

BAB 4

METODOLOGI PENELITIAN

4.1 Sampel, Sumber Data dan Cara Pengumpulan Data

Bab ini membahas model dan metodologi yang digunakan oleh penulis dalam penelitian, beserta penyesuaian yang dilakukan terhadap model yang digunakan. Untuk menjelaskan hasil penelitian penulis akan menggunakan analisis deskriptif dan analisis kuantitatif dalam menjelaskan probabilitas kemampuan bertahan perusahaan.

Studi ini menggunakan data sekunder yang merupakan data panel dari perusahaan-perusahaan yang ada di industri pengolahan kakao. Penulis menggunakan data Statistik Industri Menengah Besar dari BPS, dengan kode ISIC 31191 yaitu industri bubuk kakao. Pada data tersebut terdapat kurang lebih 12 perusahaan¹⁹, data BPS terdapat perbedaan dengan data perusahaan yang terdapat pada AIKI dan Departemen Perindustrian²⁰. Namun untuk konsistensi penelitian penulis memutuskan untuk menggunakan data BPS sebagai sumber penelitian. Rencana awal penelitian melingkupi analisa periode tahun 2000, 2001, 2002, 2003, 2004, 2005, dan 2006 akan tetapi terdapat kesalahan pencatatan kode perusahaan pada data BPS tahun 2000 dan 2001. Melihat keadaan tersebut penulis memutuskan untuk melakukan penelitian hanya pada periode tahun 2002 sampai dengan 2006. Selain data BPS penelitian ini menggunakan sumber data dari AIKI, Departemen Pertanian dan situs resmi ICCO.

Dalam analisis *survival model*, penghitungan hanya pada perusahaan yang terdapat di periode pertama dan bertahan pada periode berikutnya (diberi nilai 1), atau tidak ada pada periode berikutnya (diberi nilai 0)²¹. Perusahaan yang tidak terdapat pada periode pertama namun muncul pada periode berikutnya (perusahaan baru), tidak dapat dimasukkan kedalam analisis.

¹⁹ Data pada tahun 2002.

²⁰ Berdasarkan Laporan Kinerja Industri Pengolahan Kakao 2008 AIKI terdapat 16 perusahaan, sedangkan berdasarkan Road Map Kakao 2005-2025 Departemen Pertanian terdapat 17 perusahaan. Menurut Departemen Perindustrian yang terdapat pada Gambaran Sekilas Industri Kakao(2007) terdapat 56 perusahaan pengolahan kakao.

²¹ Penentuan variabel *d_survival* merupakan keberadaan (*survival*) perusahaan tiap tahunnya, yang kemudian dimasukan kedalam analisa data panel periode yang dimaksud.

Penelitian ini menggunakan alat analisa ekonometrika berupa *software* komputer, yaitu STATA 8.0. Penelitian dilakukan dengan menggunakan data panel pada rentang waktu 2002-2006.

4.2 Spesifikasi Model : Modifikasi *Survival Model*

Untuk meneliti kemampuan perusahaan bertahan dalam industri pengolahan kakao, penulis melakukan modifikasi model *firm-survival* Paul Dunes dan Alan Hughes (1994) sehingga diperoleh model akhir sebagai berikut:

$$d_survival(d=1) = f(Age, Worker, PPN, Prc_Dis, Productivity) + Error_t \quad (4.1)$$

Dimana : $d = 1$; *Survive* → Model Probit
 $d = 0$; *non-survive*

d_survival = Probabilitas perusahaan bertahan hidup (variabel dependen). Merupakan variabel yang bersifat *discrete*. Nilainya adalah 1 (satu) dan 0 (nol). Nilai satu untuk perusahaan yang bertahan dalam industri pengolahan kakao

Age = Umur Perusahaan (variabel independen). Nilainya diperoleh dari selisih tahun berdirinya perusahaan dengan tahun pada masing-masing periode pertumbuhan. Penghitungan umur perusahaan sebagai berikut;

$$A_t = Y_t - Y_0 \quad (4.2)$$

Dimana A_t : umur perusahaan di tahun t

Y_t : tahun t

Y_0 : tahun dimana perusahaan didirikan

Satuan yang digunakan adalah tahun

Worker = Jumlah tenaga kerja (variabel independen). Nilai ini didapat dari jumlah seluruh pekerja disetiap perusahaan. Satuannya jumlah jiwa.

PPN²² = PPN (variabel independen), merupakan besar PPN yang dibayar oleh perusahaan per tahun. Satuannya adalah rupiah.

Karena tidak terdapat data besaran PPN yang dibayarkan untuk setiap perusahaan maka penulis menggunakan *proxy* sebagai berikut (satuannya adalah Rupiah):

Bahan baku Dalam Negeri = Bahan baku Total–Bahan baku Impor

$$\begin{aligned} \text{PPN} &= \text{Local Raw Material} - \frac{\text{Local Raw Material } t}{(1+r)} \\ &= \text{Local Raw Material} \left[1 - \frac{1}{(1+r)} \right] \end{aligned} \quad (4.3)$$

Prc_Dspty²³ = *Price Disparity* (variabel independen), merupakan selisih harga biji kakao international dengan harga biji kakao Indonesia. Satuannya adalah Rupiah, dengan penghitungan sebagai berikut;

$$\text{Pri_Dis}_t = \text{Int_Price}_t - \text{Ind_Price}_t \quad (4.4)$$

Dimana : Pri_Dis_t = selisih harga biji kakao

Int_Price_t = harga biji kakao international

Ind_Price_t = harga biji kakao Indonesia

Productivity²⁴ = Produktifitas variabel independen, merupakan ratio antara output dan input, yaitu output yang dihasilkan untuk setiap satu satuan input. Pengukuran ini merupakan produktifitas secara total. Produktifitas juga dapat diukur secara parsial antara lain dengan melihat produktifitas tenaga kerja, yaitu nilai output yang dihasilkan untuk satu tenaga kerja. Satuannya adalah nilai ratio.

$$\text{Productivity} = \frac{\text{Output}}{\text{Input}} \quad (4.5)$$

Error_t = *Error* pada model

²² PPN yang diberlakukan terhadap kakao sebesar 10% dan berdasarkan keterangan dari pihak AIKI input bahan baku perusahaan pengolahan Kakao sebagian besar (95%) merupakan biji kakao walaupun ada input lain yaitu berupa bahan kimia pewarna untuk bubuk kakao dalam jumlah kecil.

²³ Harga kakao international dikonversi kedalam Rupiah dengan nilai kurs tengah rata-rata tahunan.

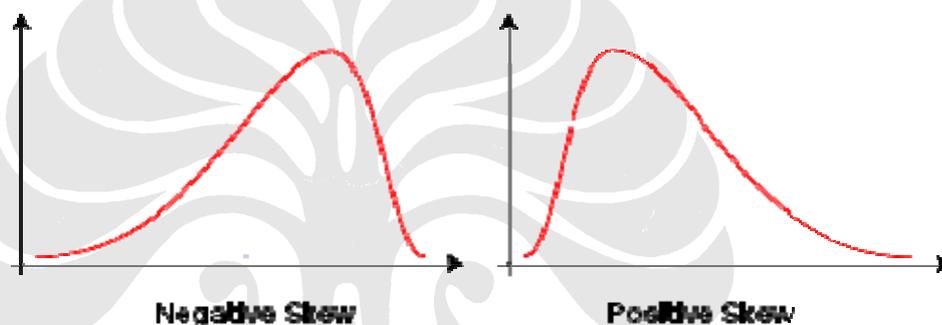
²⁴ Penelitian Bernard(2003) menyatakan bahwa probabilita perusahaan keluar dari industri meningkat pada perusahaan dengan produktifitas rendah dan tidak terhubung dengan pasar ekspor.

4.3 Normalitas Data

Asumsi mendasar pada analisa statistik ialah melihat karakter dari kumpulan data yang akan diteliti, hal ini termasuk melihat *skewness* dan *kurtosis*.

Skewness adalah pengukuran sebaran data yang simetris, pengujian ini melihat penyimpangan dari data yang simetris. Distribusi data dikatakan simetris jika terlihat sama sebarannya pada nilai terkecil dan terbesarnya dengan titik tengah. Jika bentuk distribusi sebelah kiri (*tail*) lebih besar daripada sebelah kanan maka disebut *negatif skewness* dan sebaliknya disebut *positive skewness*. Data yang terdistribusi normal berbentuk simetris dengan titik tengah, data yang terdistribusi normal disebut *zero skewness*.

Gambar 4.1 Distribusi Data (*Skewness*)



Sumber: www.wikipedia.com

Untuk data Y_1, Y_2, \dots, Y_N persamaan skewness yaitu:

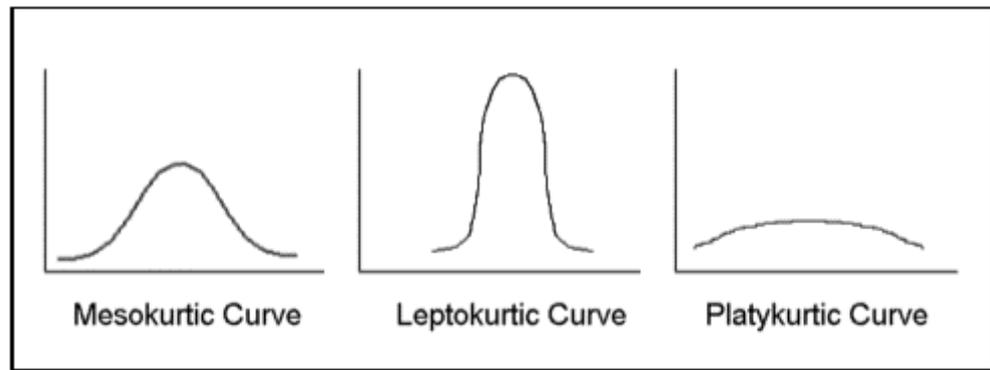
$$Skewness = \frac{\sum_{i=1}^N (Y_i - \bar{Y})^3}{N S^3} \quad (4.6)$$

adalah rata-rata, S adalah standart deviasi dan N adalah jumlah data. Skewness untuk distribusi normal adalah nol dan distribusi data yang simetris harus memiliki nilai *skewness* dekat dengan nol. *Negatif Skew* menandakan condong ke kiri dan *positive skew* menandakan condong ke kanan. *Skewed left* artinya *left tail* relatif lebih panjang daripada *right tail*, sebaliknya *Skewed right* artinya *right tail* relatif lebih panjang daripada *left tail*.

Kurtosis adalah pengukuran data dengan melihat apabila data tersebut *peaked* atau *flat* terhadap distribusi normal. Kumpulan data disebut *high kurtosis* jika cenderung memiliki puncak yang tinggi dekat dengan reratanya kemudian

menurun secara tajam dan memiliki *heavy tails*. Kumpulan data disebut *low kurtosis* jika memiliki *flat top* pada reratanya dibandingkan puncak yang tajam. Pada kasus ekstrem ialah distribusi data yang sama/seragam.

Gambar 4.2 Jenis-jenis Distribusi Kurtosis



Sumber: www.allpsych.com

Bentuk distribusi normal digambarkan dengan Mesokurtic. Kurva yang menggambarkan distribusi lebih banyak ditengah dan memiliki puncak yang lebih tinggi dinamakan Leptokurtic. Kemudian kurva yang menggambarkan distribusi yang sedikit ditengah dan lebih banyak pada sisinya dinamakan Platykurtic.

Untuk data Y_1, Y_2, \dots, Y_N formula untuk kurtosis adalah

$$Kurtosis = \frac{\sum_{i=1}^N (Y_i - \bar{Y})^4}{(N-1)S^4} \quad (4.7)$$

Dimana \bar{Y} adalah rata-rata, S adalah standart deviasi dan N adalah jumlah data. Nilai *kurtosis* untuk distribusi normal standart adalah tiga, oleh karena itu beberapa sumber menggunakan penghitungan berikut ini

$$Kurtosis = \frac{\sum_{i=1}^N (Y_i - \bar{Y})^4}{(N-1)S^4} - 3 \quad (4.8)$$

Jika menggunakan penghitungan ini maka distribusi normal ditunjukkan dengan nilai kurtosis nol, dan *positive kurtosis* menunjukkan distribusi “*peaked*” dan *negatif kurtosis* menandakan distribusi “*flat*”. Penggunaan kedua definisi ini tergantung dari konvensi, yang harus diperhatikan adalah mengetahui konvensi pada software yang kita gunakan merujuk kepada salah satu definisi. Penelitian ini menggunakan software STATA 8.0, yang merujuk kepada rumus pertama (4.7).

4.4 Pengolahan Data *Survival Model* : Model Probit

Sebagian besar model yang digunakan dalam penelitian berupa data kualitatif sebagai variabel dependen dan variabel independennya. Akan tetapi permasalahan dalam penelitian tidak selalu berupa variabel kualitatif (berupa angka) yang membutuhkan model spesifik untuk menjelaskan permasalahan tersebut. Pada bagian ini akan dijelaskan mengenai model yang menggunakan data kualitatif pada variabel dependennya.

Model dengan sifat kualitatif pada variabel dependennya atau disebut juga *binary/dichotomous* adalah model probit. Data yang digunakan berbeda dengan biasanya yang *continuous*, model probit menggunakan data *binary* sehingga menjadi *discrete*. Model probit ini salah satu yang menggunakan data bersifat *discrete* untuk variabel dependennya, variabel *discrete* ialah variabel yang nilainya tidak dapat berada diluar opsi jawaban. Contoh variabel ini adalah tingkat pendidikan, misalnya angka 1 diberikan untuk yang berpendidikan SD, 2 untuk SMP, 3 untuk SMA, 4 untuk S1. Diantara nilai-nilai tersebut tidak terdapat nilai lainnya. Contoh lain adalah variabel jenis kelamin, status eselon dalam kepegawaian dan sebagainya. Dilain pihak contoh untuk variabel yang bersifat *continuous* adalah pendapatan, tingkat konsumsi, tabungan dan sebagainya.

Dalam model probit ini menggunakan notasi 1 (satu) dan 0 (nol). Angka satu untuk kejadian yang berhasil dan angka nol untuk kejadian tidak berhasil. Sebagai contoh dapat digunakan untuk $Y=1$ untuk perusahaan yang dapat bertahan dalam industri dan $Y=0$ untuk perusahaan yang tidak bertahan/mati.

Sebelum menjelaskan lebih lanjut tentang model probit ada baiknya melihat teori yang mendasarinya terlebih dahulu.

Linier Probability Model (LPM)

$$Y_t = \beta_1 + \beta_2 X_t + u_t \quad (4.9)$$

Dimana Y_i adalah variabel kualitatif dan X_i adalah variabel independennya.

Model LPM yang digunakan adalah:

$$E(Y_t | X_t) = \beta_1 + \beta_2 X_t \quad (4.10)$$

Dimana, $E(Y_t | X_t)$ adalah ekspektasi terjadi Y_i pada saat nilai X_i . Dengan menggunakan asumsi bahwa P_i adalah ekspektasi $Y_i=1$ (kejadian berhasil) dan $(1-$

P_i) adalah ekspektasi $Y_i=0$ (kejadian tidak berhasil) maka dengan menggunakan teori kemungkinan sederhana dimana $P_i+(1-P_i)=1$, dapat ditulis sebagai berikut:

$$E(Y_t) = 0(1-P_t) + 1(P_t) = P_t \quad (4.11)$$

Maka jika mengabungkan (4.10) dan (4.11) akan diperoleh:

$$E(Y_t | X_t) = \beta_1 + \beta_2 X_t = P_t \quad (4.12)$$

Karena P_i adalah nilai dari probabilitas maka akan berada pada lingkup $0 \leq E(Y_t | X_t) \leq 1$. Sehingga dapat ditarik kesimpulan dari LPM, adalah tidak terlalu jauh berbeda dengan OLS biasa. Variabel dependen *binary* dapat diregress dengan metode OLS. Namun karena hubungan antara Y_i dengan X_i adalah linier, maka nilai Y_i sangat tergantung dari nilai X_i akan berada diluar *range* 0 sampai dengan 1. Hal ini menjadikan LPM tidak terlalu baik. Kelemahan utama dari LPM ialah nilainya Y_i yang sangat mungkin berada diluar *range* variabel dependen yang *binary*. Untuk itu dibutuhkan model yang ketika mendekati satu atau nol pergerakannya melambat sehingga tidak bisa keluar dari interval tersebut dan hubungan antara P_i dan X_i tidak linier.

Untuk mengatasi hal tersebut maka model yang selanjutnya disebut dengan *Cumulative Distribution Function* (CDF). Model CDF yang sifatnya random ini dibagi menjadi dua, yaitu model logit dan probit.

Pada model logit hubungan antara P_i dengan X_i tidak lagi linier. Persamaan yang digunakan ialah:

$$P_t = (E=1 | X_t) = \frac{1}{1 + e^{-(\beta_1 + \beta_2 X_t)}} \quad (4.13)$$

Dapat ditulis juga sebagai berikut:

$$P_t = \frac{1}{1 + e^{-z}} = \frac{e^{z_t}}{1 + e^{z_t}} \quad (4.14)$$

Dimana : $Z_t = \beta_1 + \beta_2 X_t$

Maka dari persamaan (4.14) dapat dibuat persamaan $(1-P_i)$:

$$1-P_t = 1 - \frac{e^{z_t}}{1 + e^{z_t}} = \frac{1}{1 + e^{z_t}} \quad (4.15)$$

Sehingga apabila dibandingkan persamaan (4.14) dan (4.15) akan diperoleh persamaan (4.16) sebagai berikut:

$$\frac{P_t}{1-P_t} = e^{z_t} \quad (4.16)$$

Sehingga persamaan (4.16) adalah *odds ratio*, yaitu probabilitas terjadinya kejadian sukses (satu). Kemudian untuk memperoleh nilai natural log (model

logit), akan diperoleh nilai Z_i itu sendiri seperti yang ditunjukkan pada persamaan (4.17).

$$L_t = \ln\left(\frac{P_t}{1-P_t}\right) = Z_t = \beta_1 + \beta_2 X_t \quad (4.17)$$

Dimana: L adalah log natural dari *odds ratio*. Sehingga implikasi dari model tersebut ialah besar nilai P_i dengan X_i tidak berhubungan linier seperti yang terjadi pada LPM.

Hal-hal yang perlu dicermati ialah:

1. Nilai P akan bergerak dalam rentang 0 sampai 1. Walaupun nilai logit L akan bergerak dalam rentang $-\infty$ hingga ∞ . Hal ini mengatasi permasalahan pada model LPM.
2. Walaupun nilai L_i linier dengan nilai X_i , namun tidak demikian probabilitasnya.
3. Model logit memungkinkan menambahkan variabel independen lebih dari satu.
4. Jika nilai L_i adalah positif artinya semakin besar nilai variabel independen (X_i), maka nilai P mendekati 1 (satu) artinya probabilitas kejadian sukses semakin besar. Jika nilai L_i negatif artinya semakin besar nilai variabel independen (X_i), maka nilai P semakin besar mendekati 0 (nol) artinya probabilitas kejadian tidak sukses semakin besar.
5. Intrepetasi lain untuk model logit pada persamaan (4.17): β_2 adalah slope yang besarnya merupakan perubahan dari L_i (*log odds*) setiap perubahan pada X_i . Nilai ini secara teori sulit diartikan.

Model probit dan logit adalah dua jenis pendekatan CDF untuk mengatasi masalah pada LPM. Perbedaan antara logit dan probit adalah pada perbedaan asumsi persebaran data. Pada logit diasumsikan data mengikuti persebaran distribusi e , sedangkan pada probit mengikuti distribusi normal oleh karena itu model probit disebut juga model normit (*normal distribution*).

Dalam model probit menggunakan pendekatan dari nilai utilitas atau pendekatan perilaku dalam memilih rasional (*rational choice on behavior*). Sebagai contoh misalnya peneliti ingin melihat faktor pendapatan (variabel independen) terhadap kemungkinan seseorang memiliki rumah (variabel dependen)

yang *binary*). Maka untuk melihat kemungkinan seseorang memiliki rumah model menghitung indeks utilitas (I_i) yang dipengaruhi variabel independennya. Semakin besar nilai I_i , maka semakin besar kemungkinan (P_i) seseorang memiliki rumah.

$$I_i = \beta_1 + \beta_2 X_i \quad (4.18)$$

Dimana X_i adalah pendapatan.

Untuk menentukan besarnya tingkat indeks utilitas yang menentukan apakah seseorang akan memiliki rumah atau tidak digunakan indeks kritis I_i^* . Jika I_i melebihi I_i^* seseorang memiliki rumah. Sama seperti I_i , I_i^* juga tidak dapat dihitung nilai pastinya, sehingga digunakan asumsi bahwa I_i^* dan I_i terdistribusi normal dan memiliki nilai *mean* dan *varians* yang sama. Dengan menggunakan asumsi tersebut, probabilitas dari I_i^* lebih kecil atau sama dengan I_i dapat dihitung dengan model CDF terdistribusi normal:

$$P_i = P(Y=1 | X) = P(I_i^* \leq I_i) = P(Z_i \leq \beta_1 + \beta_2 X_i) = F(\beta_1 + \beta_2 X_i) \quad (4.19)$$

Dimana $P(Y=1 | X)$ adalah kemungkinan terjadinya kejadian sukses ($Y=1$) dengan nilai variabel X tertentu. Dan Z_i adalah standart normal variabel, $Z \sim N(0, \sigma^2)$, F adalah *standart normal* CDF.

Model Probit dinotasikan sebagai berikut:

$$F(I_i) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{I_i} e^{-z^2/2} dz$$

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

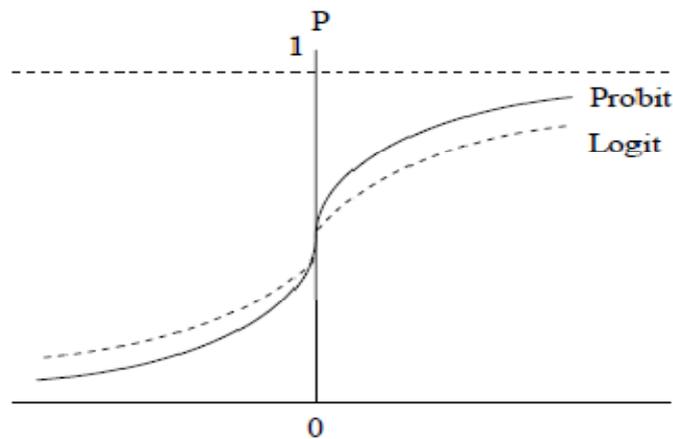
Untuk memperoleh nilai I_i dilakukan invers terhadap fungsi persamaan (4.19) :

$$I_i = F^{-1}(I_i) = F^{-1}(P_i) = \beta_1 + \beta_2 X_i \quad (4.20)$$

dimana F^{-1} adalah invers dari normal CDF.

Pertanyaan yang muncul kemudian adalah model mana yang lebih baik antara logit dengan probit. Sebenarnya kedua model ini adalah sama, perbedaannya hanya pada model probit memiliki ekor yang sedikit lebih panjang.

Gambar 4.3 Perbedaan Distribusi Logit dan Probit



Sumber : *Basic Econometrics* Damodar N.Gujarati

Untuk nilai koefisien keduanya, jika mengalihkan nilai koefisien probit dengan $\pi/\sqrt{3}$ (sebesar 1,81) maka akan diperoleh nilai koefisien logit. Sedangkan jika mengalihkan nilai koefisien logit dengan $0,55 (\sqrt{3}/\pi)$ maka akan diperoleh nilai koefisien probit.

4.5 Uji Pelanggaran Ekonometrika

Model probit jauh berbeda dengan model OLS. Dalam model OLS terdapat istilah BLUE (*Best Linier Unbiased Estimator*), namun pada model probit tidak terdapat istilah ini, karena bukan model linier. Dalam OLS hasil regresi menghasilkan sebuah garis linier yang merupakan agregat kuadrat dari simpangan masing-masing observasi yang terkecil (*least square*), sedangkan model probit adalah model yang menggunakan pendekatan *Maximum Likelihood* (ML). Oleh karena itu hasil regresi model probit akan menghasilkan kecenderungan yang paling mungkin.

Dalam regresi probit yang harus diperhatikan adalah *goodness of fit* dan normalitas data observasi yang digunakan. Uji normalitas secara statistik sederhana akan terpenuhi jika jumlah observasi yang digunakan lebih dari 20. Kemudian untuk melihat *goodness of fit* dapat melihat kepada :

- Nilai *pseudo R²* : Sesuai dengan namanya kata *pseudo* yang artinya seakan-akan, jadi nilai ini merupakan nilai yang seakan-akan nilai *R²* pada model OLS. Namun nilai *pseudo R²* disini tidak seakurat *R²* atau *adjusted*

R^2 pada model OLS. Namun nilai *pseudo R²* yang semakin tinggi pada model probit dapat diartikan model tersebut lebih baik.

- *Sensitivity* : menyatakan seberapa besar hasil observasi positif secara tepat dinyatakan positif
- *Specitivity* : menyatakan seberapa besar hasil observasi negatif secara tepat dinyatakan negatif
- Meninjau grafik antara *sensitivity/specitivity* dan *probability cut-off*, jika koordinat ($X < 0,5$ dan $Y > 0,5$) maka dapat dikatakan bahwa model semakin baik dan stabil.

