_1 + \beta_2 X_i = P_i$ .... (4)

Jadi:

Perhitungan : $Y_i = \beta_1 + \beta_2 X_i$

Interpretasi : $P_i = \beta_1 + \beta_2 X_i$

Manfaat LPM adalah untuk (1) Penjelasan; (2) Prediksi; (3) Penggolongan

Untuk penjelasan dan prediksi:

$Pi = \beta_1 + \beta_2 X_i$; $0 < \beta_1 + \beta_2 X_I < 1$; jika $Pi = 1$; $\beta_1 + \beta_2 X_I \geq 1$;

jika $Pi = 0$; $\beta_1 + \beta_2 X_I \leq 0$; sesuai dengan aksioma peluang.

Untuk penggolongan:

Alokasikan pada: kelompok 1 ; $\bar{y} > ½$; (y=1, memiliki mobil)

kelompok 2 ; $\bar{y} \leq ½$; (y=0, tidak memiliki mobil)

LPM menghadapi beberapa problem misal: (1) non-normalitas dari u_i; (2) heteroskedasitas dari u_i; (3) kemungkinan $\bar{Y}$ keluar dari range 0-1; (4) secara umum nilai R^2 rendah. Karena itu dibutuhkakan model probabilitas yang (1) saat X meningkat, $P_i = E(Y_i = 1/X)$ meningkat tetapi berhenti pada range 0-1 dan (2) hubungan antara P dan X adalah nonlinier. *Cumulative Distribution Function* (CDF) dapat menjawab batasan range 0-1, yang dipergunakan pada model (1) logistik (logit) dan (2) probit.

(2) Model Logit

Model Logit adalah bersifat nonlinier sehingga *Ordinary Least Square* (OLS) tidak dapat diaplikasikan, disebut juga sebagai fun[illegible]if LOGISTIK.

Model Logit : $P_i = E(Y_i / X_i) = \dfrac{1}{1+e^{-(\beta 1 + \beta 2\, Xi)}}$ .... (5)

Bila $Z = \beta_1 + \beta_2 X_i$, maka $P_i = \dfrac{1}{1+e^{-Z}}$ .... (6)

Persamaan (6) dikenal sebagai **fungsi distribusi logistik**. Logit adalah natural log dari odd ratio.

Gunakan sifat logaritma untuk transformasi linier:

$Ln\,(Pi / (1-Pi)) = Ln\,(e^Z)$

$Ln\,(Pi / (1-Pi)) = Z = \beta_1 + \beta_2 X_i$, ... (7)

Keterangan:

$Pi / (1-Pi)$ = odd ratio; Odd ratio tidak hanya linier dalam X tetapi juga linier dalam parameter. L dinamakan logit, dengan demikian pers.(7) dinamakan model logit.

(3) Model Probit

Untuk menjelaskan sifat dikotomi variabel dependen dapat digunakan dengan Cumulative Distribution Functions (CDF). Regresi probit adalah alternatif pendekatan log-linier untuk mengatasi variabel dependen kategorikal (www2.chass.ncsu.edu/garson/pa765/logit.htm). Model estimasi dari CDF normal dikenal sebagai model probit, kadang-kadang dikenal dengan model normit.

$$F(Z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\beta_1 + \beta_2 X_i} e^{-1/2\,(z^2)}\, dz \quad \ldots\ldots\ldots\ldots (8)$$

Misal untuk pi = 0.2, Z = ?

F^{-1} (pi) = Z, berkisar antara –3,5 s/d 3,4

disebut Normit (I)

Model Probit : $Z_i = \beta_1 + \beta_2 X_I$

(4) Model Tobit

Model Tobit adalah perluasan dari model Probit yang dikembangkan oleh James Tobit. Dikenal juga sebagai model regresi *censored* atau model variabel dependen terbatas (limited) karena restriksi pengambilan nilai oleh *regressand*. Secara matematis diformulasikan sebagai berikut:

$Y_i = \beta_1 + \beta_2 X_{2i} + u$ jika RHS > 0, ... (9)

= 0, otherwise ; dimana RHS = right-hand side

Definisi Operasional Variabel Penelitian

Penghasilan (X_i) adalah jumlah penghasilan rumah tangga rata-rata tiap bulan.

Jumlah responden yang memiliki mobil dalam X_i dengan notasi N_i

Probabilitas (P_i) adalah probabilitas jika responden yang memiliki mobil, dinotasikan P_i = 1 ; Probabilitas ($1-P_i$) adalah probabilitas responden yang tidak memiliki mobil, dinotasikan $P_i = 0$; Jumlah responden yang memiliki mobil dengan notasi n_I . Frekuensi relatif ($\hat{P}_i$) = n_i / N_i

METODOLOGI PENELITIAN

Metode pengumpulan data menurut Nam Lim dalam Sigit,2001 ada empat, yaitu

(1) metode observasi;
(2) metode *documentary-historical*;
(3) metode survey; dan
(4) metode eksperimental.

Metode pengumpulan data yang digunakan dalam penelitian ini adalah metode survey (survai) karena digunakan instrumen kuesioner dan wawancara untuk mendapatkan tanggapan dari responden yang disampel (Singarimbun, 1995 dan Sigit,2001).

Ciri-ciri metode survai (Sigit,2001) adalah:

(1) ada sampel dan populasi;
(2) mencari tanggapan langsung dari responden;
(3) relatif banyak respondennya dan

(4) dalam situasi yang alami.

Penelitian survai termasuk dalam penelitian deskriptif, karena penelitian deskriptif menentukan dan melaporkan keadaan yang ada menurut kenyataannya dengan mengukurnya.

HASIL DAN PEMBAHASAN

Untuk mengestimasi model, dihubungkan dengan *income* (X_i) dengan N_i = jumlah keluarga dalam X_i ; dan n_i = jumlah responden yang memiliki mobil dalam X_i. Karena itu untuk mengestimasi P_i yang berhubungan dengan X_i digunakan:

$$\hat{P}_i = n_i / N_I;\ \hat{P}_i = \text{frekuensi relatif}$$

Untuk pendugaan parameter akan digunakan WLS (bobot (w_i)), karena masalah heteroskedasitas (Gujarati, 1995). $W_i = N_i\ p\hat{}\ (1-p\hat{})$

Dengan demikian bila ada data sebagai berikut:

Xi (dalam 100.000)	Ni	ni	pi^	Wi
6	40	8	0,2	6,4
8	50	12	0,24	9,12
10	60	18	0,30	12,6
13	80	28	0,35	18,2
15	100	45	0,45	24,75
20	70	36	0,51	17,49
25	65	39	0,60	15,6
30	50	33	0,66	11,2
35	40	30	0,75	7,5
40	25	20	0,80	4,0

Dari data diatas, responden dengan pendapatan rata-rata Rp. 3.700.000,- per bulan peluang memiliki mobil adalah

Logit = Li = ln (p^ / (1-p^))

$$\text{WLS}: Li^* = \sqrt{Wi}\ L_i;\ Xi^* = \sqrt{Wi}\ X_i;$$

$$:Li^* = -1{,}5942\sqrt{Wi} + 0{,}0786\ X^*_I$$

$$R^2 = 0{,}9650$$

$X_i = 37$

$$\beta_1^* = -1{,}5942\sqrt{7{,}5} = -4{,}366$$

$$Li^* / (X_i = 37) = -4{,}366 + 0{,}0786\ (37)$$

$$\sqrt{7{,}5} \qquad = -3{,}598$$

$$Li = Li^* / \sqrt{Wi} = -3{,}598 / \sqrt{7{,}5} = -1{,}314$$

Peluang :

$$P_i = 1/(1 + e^{-(Z)}) = 1/(1 + e^{-(-1{,}314)}) = 0{,}213$$

$$\text{OLS}: L_i = -1{,}6604 + 0{,}0792\ X_i$$

$$R^2 = 0{,}9791$$

X=37

$L_i = -1,6604 + 0,0792\ (37)$

$= 1,27$

$P_i = 1/(1+e^{-(1,27)})$

$= 0,7807$

Suatu keluarga dengan penghasilan Rp. 3.700.000,- dengan data seperti pada tabel 1, berpeluang memiliki mobil sebesar 78,07 %

KESIMPULAN

Pengenalan variabel dependen kualitatif membuat regresi linier dan nonlinier menjadi suatu alat yang sangat fleksibel yaitu mampu untuk menangani banyak masalah menarik yang dijumpai dalam studi empiris. Model-model dengan variabel dependen dummy jika dinyatakan sebagai fungsi linier dari variabel yang menjelaskan (yang mungkin bersifat kuantitatif atau kualitatif atau kedua-duanya) disebut model probabilitas linier. Model logit dan probit adalah model non-linier, sehingga perlu ditransformasikan kedalam bentuk linier dalam parameter lebih dahulu. Kedua model ini dapat mengatasi masalah yang terjadi pada LPM yaitu (1) ketidak-normalan disturbansi; (2) varians heteroskedasitas disturbansi; (3) tidak dipenuhinya $0 \leq E(Y_i / X_i) \leq 1$. Model logit yang menggunakan WLS sebagai penduga parameter, dijumpai dalam Teknik Industri sebagai model logistik atau lengkapnya fungsi distribusi kumulatif logistik. Model logit banyak digunakan dalam memberikan (1) penjelasan dan (2) prediksi, seperti dicontohkan dalam kasus prediksi hubungan pendapatan dengan kepemilikan mobil.

DAFTAR PUSTAKA

Anonim, ***Logistic Regression***, http://www2.chass.ncsu.edu/

Anonim, ***Log-Linear, Logit, and Probit Models***, http://www2.chass.ncsu.edu/garson/pa765/logit.htm

Anonim, ***P* Logit Model for Social Networks***, http://kentucky.psych.uiuc.edu/pstar/

Anonim, ***Dichotomous Association: Percent Difference, Yule's Q, Yule's Y, Risk*** http://kentucky.psych.uiuc.edu/pstar/

Anonim, Ordinal Association : Gamma, Kendall's tau-b and tau-c, Somers'd http://kentucky.psych.uiuc.edu/pstar/

Gujarati, Damodar (1995), ***Basic Econometrics***, McGraw-Hill, Third Edition, New York.

Hair et al, ***Multivariate Data Analysis with Readings***, Macmillan Publishing Co., New York.

Hosmer. Jr and Lemeshow (1989) ***Applied Logistic Regression***, John Wiley & Sons, New York.