



UNIVERSITAS
ISLAM
INDONESIA

JAKA NUGRAHA

**METODE
MAKSIMUM
LIKELIHOOD
DALAM
MODEL
PEMILIHAN
DISKRIT**

METODE MAKSIMUM LIKELIHOOD DALAM MODEL PEMILIHAN DISKRIT

Penulis:
Jaka Nugraha

Universitas Islam Indonesia
©2017

Perpustakaan Nasional RI:
Katalog Dalam Terbitan (KDT)

Nugraha, Jaka

Metode Maksimum Likelihood Dalam Model
Pemilihan Diskrit / Jaka Nugraha; --Yogyakarta:
Universitas Islam Indonesia, 2017.

--.141 hlm. ; 24 cm

ISBN 978-602-450-019-1

e-ISBN 978-602-450-020-7

© 2017 Jaka Nugraha

Hak cipta dilindungi Undang-undang.

Dilarang memperbanyak atau memindahkan seluruh atau sebagian isi buku ini dalam bentuk apapun, baik secara elektronik ataupun mekanik termasuk memfotokopi, tanpa izin tertulis dari Penulis.

Untuk kalangan sendiri sebagai materi proses pembelajaran dan tidak diperjualbelikan.

METODE MAKSIMUM LIKELIHOOD DALAM MODEL PEMILIHAN DISKRIT

Penulis :
Jaka Nugraha

Terbitan:
Februari 2017 M / Rabi'atul Akhir 1438 H

Cetakan Ke-1

Penerbit:



**UNIVERSITAS
ISLAM
INDONESIA**

Kampus Terpadu UII, Jl. Kaliurang Km 14,5,
Yogyakarta 55584, INDONESIA
Tel. (0274) 896 447 Ext. 1301; Fax. (0274) 896 445
<http://www.uui.ac.id>; e-mail: perpustakaan@uui.ac.id

Kata Pengantar

Buku ini menjelaskan beberapa model yang dapat digunakan dalam menyusun pola pemilihan responden (konsumen) dalam menentukan pilihan barang yang dibeli. Model yang dibahas di dalam buku ini adalah model Logit, model *Generalized Extreme Value*, model Probit dan model *Mixed Logit*. Pembahasan ditekankan bagaimana mengestimasi parameter yang terdapat dalam model tersebut dan membahas kelebihan-kelemahan masing-masing model melalui data simulasi. Estimasi parameter dilakukan dengan menggunakan metode *Maximum Likelihood*.

Pada bab satu membahas model Logit yang mencakup struktur dalam model Utilitas dan keterkaitannya dengan model Logit, estimasi parameter, sifat model Logit dan keterbatasannya. Bab kedua membahas model Logit dan model Probit untuk respon biner. Bab ketiga dan keempat membahas pengembangan model Logit, yaitu model Probit pada bab 3 dan model *Mixed Logit* pada bab 4. Bab kelima, keenam dan ketujuh membahas sifat dari metode *Maximum Likelihood* dan statistik uji yang dapat digunakan untuk inferensi parameter maupun menguji kecocokan model.

Penulis berharap, tulisan ini dapat membantu para peneliti di bidang pemodelan perilaku konsumen yang telah banyak berkembang dan diaplikasikan di berbagai bidang termasuk transportasi, energi, perumahan dan pemasaran. Buku ini dapat digunakan sebagai referensi bagi peneliti yang tertarik pada pengembangan model diskrit yang permasalahannya masih sangat terbuka lebar meliputi inovasi model, metode Estimasi dan metode Simulasi dalam menghitung nilai probabilitas.

Yogyakarta, 20 Oktober 2016

Penulis

Daftar Isi

| | |
|---|-----|
| Kata Pengantar | iii |
| Daftar Isi | iv |
| Daftar Tabel | vi |
| Daftar Gambar | vii |
| 1. Model Pemilihan Diskrit Dan Model Logit | |
| 1.1 Pendahuluan | 1 |
| 1.2 Model utilitas | 3 |
| 1.3 Model <i>logit</i> | 5 |
| 1.4 Estimasi parameter model logit | 9 |
| 1.5 Model multinomial <i>logit</i> | 12 |
| 1.6 Contoh kasus model <i>multinomial logit</i> | 15 |
| 1.7 Sifat penaksir MLE dalam model <i>logit</i> | 17 |
| 1.7.1 Studi data simulasi | 18 |
| 1.7.2 Bias dan nilai pseudo R^2 pada korelasi tinggi | 24 |
| 1.7.3 Kesalahan tipe I | 27 |
| 2. Model Probit Dan Model Logit Pada Respons Biner | |
| 2.1 Spesifikasi model <i>probit</i> dan model <i>logit</i> | 36 |
| 2.2 MLE dan uji hipotesis pada model <i>logit</i> dan <i>probit</i> | 39 |
| 2.3 Interpretasi estimator | 40 |
| 2.4 Contoh kasus | 42 |
| 3. Model Probit | |
| 3.1 Pendahuluan | 47 |
| 3.2 Model multinomial <i>probit</i> (MNP) | 47 |
| 3.3 Identifikasi parameter | 52 |
| 3.4 Variasi individu | 57 |
| 3.5 Estimasi parameter | 59 |
| 3.5.1 Simulasi <i>monte carlo</i> | 59 |
| 3.5.2 Metode simulasi GHK | 61 |
| 3.6 Simulasi GHK dalam MSLE | 65 |
| 4. Model Mixed Logit | |
| 4.1 Pendahuluan | 71 |
| 4.2 Model <i>mixed logit</i> | 72 |
| 4.3 Metode <i>maximum likelihood estimators</i> (MLE) | 73 |
| 4.4 Barisan halton | 74 |
| 4.5 Studi simulasi | 74 |

| | | |
|-----------|--|-----|
| 4.5.1 | Menghitung korelasi utilitas | 75 |
| 4.5.2 | Program estimasi parameter | 77 |
| 4.5.3 | Hasil simulasi | 78 |
| 5. | Sifat-Sifat Penaksir Maximum Likelihood dan Simulated MLE Pada Respon Multinomial | |
| 5.1 | Penaksir <i>maximum likelihood</i> | 85 |
| 5.2 | MLE pada distribusi multinomial | 94 |
| 5.2.1 | Model <i>logit</i> multinomial | 97 |
| 5.2.2 | Model <i>nested logit</i> | 97 |
| 5.2.3 | Model <i>probit</i> | 98 |
| 5.3 | <i>Maximum simulated likelihood estimator</i> (MSLE) | 99 |
| 5.4 | Prosedur uji MLE | 102 |
| 5.4.1 | Uji rasio <i>likelihood</i> | 102 |
| 5.4.2 | Uji <i>wald</i> | 104 |
| 5.4.3 | <i>Lagrange multiplier test</i> | 105 |
| 5.5 | Simulasi <i>monte carlo</i> | 105 |
| 5.6 | Metode optimisasi | 106 |
| 5.6.1 | Metode <i>newton raphson</i> (NR) dan <i>fisher scoring</i> | 106 |
| 5.6.2 | Metode BHHH | 108 |
| 5.6.1 | Metode BFGS dan DFP | 109 |
| 6. | Menghitung Nilai Probabilitas Pada Distribusi Normal Multivariat | |
| 6.1 | Pendahuluan | 111 |
| 6.2 | Distribusi normal univariat | 111 |
| 6.3 | Distribusi normal bivariat | 112 |
| 6.4 | Probabilitas normal trivariat | 116 |
| 6.5 | Distribusi multivariat normal | 118 |
| 6.6 | Menghitung nilai probabilitas dalam program R | 121 |
| 7. | Uji Goodness Of Fit Dalam DCM | |
| 7.1. | Pendahuluan | 123 |
| 7.2. | Uji <i>goodness of fit</i> | 125 |
| 7.3. | Studi kasus | 127 |
| | Referensi | 131 |
| | Glosarium | 135 |
| | Indeks Subjek | 137 |

Daftar Tabel

| | | |
|-----------|---|-----|
| Tabel 1.1 | Data Survei Demografi dan Kesehatan di El Salvador | 15 |
| Tabel 1.2 | Hasil estimasi parameter model multinomial | 16 |
| Tabel 1.3 | Estimator pada model 1 | 24 |
| Tabel 1.4 | Nilai estimasi parameter untuk replikasi 20 dengan $n=100$ | 24 |
| Tabel 2.1 | Estimasi parameter dari model linear, model <i>logit</i> dan model <i>probit</i> | 42 |
| Tabel 2.2 | Estimasi parameter setelah distandardisasi | 43 |
| Tabel 2.3 | Estimasi efek parsial variabel x_5 | 44 |
| Tabel 4.1 | Konversi variansi dan korelasi | 76 |
| Tabel 4.2 | Hasil estimasi korelasi | 78 |
| Tabel 4.3 | Hasil estimasi parameter regresi | 82 |
| Tabel 7.1 | Data hasil pendampingan | 127 |
| Tabel 7.2 | Nilai probabilitas model <i>probit</i> | 127 |
| Tabel 7.3 | Nilai probabilitas model <i>logit</i> | 128 |

Daftar Gambar

| | | |
|-------------|---|-----|
| Gambar 1.1 | Grafik fungsi densitas distribusi normal standar dan distribusi extreme value | 5 |
| Gambar 1.2 | Grafik fungsi densitas kumulatif distribusi normal standar dan distribusi extreme value | 6 |
| Gambar 1.3 | Plot usia terhadap <i>logit</i> | 16 |
| Gambar 1.4 | Plot usia terhadap <i>logit</i> pada nilai aktual dan nilai prediksi | 17 |
| Gambar 1.5 | Grafik nilai R^2 pada $n=500$ replikasi 20 | 25 |
| Gambar 1.6 | Grafik nilai $V(x)$ terhadap $p(x)$ | 26 |
| Gambar 1.7 | Grafik kesalahan tipe I (α) pada $n=500$ replikasi 20 untuk parameter B01 | 28 |
| Gambar 1.8 | Grafik kesalahan tipe I (α) pada $n=500$ replikasi 20 untuk parameter B02 | 29 |
| Gambar 1.9 | Grafik kesalahan tipe I (α) pada $n=500$ replikasi 20 untuk parameter B1 | 30 |
| Gambar 1.10 | Grafik kesalahan tipe I (α) pada $n=500$ replikasi 20 untuk parameter B2 | 31 |
| Gambar 1.11 | Grafik kesalahan tipe I (α) pada $n=500$ replikasi 20 untuk parameter γ | 32 |
| Gambar 2.1 | Grafik distribusi kumulatif normal standar dan distribusi logistik | 36 |
| Gambar 4.1 | Bias estimator parameter α_1 | 79 |
| Gambar 4.2 | Bias estimator parameter α_2 | 80 |
| Gambar 4.3 | Bias estimator parameter β_1 | 80 |
| Gambar 4.4 | Bias estimator parameter β_2 | 81 |
| Gambar 4.5 | Bias estimator parameter γ | 81 |
| Gambar 7.1 | Prediksi menggunakan model <i>probit</i> maupun <i>logit</i> | 129 |

Bab 1

Model Pemilihan Diskrit dan Model Logit

1.1 Pendahuluan

Setiap manusia, lembaga, perusahaan akan dihadapkan pada pilihan-pilihan. Kita harus memilih satu dari semua alternatif yang tersedia, mulai dari satu alternatif saja (tidak ada alternatif lain), dua alternatif sampai dengan banyak alternatif. Banyak hal/faktor yang digunakan untuk mengambil keputusan. Pertama, faktor internal yaitu sifat-sifat atau karakteristik yang melekat pada diri pembuat keputusan. Kedua, faktor eksternal yaitu keadaan/faktor yang berasal dari luar/lingkungan pembuat keputusan. Kedua faktor tersebut dapat bersifat kualitatif maupun kuantitatif.

Sementara itu keputusan/pilihan dilakukan berdasarkan atas asas manfaat dan mudarat (risiko) bagi pembuat keputusan. Suatu pilihan dilakukan jika pilihan itu paling menguntungkan dibandingkan dengan semua alternatif pilihan yang lain, atau bisa juga karena pilihan itu paling kecil risikonya dibanding dengan alternatif yang lain. Di samping itu, kadang kita tertarik untuk menduga nilai respon yang dikaitkan dengan sekumpulan *covariate* dengan menggunakan fungsi penghubung tertentu. Ketika respon kontinu, umumnya dipilih fungsi penghubung identitas, sedangkan untuk respons biner dipilih fungsi penghubung *Logit*. Terdapat perbedaan antara analisis regresi logistik ganda dengan analisis regresi ganda. Analisis regresi ganda digunakan pada distribusi normal, sedangkan analisis regresi logistik digunakan untuk data yang berdistribusi binomial. Regresi Logistik pada respons dikotomis tersebut biasa dinamakan Regresi Logistik Biner. Sering kali respons mempunyai lebih dari dua alternatif jawaban misalkan tentang warna kesukaan (merah, kuning, hijau, dsb.) atau contoh yang lain tentang penilaian (sangat baik, baik, cukup, buruk, sangat buruk). Pada data polikotomis tersebut dapat digunakan analisis Regresi Multinomial.

Model pemilihan diskrit (*Discrete Choice Model: DCM*) adalah pemodelan dengan mengasumsikan bahwa pembuat keputusan menentukan pilihan di antara sekumpulan pilihan berdasarkan nilai utilitasnya. Pilihan atau respons berupa data nominal dan dipilih salah satu yang mempunyai nilai utilitas tertinggi (maksimum) oleh pembuat keputusan. DCM dapat menggambarkan pembuat keputusan dalam menentukan pilihan. Pembuat keputusan atau responden di sini dapat berupa orang, keluarga, perusahaan atau unit pembuat keputusan yang lain. DCM mulai

berkembang dan banyak diaplikasikan di bidang ekonomi dan transportasi setelah ditemukannya model *Multinomial Logit* (MNL) oleh Mc Fadden pada tahun 1974. DCM menyangkut dua aktivitas yang saling berhubungan, yaitu penentuan model dan perhitungan proporsi masing-masing pilihan. Model yang telah banyak dibahas antara lain model *Logit*, model *Generalized Extreme Value* (GEV), dan model *Probit*.

Tujuan analisis DCM adalah menghitung probabilitas responden dalam memilih sebuah pilihan. Persamaan probabilitas pada model *Logit* berupa persamaan tertutup yang perhitungannya relatif sederhana dibandingkan persamaan probabilitas dalam model *Probit* dan model *Mixed Logit* yang berbentuk persamaan terbuka dengan melibatkan integral rangkap. Dalam menghitung proporsi masing-masing pilihan, khususnya pada model *Probit* dan model *Mixed Logit*, simulasi memainkan peranan penting.

Beberapa metode simulasi yang dapat digunakan untuk menghitung proporsi dalam model *Probit* di antaranya adalah metode *Accept-reject*, metode *Smoothed Accept-reject*, dan metode *Geweke-Hajivassiliou-Keane* (GHK). Dari beberapa metode simulasi untuk model *Probit*, Hajivassilio, dkk. (1996) menemukan bahwa GHK adalah yang paling akurat. Geweke, dkk. (1994) menemukan bahwa GHK lebih baik dibanding *Smoothed AR*. Harris, dkk. (2000) telah melakukan pengkajian perbandingan metode estimasi model *Probit* biner pada data panel dan diperoleh hasil bahwa estimasi menggunakan metode Maksimum *Likelihood* adalah yang terbaik dibanding dengan pendekatan *Solomon-Cox* untuk fungsi *Likelihood* dan *Gibbs sampler* dalam estimasi *Bayesian*.

Sebelum menghitung proporsi untuk masing-masing pilihan, perlu dilakukan estimasi parameter dalam model. Metode estimasi parameter pada DCM yang berdasarkan simulasi terhadap fungsi probabilitasnya antara lain *Maximum Simulated Likelihood* (MSL), Metode *Simulated Score* (MSS) dan Metode *Simulated Moment* (MSM) (Contoyannis, dkk., 2001). Metode-metode estimasi tersebut analog dengan metode estimasi klasik yaitu metode *Maksimum Likelihood* dan metode *Moment*, hanya saja proporsi masing-masing pilihan dihitung secara simulasi. Metode *Maksimum Likelihood* menghasilkan penaksir yang mempunyai sifat-sifat asimtotis yang baik yaitu konsisten, normal asimtotis dan efisien (Maddala, dkk., 1993).

Dalam penelitian bidang transportasi, riset pasar maupun politik, seringkali pengamatan melibatkan respons multivariat. Misalnya dalam pemilihan umum, setiap orang memilih 3 anggota wakil rakyat masing-masing untuk DPRD I, DPRD II dan DPR. Contoh lain dalam riset pasar, bagaimana pola konsumen dalam membeli

kebutuhan rumah tangga seperti sabun mandi, sabun cuci, pasta gigi. Penelitian mengenai pola konsumen dalam membeli jenis *handphone* beserta jenis kartunya. Analisis logistik maupun *Probit* pada data panel dengan menggunakan pendekatan univariat, yakni mengabaikan adanya korelasi akan menghasilkan penaksir yang masih konsisten tetapi jika terdapat korelasi yang besar maka penaksir tersebut menjadi tidak efisien. Pada data biner panel, metode *Generalized Estimating Equation* (GEE) dapat digunakan untuk mendapatkan penaksir yang konsisten dan efisien (Liang dan Zeger, 1986). Demikian juga pada respons multivariat, jika dianalisis menggunakan pendekatan univariat maka akan mempengaruhi sifat-sifat penaksirnya. Umumnya pada kasus multivariat ini, diatasi menggunakan pendekatan univariat yaitu menyusun pilihan berupa kombinasi di antara semua alternatifnya. Misalkan untuk kasus dua respons, respons pertama terdiri dari 3 alternatif dan respons kedua terdiri dari 4 alternatif. Jika digunakan pendekatan univariat maka akan terdapat kombinasi pilihan sebanyak 12 alternatif. Pendekatan univariat ini mengakibatkan sifat *Independence from Irrelevant Alternatives* (IIA) tidak terpenuhi. Di samping itu, peneliti tidak dapat menyusun model marginalnya.

1.2 Model utilitas

Pemodelan pilihan diskrit atau DCM menggambarkan pembuat keputusan (responden) dalam menentukan sebuah pilihan di antara sekumpulan pilihan yang tersedia. Responden dapat berupa orang, rumah tangga, perusahaan atau unit pembuat keputusan yang lain. Himpunan semua pilihan atau alternatif disebut *choice set*. DCM digunakan untuk menguji pilihan “yang mana”, sedangkan model regresi dipakai untuk menguji “berapa banyak”. DCM diturunkan di bawah asumsi manfaat maksimum bagi responden (pembuat keputusan). *Choice set* memenuhi tiga sifat. Pertama, semua pilihan harus *mutually exclusive*. Setiap responden hanya memilih tepat satu pilihan. Kedua, *exhaustive*, yaitu semua pilihan tersedia untuk dipilih. Ketiga, jumlah pilihan adalah *finite* (Train, 2003).

Seorang responden dinotasikan dengan i , yang berhadapan dengan pilihan sebanyak J pilihan. Responden mempunyai tingkat utilitas (keuntungan) untuk setiap pilihan. Misalkan U_{ij} untuk $j=1, \dots, J$ adalah utilitas responden i jika memilih pilihan j . Peneliti tidak mengetahui nilai utilitas U_{ij} bagi responden pada masing-masing pilihan. Peneliti hanya mengamati keputusan responden (Y_{ij}) dan atribut yang ada untuk masing-masing pilihannya yang dinotasikan dengan Z_{ij} dan atribut responden yang dinotasikan dengan X_i . Diasumsikan, kedua atribut tersebut mempunyai persamaan

$$V_{ij} = f(Z_{ij}, X_i)$$

$$= \alpha_j + \beta_j X_i + \gamma_j Z_{ij}$$

atau dapat dinyatakan sebagai

$$V_{ij} = \theta_j W_{ij} \quad (1.1)$$

dengan $\theta_j = (\alpha_j, \beta_j, \gamma)$ dan $W_{ij} = (1, X_i, Z_{ij})'$ dan $i=1, \dots, n$ dan $j=1, \dots, J$. V_{ij} dinamakan *representative utility*. Model (1.1) merupakan model yang disampaikan oleh Boulduc (1999).

Nilai utilitas U_{ij} tidak diketahui oleh peneliti, oleh karena itu dapat disusun model

$$U_{ij} = V_{ij} + \varepsilon_{ij} \quad (1.2)$$

$\varepsilon_i = (\varepsilon_{i1}, \dots, \varepsilon_{ij})'$ adalah variabel *random* yang mempunyai densitas $f(\varepsilon_i)$. V_{ij} merupakan faktor terobservasi dan ε_{ij} merupakan faktor tidak terobservasi dalam utilitas. Train (2003) menjelaskan bahwa variabel *random* U_{ij} diasumsikan memiliki sifat kontinu, $U_{ij} \in \mathbb{R}$ (\mathbb{R} : real) *completeness*, yaitu untuk sebarang U_{ij} dan U_{ik} dengan $j \neq k$, akan memiliki salah satu sifat $U_{ij} > U_{ik}$ atau $U_{ij} < U_{ik}$. *transitivity*, disebut juga "*rationality*". Jika $U_{is} < U_{ij}$ dan $U_{ij} < U_{ik}$ maka $U_{is} < U_{ik}$.

Beberapa sifat utilitas yang berkaitan dengan spesifikasi dan estimasi parameter dalam DCM adalah sifat "*Only differences in utility matter*". Penambahan dengan konstanta tertentu terhadap semua U_{ij} , tidak akan merubah utilitas tertingginya (peringkat utilitas). Sifat yang lain adalah "*The scale of utility is arbitrary*." Dengan mengalikan setiap U_{ij} dengan bilangan positif λ tidak akan merubah peringkat utilitasnya.

Responden memilih pilihan yang mempunyai utilitas terbesar, sehingga memilih pilihan k jika dan hanya jika $U_{ik} > U_{ij} \forall j \neq k$. Probabilitas responden i memilih pilihan k dapat dinyatakan sebagai

$$\begin{aligned} \pi_{ik} &= P(U_{ik} > U_{ij}) \quad \forall j \neq k \\ &= P(\varepsilon_{ij} - \varepsilon_{ik} < V_{ik} - V_{ij}) \quad \forall j \neq k \\ &= \int_{\varepsilon} I(\varepsilon_j - \varepsilon_k < V_k - V_j) f(\varepsilon_i) d\varepsilon_{i1} \dots d\varepsilon_{ij} \quad \forall j \neq k \end{aligned} \quad (1.3)$$

$I(\cdot)$ adalah fungsi indikator, yang bernilai 1 jika pernyataan dalam kurung benar dan

bernilai 0 jika pernyataan salah. Selanjutnya dapat dipilih atau ditentukan densitas $f(\cdot)$ yang sesuai/tepat.

1.3 Model *logit*

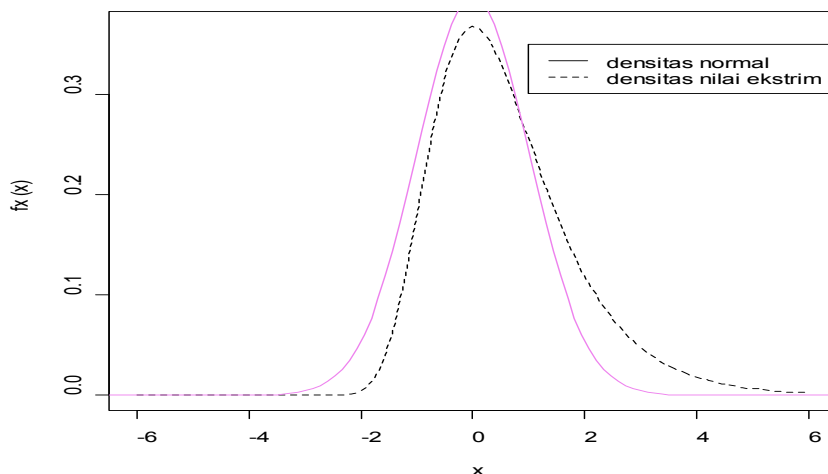
Model *Logit* diturunkan dengan asumsi bahwa ε_{ij} berdistribusi *extreme value Type I* untuk semua j . Fungsi densitas *extreme value type I* adalah

$$f(\varepsilon_j) = \exp(-\varepsilon_j) \exp(-\exp(-\varepsilon_j))$$

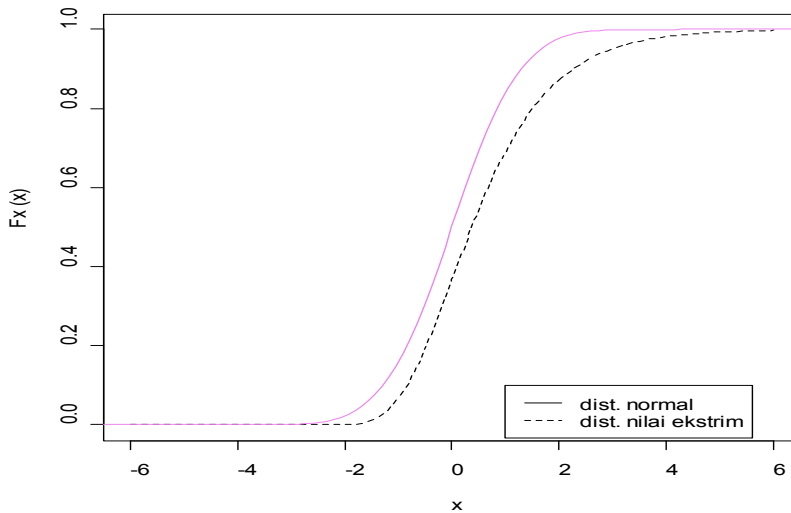
dan distribusi kumulatifnya adalah

$$F(\varepsilon_j) = \exp(-\exp(-\varepsilon_j)) \quad (1.4)$$

Mean dan variansi dari distribusi ini masing-masing adalah 0,5772 dan $\pi^2/6$ (Weisstein, 1999). Fungsi densitas *Extreme Value* ini hampir simetrik dengan distribusi normal standar, hanya saja distribusi ini mempunyai ekor yang lebih tebal dibanding dengan distribusi normal. Grafik fungsi densitas dan fungsi distribusi *Extreme Value* jika dibandingkan dengan distribusi normal standar ditampilkan di Gambar 1.1 dan Gambar 1.2.



Gambar 1.1. Grafik fungsi densitas distribusi normal standar dan distribusi *extreme value*



Gambar 1.2. Grafik fungsi densitas kumulatif distribusi normal standar dan distribusi *extreme value*

Teorema 1.1 (Train, 2003)

Jika $\epsilon_i = (\epsilon_{i1}, \dots, \epsilon_{ij})'$ adalah variabel random dan saling independen yang mempunyai densitas *Extreme Value* tipe I maka probabilitas responden i memilih pilihan k adalah

$$\pi_{ik} = \frac{\exp(V_k)}{\sum_j \exp(V_j)} \text{ untuk } k=1,2,\dots,J$$

dengan $V_{ij} = \alpha_j + \beta_j X_i + \gamma Z_{ij}$.

Bukti:

Probabilitas pembuat keputusan i memilih alternatif k yang dinyatakan sebagai berikut:

$$\begin{aligned} P_{ik} &= \Pr(\epsilon_{ij} - \epsilon_{ik} < V_{ik} - V_{ij} \quad \forall j \neq k) \\ &= \Pr(\epsilon_{ij} < \epsilon_{ik} + V_{ik} - V_{ij} \quad \forall j \neq k) \end{aligned}$$

Jika ε_{ik} diketahui dan saling independen, maka

$$P_{ik} | \varepsilon_{ik} = \prod_{j \neq k} \exp(-\exp(-(\varepsilon_{ik} + V_k - V_j)))$$

Oleh karena nilai ε_{ik} tidak diketahui maka P_{ik} merupakan integral $P_{ik} | \varepsilon_{ik}$ atas seluruh nilai ε_{ik} terbobot densitasnya, yaitu

$$P_{ik} = \int \prod_{j \neq k} (\exp(-\exp(-(\varepsilon_{ik} + V_k - V_j)))) [\exp(-\varepsilon_{ik})] \exp(-\exp(-\varepsilon_{ik})) d\varepsilon_{ik}$$

Dengan mengambil $s = \varepsilon_{ik}$ dan oleh karena $V_{ik} - V_{ik} = 0$ maka P_{ik} dapat dinyatakan sebagai

$$\begin{aligned} P_{ik} &= \int_{s=-\infty}^{\infty} \prod_j (\exp(-\exp(-(s + V_{ik} - V_{ij})))) [\exp(-s)] ds \\ &= \int_{s=-\infty}^{\infty} \exp\left(-\sum_j \exp(-(s + V_{ik} - V_{ij}))\right) \exp(-s) ds \\ &= \int_{s=-\infty}^{\infty} \exp\left(-e^{-s} \sum_j \exp(-(V_{ik} - V_{ij}))\right) e^{-s} ds \end{aligned}$$

Misalkan $t = \exp(-s)$ sehingga $dt = -\exp(-s) ds$

$$\begin{aligned} P_{ik} &= \int_{\infty}^0 \exp\left(-t \sum_j \exp(-(V_{ik} - V_{ij}))\right) (-dt) \\ &= \int_0^{\infty} \exp\left(-t \sum_j \exp(-(V_{ik} - V_{ij}))\right) dt \\ &= \frac{\exp\left(-t \sum_j \exp(-(V_{ik} - V_{ij}))\right)}{-t \sum_j \exp(-(V_{ik} - V_{ij}))} \Bigg|_0^{\infty} \end{aligned}$$

$$\begin{aligned}
&= \frac{1}{\sum_j \exp(-(V_{ik} - V_{ij}))} \\
&= \frac{\exp(V_{ik})}{\sum_j \exp(V_{ij})} \qquad \qquad \qquad \text{(terbukti)}
\end{aligned}$$

Probabilitas responden i memilih pilihan k yang dinyatakan sebagai:

$$\pi_{ik} = \frac{\exp(V_{ik})}{\sum_{j=1}^J \exp(V_{ij})} \qquad \qquad \qquad (1.5)$$

dinamakan model *Logit*. Untuk sebarang dua alternatif j dan k, rasio probabilitas *Logitnya* dapat dinyatakan sebagai

$$\frac{\pi_j}{\pi_k} = \exp(V_j - V_k) \qquad \qquad \qquad (1.6)$$

Rasio ini tidak tergantung pada alternatif lain selain j dan k. Sifat ini dinamakan *Independence from Irrelevant Alternatives (IIA)*. Formula untuk P_{ik} tersebut dinamakan probabilitas *Logit*. Jika V_{ij} merupakan fungsi linear dari x_{ij} maka dapat dinyatakan menjadi

$$P_{ik} = \frac{\exp(\beta^t x_{ij})}{\sum_{j=1}^J \exp(\beta^t x_{ij})} \qquad \qquad \qquad (1.7)$$

Untuk sebarang dua alternatif k dan r, rasio probabilitas *Logitnya* dapat dinyatakan sebagai

$$\frac{P_k}{P_i} = \frac{e^{V_{ik}}}{\sum_{j=1}^J e^{V_{ij}}} \bigg/ \frac{e^{V_{ir}}}{\sum_{j=1}^J e^{V_{ij}}} = \exp(V_{ik} - V_{ir}) \quad (1.8)$$

Rasio ini tidak tergantung pada alternatif lain selain i dan k. Sifat ini dinamakan *Independence from Irrelevant Alternatives* (IIA).

1.4 Estimasi parameter model logit

Dalam persamaan (1.8), P_{ik} merupakan sebagaimana dalam persamaan (1.8) fungsi linear dari x_{ij} dan memuat parameter β . Selanjutnya estimasi parameter β dapat dilakukan dengan berbagai metode, salah satunya adalah prosedur Maksimum *Likelihood*. Misalkan N sampel dari individu yang membuat keputusan, probabilitas individu n memilih sebuah alternatif dapat dinyatakan sebagai

$$\prod_j (P_j)^{y_j}$$

Dengan $y_{ij} = 1$ jika individu i memilih j dan nol jika memilih yang lainnya. Dengan mengasumsikan bahwa setiap keputusan antar individu saling independen maka probabilitas masing-masing individu dalam sampel memilih sebuah alternatif adalah

$$L(\beta) = \prod_{i=1}^n \prod_j (P_j)^{y_j} \quad (1.9)$$

Dengan β merupakan vektor parameter dalam model. Fungsi *Log Likelihood* nya menjadi

$$L(\beta) = \sum_{i=1}^n \sum_j y_j \ln(P_j) \quad (1.10)$$

Penaksir β adalah nilai β yang memaksimumkan fungsi $LL(\beta)$.

Teorema 1.2

Penaksir β dengan menggunakan prosedur Maksimum *Likelihood* adalah penyelesaian dari persamaan

$$\sum_{i=1}^n \sum_{j=1}^J (y_{ij} - P_j) x_j = 0$$

Bukti:

$$\begin{aligned} L(\beta) &= \sum_{i=1}^n \sum_j y_{ij} \ln(P_j) \\ &= \sum_{i=1}^n \sum_j y_{ij} \ln \left(\frac{\exp(\beta^t x_{ij})}{\sum_j \exp(\beta^t x_{ij})} \right) \\ &= \sum_{i=1}^n \sum_i y_{ij} (\beta^t x_{ij}) - \sum_{i=1}^n \sum_j y_{ij} \ln \left(\sum_j \exp(\beta^t x_{ij}) \right) \end{aligned}$$

Derivatif $LL(\beta)$ terhadap β adalah

$$\begin{aligned} \frac{\partial L(\beta)}{\partial \beta} &= \sum_{i=1}^n \sum_j y_{ij} x_j - \sum_{i=1}^n \sum_j y_{ij} \sum_j P_j x_j \\ &= \sum_{i=1}^n \sum_j y_{ij} x_j - \sum_{i=1}^n \left(\sum_j P_j x_j \right) \sum_j y_{ij} \\ &= \sum_{i=1}^n \sum_j y_{ij} x_j - \sum_{i=1}^n \sum_j P_j x_j \\ &= \sum_{i=1}^n \sum_j (y_{ij} - P_j) x_j = 0 \end{aligned} \quad \text{(terbukti)}$$

Penyelesaian persamaan penaksir tersebut dapat diselesaikan secara iterasi dengan metode *Newton-Raphson* maupun metode *Scoring*. Selanjutnya untuk melakukan menguji kesesuaian model dengan data (uji *goodness of Fit*) dapat digunakan *Likelihood Ratio Index*, yang didefinisikan sebagai

$$\rho = 1 - \frac{L(\hat{\beta})}{L(0)} \quad (1.11)$$

dengan $L(0)$ adalah nilai *Log Likelihood* untuk $\beta=0$.

Uji hipotesis dan interval konfidensi untuk parameter (*slope*) adalah (Koppelman, dkk., 2006).

a. Uji untuk masing-masing *slope*

$H_0 : \beta_j = \beta_j^{(0)}$ didasarkan pada statistik Wald:

$$Z_0 = \frac{B_j - \beta_j^0}{\hat{SE}(B_j)} \quad (1.12)$$

b. Uji untuk beberapa *slope*

$H_0 : \beta_j = \dots = \beta_q = 0$ didasarkan pada statistik

$$X^2 = G_{\text{model 1}}^2 - G_{\text{model 2}}^2 \quad (1.13)$$

yang berdistribusi Chi kuadrat dengan derajat bebasnya sama dengan selisih banyaknya parameter dari kedua model. G^2 adalah deviance yang mempunyai nilai $-2\log L$

Untuk menguji kecocokan model dapat digunakan statistik Pseudo R^2 yang identik dengan nilai R^2 (koefisien deterministik).

$$\text{Pseudo } R^2 = 1 - \frac{G_1^2}{G_0^2} \quad (1.14)$$

Jika model secara sempurna memprediksi nilai Y ($P_i = 1$ maka $y_i = 1$ dan jika $P_i = 0$ maka $y_i = 0$) maka $\log L = 0$ (atau nilai *devianceny* nol). Sehingga nilai maksimum dari Pseudo R^2 adalah satu. Statistik Pseudo R^2 secara luas digunakan untuk menjelaskan kecocokan model dalam DCM secara intuitif. Pemasalahan dalam penggunaan pseudo R^2 ini adalah tidak adanya kaidah untuk menyatakan pada nilai berapa sedemikian hingga model dikatakan baik. Permasalahan kedua adalah peningkatan nilai pseudo R^2 pada penambahan variabel independen tidak dapat menjelaskan seberapa penting variabel tersebut (Koppelman, dkk., 2006).

1.5 Model multinomial logit

Model Regresi Logistik digunakan untuk menggambarkan hubungan antara variabel independen/ prediktor dengan variabel dependen/ respon yang berupa data biner/ dikotomi (Hosmer dan Stanley, 1989) Variabel responnya (Y), biasanya digunakan nilai 0 (gagal) dan 1 (sukses). Fungsi distribusi yang digunakan adalah distribusi logistik dengan notasi $p = \pi(X) = E(Y|X)$ untuk menyatakan mean bersyarat dari Y jika diberikan vektor *covariate* $X_i = (x_{i1}, x_{i2}, \dots, x_{ip})^t$. Model regresi logistik dapat dinyatakan dalam bentuk

$$\pi(X_i) = \frac{\exp(X_i^t \beta)}{1 + \exp(X_i^t \beta)} \quad (1.15)$$

dengan $\beta = (\beta_1, \dots, \beta_p)^T$ adalah vektor parameter dan $0 \leq \pi(X) \leq 1$ dan $i=1,2,\dots,n$.

Untuk menentukan model regresi, harga β ditaksir lebih dahulu dengan menggunakan Metode Kemungkinan Maksimum dan dilanjutkan dengan pengujian terhadap model (uji kecocokan model dan uji koefisien regresi) (Agresti, 1990).

Pengembangan atas Regresi Logistik Biner adalah Regresi Logistik Multinomial, yaitu variabel dependen Y bernilai diskrit (1, 2, ...J). Jenis data untuk variabel Y dapat berbentuk data nominal maupun ordinal.

Variabel random Y_i bernilai diskrit dengan indeks 1, 2, ...J.

$$\pi_{ij} = P\{Y_i = j\} \quad (1.16)$$

merupakan probabilitas responden ke-i memilih kategori ke-j untuk $i=1,2,\dots,n$ dan $j=1,2,\dots,J$. Diasumsikan antar pilihan adalah saling asing sehingga untuk setiap i.

Sehingga $\sum_{j=1}^J \pi_{ij} = 1$ hanya mempunyai J-1 parameter.

Dapat disusun distribusi Multinomial

$$P\{Y_{i1}=y_{i1}, \dots, Y_{ij}=y_{ij}\} = \binom{n_i}{y_{i1}, \dots, y_{ij}} \pi_{i1}^{y_{i1}} \dots \pi_{ij}^{y_{ij}} \quad (1.17)$$

Untuk menyusun model *Logit*, dilakukan dengan cara salah satu kategori (biasanya kategori terakhir) dijadikan *baseline*, sehingga model logitnya adalah

$$\eta_{ij} = \log\left(\frac{\pi_{ij}}{\pi_{iJ}}\right) = X_{ij}^t \beta = \beta^t X_{ij} \quad (1.18)$$

untuk $j=1,2,\dots,(J-1)$. (Rodriguez, 2001)

Teorema 1.3

Model probabilitas responden/individu ke i memilih alternatif j adalah

$$\pi_j = \frac{\exp(\eta_j)}{\sum_{k=1}^J \exp(\eta_k)} \text{ untuk suatu } i \text{ dan suatu } j \text{ di mana } i=1,2,\dots,n \text{ dan } j=1,2,\dots,J$$

Bukti:

$$\eta_{ij} = \log\left(\frac{\pi_{ij}}{\pi_{iJ}}\right) = X_{ij}^t \beta$$

$$\Leftrightarrow \frac{\pi_{ij}}{\pi_{iJ}} = \exp(\eta_{ij})$$

$$\Leftrightarrow \pi_{ij} = \pi_{iJ} \exp(\eta_{ij})$$

Karena $\sum_{j=1}^J \pi_j = 1$ dan $\eta_{iJ} = 0$ maka

$$\sum_{j=1}^J \pi_j = \sum_{j=1}^J \pi_{iJ} \exp(\eta_j) = 1$$

$$\Leftrightarrow \sum_{j=1}^J \pi_j = \pi_{iJ} \sum_{j=1}^J \exp(\eta_j) = 1$$

$$\Leftrightarrow \pi_{iJ} = \frac{1}{\sum_{k=1}^J \exp(\eta_{ik})} \text{ dan } \pi_{ij} = \frac{\exp(\eta_{ij})}{\sum_{k=1}^J \exp(\eta_{ik})}$$

(terbukti)

Untuk menentukan model regresi, harga β ditaksir lebih dahulu dengan menggunakan Metode Kemungkinan Maksimum.

Teorema 1.4

Misalkan suatu sampel terdiri n observasi dari pasangan observasi (X_i, y_i) , $i=1, \dots, n$ dan $Y_i = (y_{i1}, y_{i2}, \dots, y_{ij})$ dengan $y_{ij} = 0, 1$ untuk $j=1, 2, \dots, J$. Model Regresi Logistik

$$\pi_j = \frac{\exp(\eta_j)}{\sum_{k=1}^J \exp(\eta_k)}$$

penaksir $\beta = (\beta_1, \dots, \beta_p)^t$ menggunakan metode Kemungkinan Maksimum adalah penyelesaian dari persamaan kemungkinan:

$$\sum_{i=1}^n [y_{ij} - \pi_{ij}] = 0 \text{ dan } \sum_{i=1}^n \sum_{j=1}^J X_{ij} [y_{ij} - \pi_{ij}] = 0$$

Bukti:

Fungsi *Likelihood* dari sampel random berukuran n (n responden) adalah

$$L(\beta) = \prod_{i=1}^n \pi_{i1}^{y_{i1}} \dots \pi_{ij}^{y_{ij}}$$

dengan $y_{ij} = 1$ jika responden i memilih j dan $y_{ij} = 0$ jika responden i memilih selain j. Log dari fungsi *Likelihood* tersebut adalah

$$\begin{aligned} LL(\beta) &= \sum_{i=1}^n \sum_{j=1}^J y_{ij} \ln(\pi_{ij}) \\ &= \sum_{i=1}^n \sum_{j=1}^J y_{ij} \left(X_{ij}^t \beta - \ln \left(\sum_{j=1}^J \exp(X_{ij}^t \beta) \right) \right) \\ &= \sum_{i=1}^n \sum_{j=1}^J y_{ij} X_{ij}^t \beta - \sum_{i=1}^n \left(\ln \left(\sum_{j=1}^J \exp(X_{ij}^t \beta) \right) \sum_{j=1}^J y_{ij} \right) \\ &= \sum_{i=1}^n \sum_{j=1}^J y_{ij} X_{ij}^t \beta - \sum_{i=1}^n \ln \left(\sum_{j=1}^J \exp(X_{ij}^t \beta) \right) \end{aligned}$$

Derivatif $LL(\beta)$ terhadap β adalah

$$\begin{aligned} \frac{\partial L(\beta)}{\partial \beta} &= \sum_{i=1}^n \sum_{j=1}^J X_{ij} y_j - \sum_{i=1}^n \sum_{j=1}^J X_{ij} \left(\frac{\exp(X_j^t \beta)}{\sum_{j=1}^J \exp(X_j^t \beta)} \right) \\ &= \sum_{i=1}^n \sum_{j=1}^J X_{ij} [y_j - \pi_j] = 0 \end{aligned} \quad \text{(terbukti)}$$

Penyelesaian dari persamaan penaksir dari Teorema 1.4 dapat digunakan iterasi dengan menggunakan metode *Newton Raphson* dan metode *Scoring*.

Analisis regresi untuk respon nominal dengan menggunakan model logistik maupun model *Discrete Choice* yang didasarkan pada distribusi nilai ekstrim menghasilkan fungsi probabilitas yang sama. Penaksir parameter didapatkan dengan cara menyelesaikan persamaan penaksir secara iterasi.

1.6 Contoh kasus model *multinomial logit*

Diambil data dari survei Demografi dan Kesehatan di El Salvador (Rodriguez, 2001). Survei terhadap 3165 wanita yang masih bersuami yang diklasifikasikan ke dalam kelompok usia (interval 5 tahun) dan penggunaan alat kontrasepsi yang diklasifikasikan dalam "steril", "lainnya" dan "none". Data disajikan dalam tabel berikut.

Tabel 1.1. Data Survei Demografi dan Kesehatan di El Salvador.

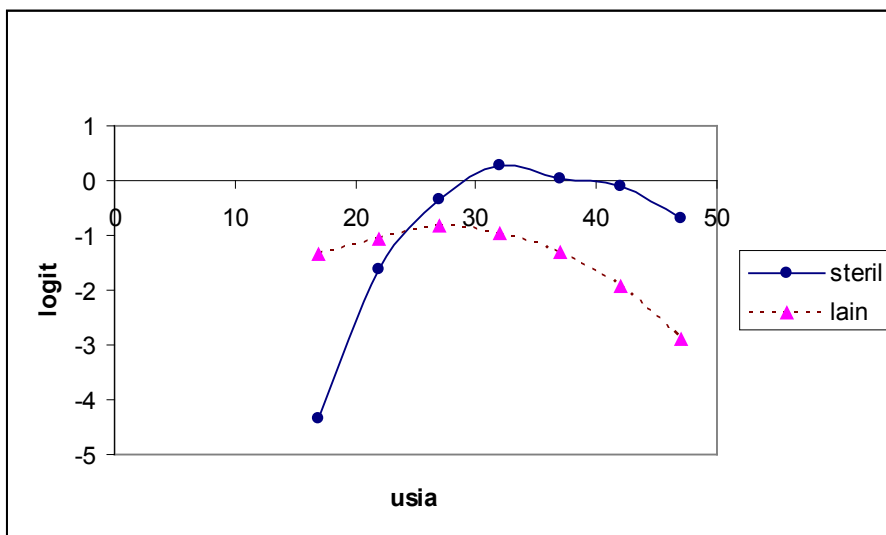
| usia | kontrasepsi | | |
|-------|-------------|---------|------|
| | Steril | lainnya | none |
| 15-19 | 3 | 61 | 232 |
| 20-24 | 80 | 137 | 400 |
| 25-29 | 216 | 131 | 301 |
| 30-34 | 268 | 76 | 203 |
| 35-39 | 197 | 50 | 188 |
| 40-44 | 150 | 24 | 164 |
| 45-49 | 91 | 10 | 183 |

Dari data pada Tabel 1.1, jika dilakukan uji independensi diperoleh nilai $\chi^2 = 521,1$ dengan derajat bebas 12. Jelas bahwa terdapat hubungan atau pengaruh usia

terhadap penggunaan kontrasepsi. Selanjutnya jika dilakukan analisis dengan model regresi multinomial maka dapat digunakan model *Logit* dengan fungsi *Logit*

$$\log\left(\frac{\pi_j}{\pi_{i3}}\right) = \beta_0 + \beta_1 Usia + \beta_2 (Usia)^2$$

Kategori ke-3 (*none*) dipakai sebagai *baseline*. Nilai untuk variabel usia diambil nilai tengah untuk masing-masing kelompok usia. Pengaruh usia terhadap *Logit* diduga berbentuk polinomial orde dua, sebab dari Gambar 1.3. plot usia terhadap nilai logit berbentuk kurva.



Gambar 1.3. Plot usia terhadap *logit*.

Beberapa paket program statistika telah menyediakan fasilitas analisis untuk model regresi multinomial, misalnya program SPSS dan MINITAB. Berikut hasil estimasi parameter beserta statistik ujinya:

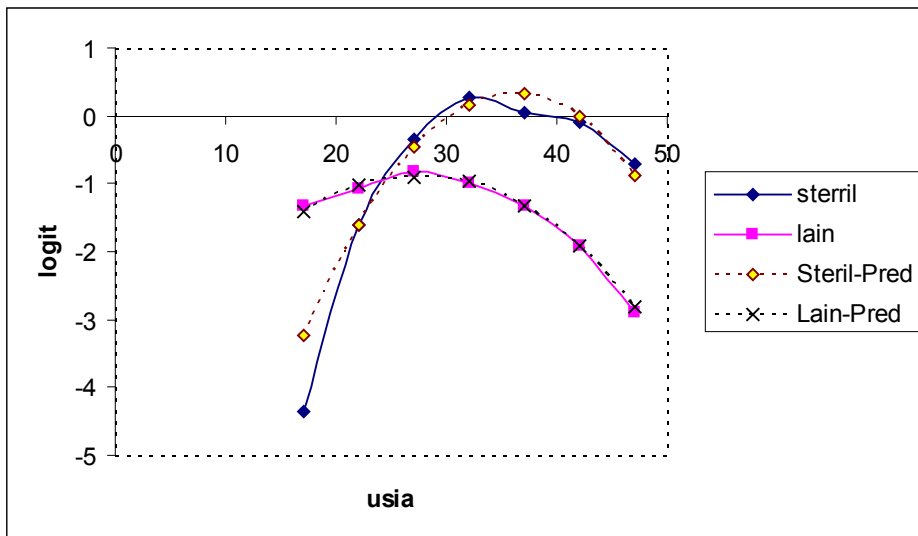
Tabel 1.2. Hasil estimasi parameter model multinomial.

| Respon | Variabel | β | Std. Error | Wald | df | P-Value | Exp(B) | 95% Confidence Interval for Exp(B) | |
|--------|-----------|---------|------------|---------|----|---------|--------|------------------------------------|-------------|
| | | | | | | | | Lower Bound | Upper Bound |
| 1.00 | Intercept | -12.266 | .735 | 278.468 | 1 | .000 | | | |
| | Usia | .700 | .045 | 243.551 | 1 | .000 | 2.014 | 1.844 | 2.199 |
| | Usia^2 | -.010 | .001 | 218.250 | 1 | .000 | .990 | .989 | .992 |

| Respon | Variabel | β | Std. Error | Wald | df | P-Value | Exp(B) | 95% Confidence Interval for Exp(B) | |
|--------|-----------|---------|------------|--------|----|---------|--------|------------------------------------|-------------|
| | | | | | | | | Lower Bound | Upper Bound |
| 2.00 | Intercept | -4.419 | .671 | 43.398 | 1 | .000 | | | |
| | Usia | .259 | .046 | 31.340 | 1 | .000 | 1.296 | 1.184 | 1.419 |
| | Usia^2 | -.005 | .001 | 39.233 | 1 | .000 | .995 | .994 | .997 |

Dari Tabel 1.2 dapat disimpulkan bahwa pengaruh usia terhadap logit berbentuk linear dan kuadratis. Nilai *Chi-Square* untuk statistik *Pearson* adalah 18,869 dengan derajat bebas 8 dan untuk statistik *Deviance* besarnya 20,475 dengan derajat bebas 8. Jadi dari nilai statistik ini berarti model layak digunakan.

Perbandingan *Logit* aktual dan *Logit* prediksi disajikan dalam Gambar 1.4. Secara umum model yang diperoleh dapat mempredikasikan dengan baik.



Gambar 1.4. Plot usia terhadap *logit* pada nilai aktual dan nilai prediksi

1.7 Sifat penaksir MLE dalam model *logit*

Model Multinomial Logit (MNL) didasarkan atas asumsi bahwa komponen errornya berdistribusi *Extreme Value* tipe I (Gumbel) dan saling independen. Dengan asumsi ini akan menghasilkan persamaan probabilitas masing-masing alternatif dalam bentuk persamaan tertutup dan konsisten dengan sifat *random utility*. MNL memenuhi sifat DCM, yaitu pembuat keputusan memilih alternatif yang memberikan utiliti terbesar.

Estimasi parameter dalam MNL dapat menggunakan metode *Maximum Likelihood Estimator* (MLE). MLE ini mempunyai sifat-sifat yang baik untuk sampel besar, khususnya *asymptotically efficient* (Horowitz, 2001). Salah satu sifat penting dalam MNL adalah rasio probabilitas dua buah alternatif tidak tergantung pada alternatif lain selain kedua alternatif tersebut. Sifat ini dinamakan *Independence from Irrelevant Alternatives* (IIA).

Sifat IIA didasarkan pada asumsi bahwa ε_{ij} saling independen untuk semua i dan j . Munizaga, Heydecker dan Ortuzar (1997) telah menunjukkan bahwa dengan mengabaikan korelasi antar alternatif akan menghasilkan estimator bias. Train (2003) menyampaikan bahwa MNL dapat merepresentasikan variasi sistematis dari karakteristik pembuat keputusan tetapi tidak dapat menangani adanya korelasi antar pilihan. Untuk data respon biner, Prentice (1988) telah menyampaikan strategi untuk menangani adanya korelasi antar respon dengan menggunakan *Generalized Estimating Equation* (GEE) untuk mendapatkan estimator/ penaksir parameter yang bersifat konsisten dan asimtotis normal.

Dalam sub bab ini dibahas bagaimana pengaruh adanya korelasi antar alternatif dalam MNL terhadap sifat estimator parameter menggunakan metode *Maksimum Likelihood Estimator*. Pengamatan didasarkan pada data simulasi dan estimasi parameter maupun perhitungan probabilitas alternatifnya menggunakan program R.

1.7.1 Studi data simulasi

Tujuan dari penelitian ini adalah mengamati pengaruh adanya korelasi antar alternatif pada MNL terhadap estimatornya. Data diperoleh dari membangkitkan data dengan nilai korelasi antar alternatifnya ditentukan. Pengamatan dilakukan dengan mengambil model untuk tiga alternatif dengan memasukan variabel atribut pembuat keputusan (X_i) dan variabel atribut masing masing alternatif (Z_{ij}). Variabel X biasa disebut variabel sosio ekonomik/geografi, misalnya penghasilan, jenis kelamin, asal daerah, jumlah anak. Sedangkan variabel Z_{ij} misalkan untuk pilihan penggunaan alat transportasi (Bus, mobil pribadi, sepeda motor) maka Z_{ij} dapat berupa waktu tempuh, biaya.

Model utilitinya adalah

$$U_{ij} = X_i\beta_j + Z_{ij}\gamma + \varepsilon_{ij} \text{ untuk } i=1,2,\dots,n \text{ dan } j=1,2,3.$$

$$\begin{aligned}
U_{i1} &= \beta_{01} + X_i\beta_1 + Z_{i1}\gamma + \varepsilon_{i1}, \\
U_{i2} &= \beta_{02} + X_i\beta_1 + Z_{i2}\gamma + \varepsilon_{i2} \\
U_{i3} &= \beta_{03} + X_i\beta_3 + Z_{i3}\gamma + \varepsilon_{i3}
\end{aligned}
\tag{1.19}$$

Dengan mengambil alternatif ke-tiga sebagai *baseline*, maka model terestimasi menjadi

$$\begin{aligned}
U_{i1}^* &= \beta_{01}^* + X_i\beta_{13} + Z_{i1}\gamma + \varepsilon_{i1}; \\
U_{i2}^* &= \beta_{02}^* + X_i\beta_{23} + Z_{i2}\gamma + \varepsilon_{i2}; \\
U_{i3}^* &= Z_{i3}\gamma + \varepsilon_{i3}
\end{aligned}
\tag{1.20}$$

dimana $\beta_{13} = \beta_1 - \beta_3$, $\beta_{23} = \beta_2 - \beta_3$, $\beta_{01}^* = \beta_{01} - \beta_{03}$ dan $\beta_{02}^* = \beta_{02} - \beta_{03}$. Jadi terdapat 5 buah parameter yang akan diestimasi.

Data dibangkitkan pada nilai parameter $\beta_{01} = 2$, $\beta_{02} = 1$, $\beta_{03} = 0.2$, $\beta_1 = -3$, $\beta_2 = -2$, $\beta_3 = -1$ dan $\gamma = 0.8$. Diambil 6 struktur kovariansi ε_i , $\text{Cov}(\varepsilon) = \Sigma$ yaitu $\Sigma_A, \Sigma_B, \Sigma_C, \Sigma_D, \Sigma_E, \Sigma_F$.

Model 1 dibangkitkan pada nilai kovariansi

$$\Sigma_A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Model 2 dibangkitkan pada nilai kovariansi

$$\Sigma_B = \begin{pmatrix} 1 & 0.25 & 0.25 \\ 0.25 & 1 & 0.25 \\ 0.25 & 0.25 & 1 \end{pmatrix}$$

Model 3 dibangkitkan pada nilai kovariansi

$$\Sigma_C = \begin{pmatrix} 1 & 0.5 & 0.5 \\ 0.5 & 1 & 0.5 \\ 0.5 & 0.5 & 1 \end{pmatrix}$$

Model 4 dibangkitkan pada nilai kovariansi

$$\Sigma_D = \begin{pmatrix} 1 & 0.9 & 0.9 \\ 0.9 & 1 & 0.9 \\ 0.9 & 0.9 & 1 \end{pmatrix}$$

Model 5 dibangkitkan pada nilai kovariansi

$$\Sigma_E = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0.5 \\ 0 & 0.5 & 1 \end{pmatrix}$$

Model 6 dibangkitkan pada nilai kovariansi

$$\Sigma_F = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0.9 \\ 0 & 0.9 & 1 \end{pmatrix}$$

$X_i \sim \text{NID}(0,1)$, $Z_{ij} \sim \text{NID}(0,1)$ dan $\varepsilon_i \sim N(0, \Sigma)$

Untuk variabel X_i dan Z_{ij} yang berupa data diskrit tidak disampaikan di sini karena menghasilkan kesimpulan yang sama dengan data kontinu. Dipilih $n=100$ dan replikasi masing-masing kovariansi sebanyak 20, 50 dan 100.

Untuk mendapatkan data multivariat normal tersebut digunakan persamaan

$$\varepsilon = L\eta \text{ dan } \Sigma = LL^t$$

di mana $\eta \sim N(0,1)$ dan L didefinisikan sebagai matrik segitiga bawah dari faktor *Cholesky*. Program R menyediakan fasilitas membangkitkan data multivariat normal dalam *library "MASS"* dan program estimasi MLE terdapat dalam *library "MicEcon"* (Henningsen, 2007).

Karena data ε_i dibangkitkan dari distribusi multivariat normal, sementara itu MNL didasarkan pada distribusi nilai ekstrim, maka diperlukan normalisasi sebagai berikut:

$$\tilde{U}_i = U_{i^*}^* \cdot \sqrt{\pi^2/6} = U_{i^*}^* \cdot \sqrt{1.6}$$

Sehingga persamaan (1.19) menjadi

$$\tilde{U}_{i1} = (\beta_{01}^* + X_i\beta_{13} + Z_{i1}\gamma) \sqrt{1.6} + \varepsilon_{i1} \sqrt{1.6};$$

$$\tilde{U}_{i2} = (\beta_{02}^* + X_i\beta_{23} + Z_{i2}\gamma) \sqrt{1.6} + \varepsilon_{i2} \sqrt{1.6};$$

$$\tilde{U}_{i3} = Z_{i3}\gamma \sqrt{1.6} + \varepsilon_{i3} \sqrt{1.6}$$

Data dibangkitkan pada nilai parameter

$\beta_{01} = 2, \beta_{02} = 1, \beta_{03} = 0.2, \beta_1 = -3, \beta_2 = -2, \beta_3 = -1$ dan $\gamma = 0.8$.

Jadi

$\beta_{01}^* = 1.8, \beta_{02}^* = 0.8, \beta_{13} = -2, \beta_{23} = -1$ dan $\gamma = 0.8$.

Dengan adanya faktor pengali $\sqrt{1.6}$ maka estimator targetnya (yang diharapkan) adalah

$B_{01} = 2.27684, B_{02} = 1.011929, B_1 = -2.529822, B_2 = -1.264911$ dan $C = 1.011929$.

Program disusun dalam dua tahap, pertama adalah proses membangkitkan data dengan distribusi dan struktur kovariansi tertentu. Kedua, adalah melakukan estimasi parameter untuk mendapatkan model MLN. Program dapat disusun sebagai berikut:

```
#PROGRAM R
npara=5 #jumlah parameter yang diestimasi
N=500 #jumlah sampel
X=rnorm(N,3)
E=matrix(N,3)
Z=matrix(N,3)
U=matrix(N,3)
V=matrix(N,3)
Rep=20 #jumlah replikasi
esti=array(0,dim=c(npara,4,Rep))
s12=0 #korelasi 1 dan 2
s13=0 #korelasi 1 dan 3
s23=0 #korelasi 2 dan 3
sigma=matrix(c(1,s12,s13,s12,1,s23,s13,s23,1),3,3)
```

```

for (i in 1:3){Z[,i]=rnorm(N)}
B01=2;
B02=1 ;
B03=0.2 ;
B1=-3 ;
B2=-2 ;
B3=-1 ;
G=0.8
V[,1]=B01-B03 + X*(B1-B3) + Z[,1]*G
V[,2]=B02-B03 + X*(B2-B3) + Z[,2]*G
V[,3]= Z[,3]*G
n=1
EE=array(0,dim=c(N,3,Rep))
for (i in 1:Rep)
{
E=mvrnorm(N,c(0,0,0),sigma)
EE[,i]=E
U[,1]=V[,1]+ E[,1]
U[,2]=V[,2]+ E[,2]
      U[,3]=V[,3]+ E[,3]
Y=matrix(0,N,3)
for(i in 1:N){if (U[i,1] > U[i,3] & U[i,1] >U[i,2]) Y[i,1]=1 else (if (U[i,2] > U[i,1] & U[i,2] >
U[i,3] ) Y[i,2]=1 else Y[i,3]=1))
hasil=summary( maxLik(MLE.logit, start=c(0,0,0,0,0)))
hasil2=hasil[6]$estimate
esti[,i]<-hasil2
n=n+1
}

```

```

MLE.logit<-function(a)
{
b01=a[1];b02=a[2];b1=a[3]
b2=a[4]
g=a[5]
p1=sum(Y[,1]*(b01 + X*b1 + Z[,1]*g)- Y[,1]*log(exp(b01 + X*b1 + Z[,1]*g)+exp(b02 +
X*b2 + Z[,2]*g)+exp(Z[,3]*g)))
p2=sum(Y[,2]*(b02 + X*b2 + Z[,2]*g)- Y[,2]*log(exp(b01 + X*b1 + Z[,1]*g)+exp(b02 +
X*b2 + Z[,2]*g)+exp(Z[,3]*g)))
p3=sum(Y[,3]* (Z[,3]*g)- Y[,3]*log(exp(b01 + X*b1 + Z[,1]*g)+exp(b02 + X*b2 +

```


$$Z_{[2]} * g) + \exp(Z_{[3]} * g))$$

$$lg = p_1 + p_2 + p_3$$

$$lg$$

$$\}$$

Menyusun nilai utiliti agar parameternya *estimable* sangat penting untuk dilakukan. Pada utilitas asli mempunyai 7 parameter, tetapi hanya 5 yang terestimasi. Pengurangan jumlah parameter ini tidak mengubah probabilitas pilihan maupun interpretasinya. Dengan mengambil parameter

$$B_{01}=2, B_{02}=1, B_{03}=0.2, B_1=-3, B_2=-2, B_3=-1 \text{ dan } \gamma=0.8$$

maka utilitas masing-masing alternatif dan individu adalah

$$U_{i1}^* = 1.8 - 2X_i + 0.8Z_{i1} + \epsilon_{i1}$$

$$U_{i2}^* = 0.8 - X_i + 0.8Z_{i2} + \epsilon_{i2}$$

$$U_{i3}^* = 0.8Z_{i3} + \epsilon_{i3}$$

Model probabilitas masing alternatif dan individu yang akan diestimasi adalah

$$\hat{P}_{1i} = \frac{\exp(b_{10} + b_1 X_i + c Z_{1i})}{\exp(b_{10} + b_1 X_i + c Z_{1i}) + \exp(b_{20} + b_2 X_i + c Z_{2i}) + \exp(c Z_{3i})}$$

$$\hat{P}_{2i} = \frac{\exp(b_{20} + b_2 X_i + c Z_{2i})}{\exp(b_{10} + b_1 X_i + c Z_{1i}) + \exp(b_{20} + b_2 X_i + c Z_{2i}) + \exp(c Z_{3i})}$$

$$\hat{P}_{3i} = \frac{\exp(c Z_{3i})}{\exp(b_{10} + b_1 X_i + c Z_{1i}) + \exp(b_{20} + b_2 X_i + c Z_{2i}) + \exp(c Z_{3i})}$$

dan $p_{1i} + p_{2i} + p_{3i} = 1$. Estimator yang akan dicari adalah b_{10}, b_{20}, b_1, b_2 dan c .

Data yang dibangkitkan untuk masing-masing kovariansi sebanyak 17.000 (dengan $n=100$, pada replikasi 20, 50 dan 100). Di samping itu juga dibangkitkan data $n=500$ dengan replikasi 20.

1.7.2 Bias dan nilai pseudo R² pada korelasi tinggi

Diambil tiga nilai replikasi, yaitu 20, 50 dan 100 dimaksudkan untuk melihat kestabilan masing-masing estimator. Diperoleh hasil bahwa rata-rata estimator yang dihasilkan pada ketiga nilai replikasi tidak berbeda secara signifikan. Pengujian dilakukan dengan uji anova satu arah, dengan hipotesis $H_0 : \theta_{R1} = \theta_{R2} = \theta_{R3}$. Pada model satu (Σ_A) diperoleh hasil sebagaimana Tabel 1.3.

Tabel 1.3. Estimator pada model 1

| parameter | rata-rata θ pada replikasi | | | Signifikansi |
|-----------|-----------------------------------|-----------|-----------|--------------|
| | 20 | 50 | 100 | |
| B10 | 2.445245 | 2.596763 | 2.558173 | 0.2892 |
| B20 | 1.116578 | 1.24624 | 1.231416 | 0.3738 |
| B1 | -2.732932 | -2.877022 | -2.775457 | 0.06004 |
| B2 | -1.372098 | -1.550125 | -1.429905 | 0.04827 |
| γ | 1.093627 | 1.0728 | 1.077740 | 0.0689 |

Jadi secara umum dapat disimpulkan bahwa replikasi tidak cukup berpengaruh terhadap estimator. Demikian juga pada model yang lain ($\Sigma_B, \Sigma_C, \Sigma_D, \Sigma_E, \Sigma_F$) diperoleh kesimpulan yang sama, yaitu rata-rata estimator pada ketiga nilai replikasi tidak berbeda secara signifikan.

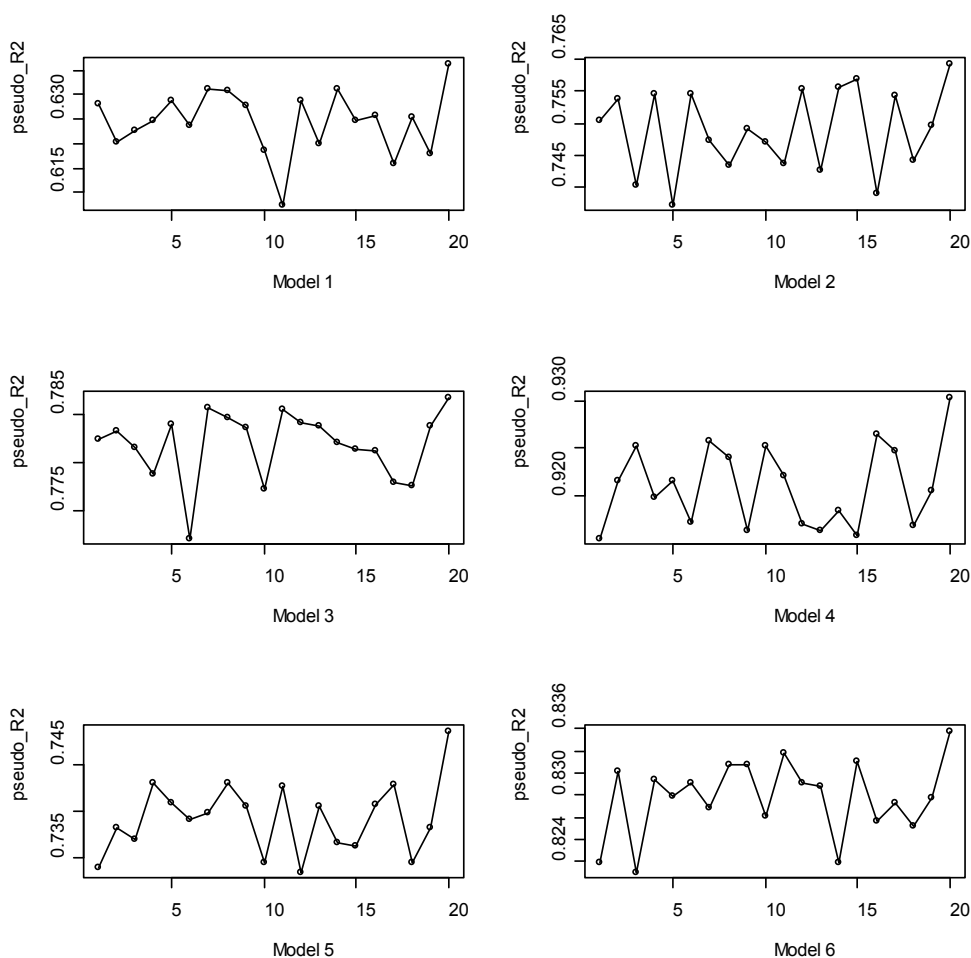
Adanya korelasi antar suku *error* (ϵ_j), dari Tabel 1.4 nampak jelas berakibat pada bias estimator terhadap parameternya, Semakin tinggi tinggi tingkat korelasinya, maka bias estimator terhadap parameter semakin besar.

Tabel 1.4. Nilai estimasi parameter untuk replikasi 20 dengan n=100

| Kovariansi | Parameter | | | | |
|--|-----------|----------|-----------|-----------|----------|
| | B01 | B02 | B1 | B2 | γ |
| $\sigma_{12} = \sigma_{13} = \sigma_{23} = 0$ | 2.445245 | 1.116578 | -2.732932 | -1.372098 | 1.093627 |
| $\sigma_{12} = \sigma_{13} = \sigma_{23} = 0.25$ | 2.971488 | 1.479699 | -3.226701 | -1.607597 | 1.214585 |
| $\sigma_{12} = \sigma_{13} = \sigma_{23} = 0.5$ | 3.848194 | 1.962332 | -4.333718 | -2.364538 | 1.567595 |
| $\sigma_{12} = \sigma_{13} = \sigma_{23} = 0.9$ | 9.361293 | 4.151584 | -1.052819 | -5.327487 | 4.140567 |
| $\sigma_{12} = \sigma_{13} = 0, \sigma_{23} = 0.5$ | 2.9512 | 1.555506 | -3.012371 | -1.595781 | 1.147604 |
| $\sigma_{12} = \sigma_{13} = 0, \sigma_{23} = 0.9$ | 3.688772 | 2.033002 | -3.475285 | -2.057577 | 1.351584 |

Yang perlu menjadi perhatian adalah pada korelasi tinggi (sekitar 0.9) maka kemungkinan tidak diperoleh estimator karena pada proses iterasi menghasilkan matrik hessian yang tidak positif definit. Pada jumlah sampel 100 dan replikasi sebanyak 100, beberapa di antaranya tidak diperoleh estimator karena matrik hessiannya tidak positif definit. Hal serupa juga terjadi ketika jumlah sampelnya relatif kecil walaupun dengan korelasi rendah. Dari beberapa percobaan, untuk 5 parameter minimum sampel yang dibutuhkan sebesar 500 agar estimatornya dapat diperoleh..

Semakin besar korelasi antar komponen *error* dari masing-masing alternatif menyebabkan semakin besar nilai bias estimator. Dari percobaan yang telah dilakukan, semakin besar korelasinya juga semakin besar nilai pseudo R2. (lihat Gambar 1.4.).

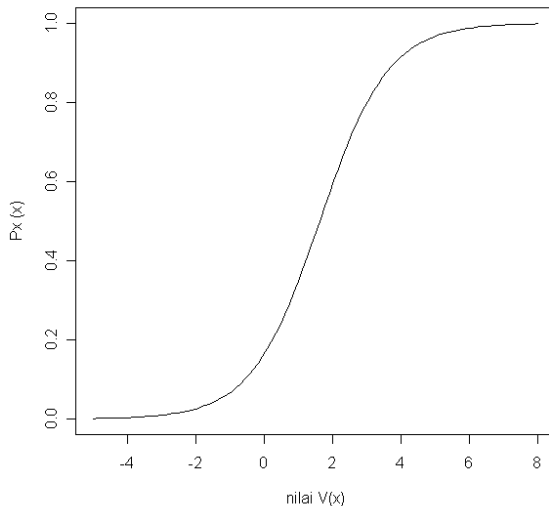


Gambar 1.5. Grafik nilai R² pada n=500 replikasi 20

Peristiwa ini dapat dijelaskan sebagai berikut: persamaan (9) dan (14) menyatakan

$$\max(P_{i1}, \dots, P_{ij}) \rightarrow 1 \Rightarrow LL(\hat{\beta}) \rightarrow 0 \Rightarrow \text{pseudo } R^2 \rightarrow 1$$

Nilai $p_{ij} \rightarrow 1$ untuk i yang memilih j dicapai pada nilai V_{ij} yang cukup besar, seperti terlihat dalam Gambar 1.5 yang merupakan hubungan antara V_{ij} dan p_{ij} pada V_{ik} tetap untuk $j \neq k$. Karena adanya korelasi antar alternatif mengakibatkan terjadinya bias dan dengan adanya bias ini berarti nilai mutlak semua estimator (koefisien) relatif lebih besar dibandingkan estimator pada korelasi rendah. Semakin besar estimator maka nilai $V(x)$ semakin besar, sebaliknya semakin kecil estimator maka nilai $V(x)$ semakin kecil.



Gambar 1.6. Grafik nilai $V(x)$ terhadap $p(x)$

Misalkan pada tiga alternatif pilihan, masing-masing mempunyai utilitas dengan komponen deterministiknya

$$V_{i1} = b_{01} + X_{i1}b_{13} + Z_{i1}c$$

$$V_{i2} = b_{02} + X_{i2}b_{23} + Z_{i2}c$$

$$V_{i3} = cZ_{i3}$$

Jika terjadi korelasi antar alternatif menghasilkan bias sebesar k (bernilai positif),

maka

$$V_{i1}^* = kb_{01} + kX_{i1}b_{13} + kZ_{i1}c = kV_{i1}$$

$$V_{i2}^* = kb_{02} + kX_{i2}b_{23} + kZ_{i2}c = kV_{i2}$$

$$V_{i3}^* = kcZ_{i3} = kV_{i3}$$

Dengan adanya faktor pengali k , maka diperoleh rasio

$$\exp(V_{i1}^*) : \exp(V_{i2}^*) : \exp(V_{i3}^*) = [\exp(V_{i1})]^k : [\exp(V_{i2})]^k : [\exp(V_{i3})]^k$$

Probabilitas masing-masing alternatifnya menjadi:

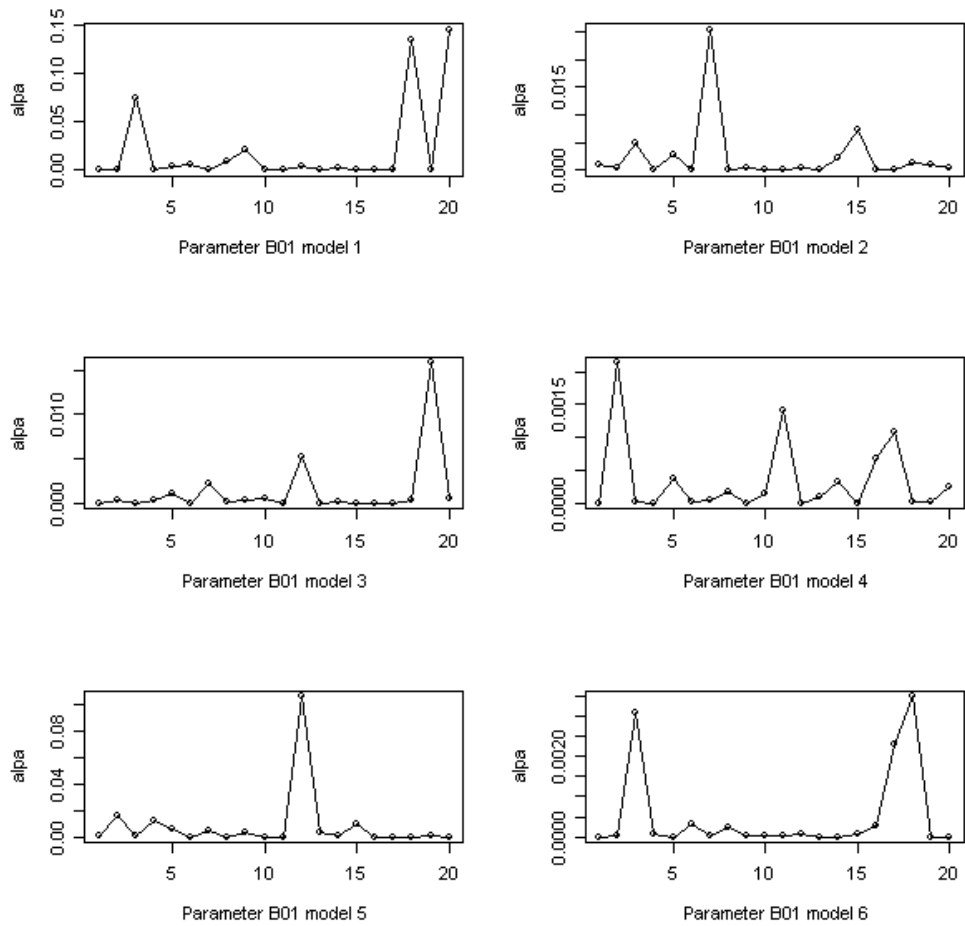
$$\hat{p}_j = \frac{[\exp(V_j)]^k}{[\exp(V_{i1})]^k + [\exp(V_{i3})]^k + [\exp(V_{i3})]^k} \text{ untuk } j=1,2,3.$$

Dengan adanya bias berakibat nilai $\min(\hat{p}_{i1}, \hat{p}_{i2}, \hat{p}_{i3}) \forall_i$ akan semakin kecil dan $\max(\hat{p}_{i1}, \hat{p}_{i2}, \hat{p}_{i3}) \forall_i$ akan semakin besar dibandingkan pada kondisi tidak ada korelasi. Selanjutnya hal ini mengakibatkan nilai psedo R^2 makin besar.

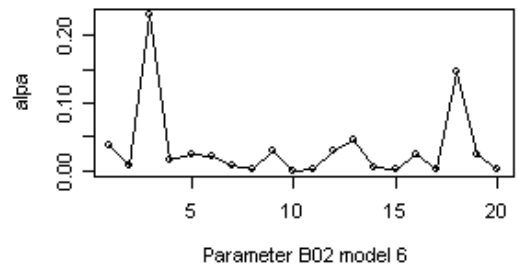
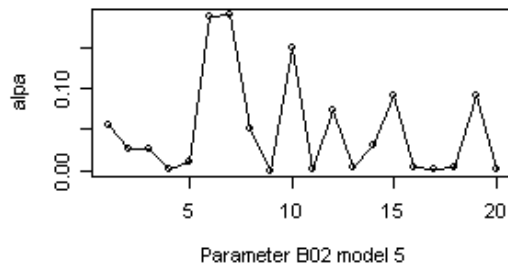
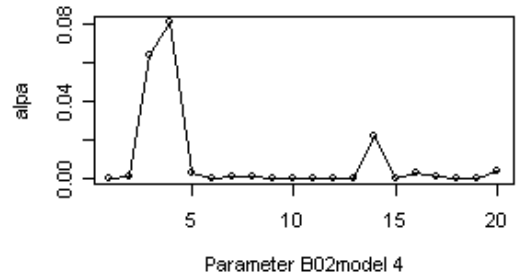
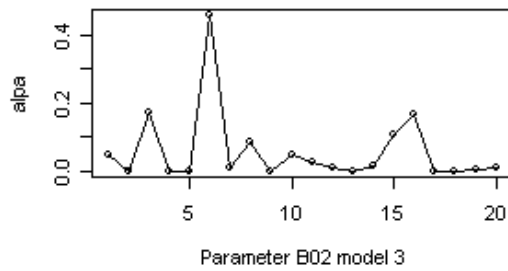
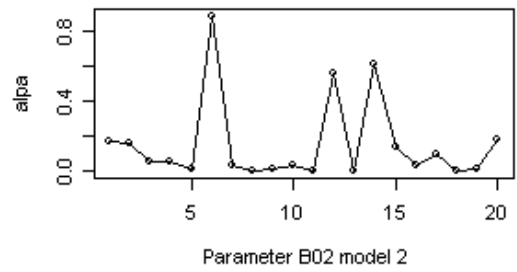
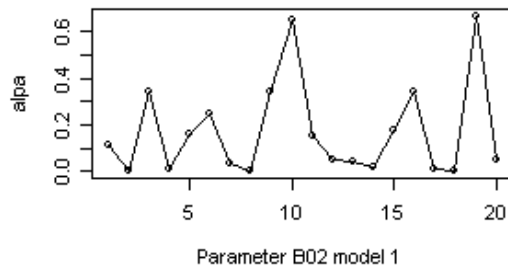
Mengingat konsekuensi adanya korelasi antar alternatif mengakibatkan nilai pseudo R^2 besar (mendekati satu), maka harus hati-hati dalam menggunakan statistik pseudo R^2 . Pada kondisi nilai $\exp(V_{i1}), \dots, \exp(V_{ij})$ atau nilai probabilitas masing-masing alternatif relatif sama, statistik *Pseudo* R^2 akan bernilai rendah.

1.7.3 Kesalahan tipe I

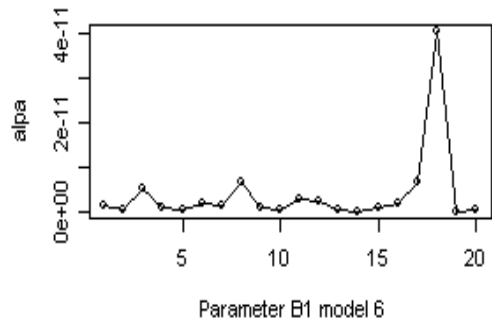
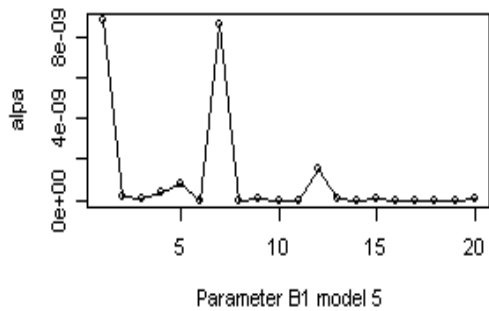
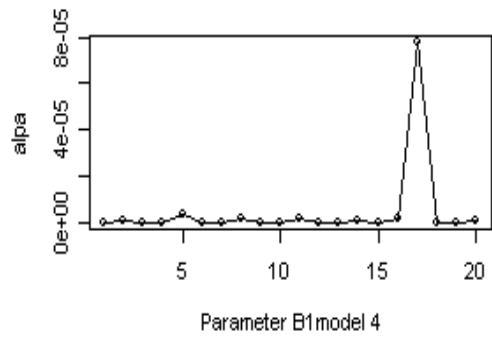
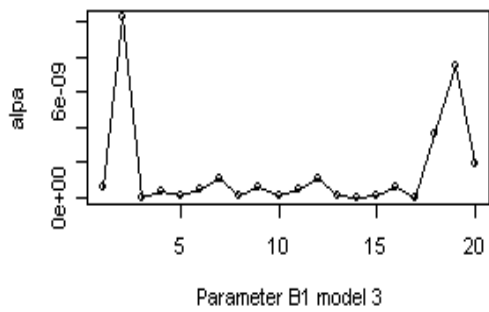
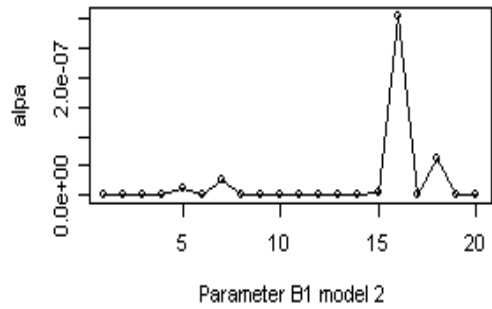
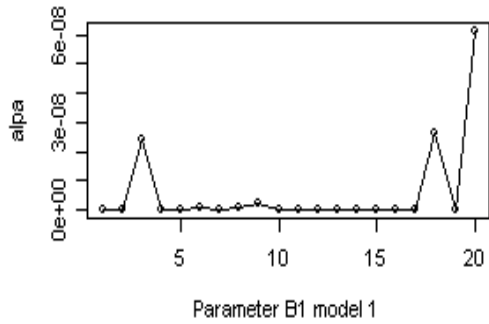
Hasil uji signifikansi parameter secara parsial dapat dilihat di Gambar 1.6 s/d Gambar 1.10. Secara umum dapat disimpulkan bahwa parameter 3, 4 dan 5 pada semua bentuk kovariansi pengaruhnya signifikan. Sementara itu untuk parameter 1 dan 2 (parameter konstanta) pada terdapat nilai alpa yang cukup besar (tidak signifikan) di antara 20 replikasi. Pada analisis berdasarkan data empirik, jika dijumpai uji parameternya hasilnya tidak signifikan sebaiknya dikeluarkan dari model. Sebagaimana pada kasus di sini, yaitu parameter konstanta (B01 dan B02) pengaruhnya tidak signifikan, maka parameter ini sebaiknya tidak masuk dalam model. Karena yang dikeluarkan dari model adalah parameter konstanta, maka akan mengakibatkan nilai Pseudo R^2 menjadi lebih rendah.



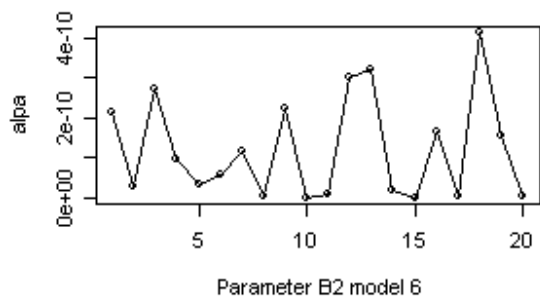
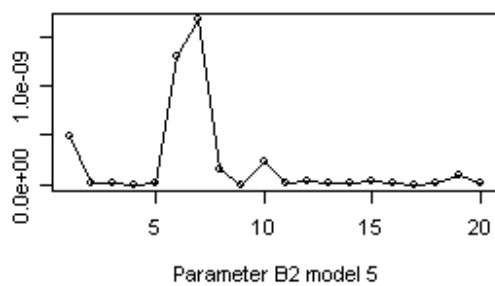
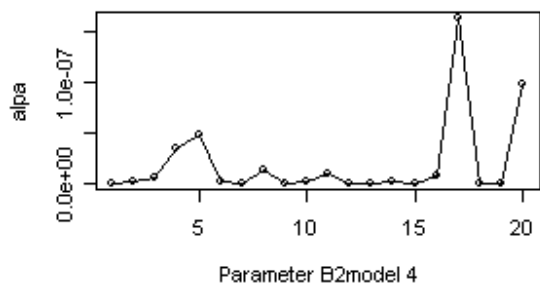
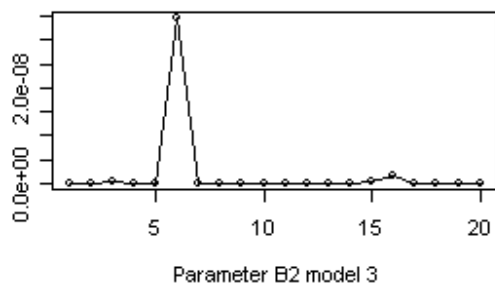
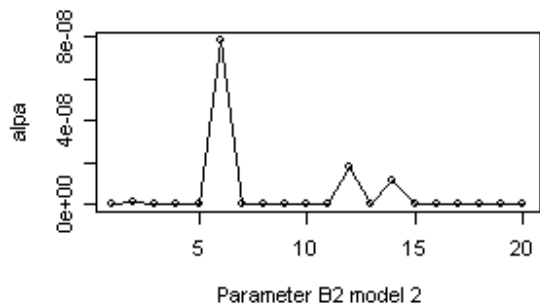
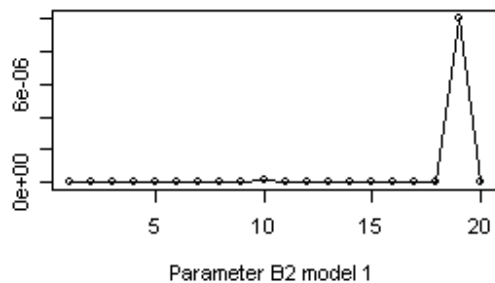
Gambar 1.7. Grafik kesalahan tipe I (alpha) pada n=500 replikasi 20 untuk parameter B01



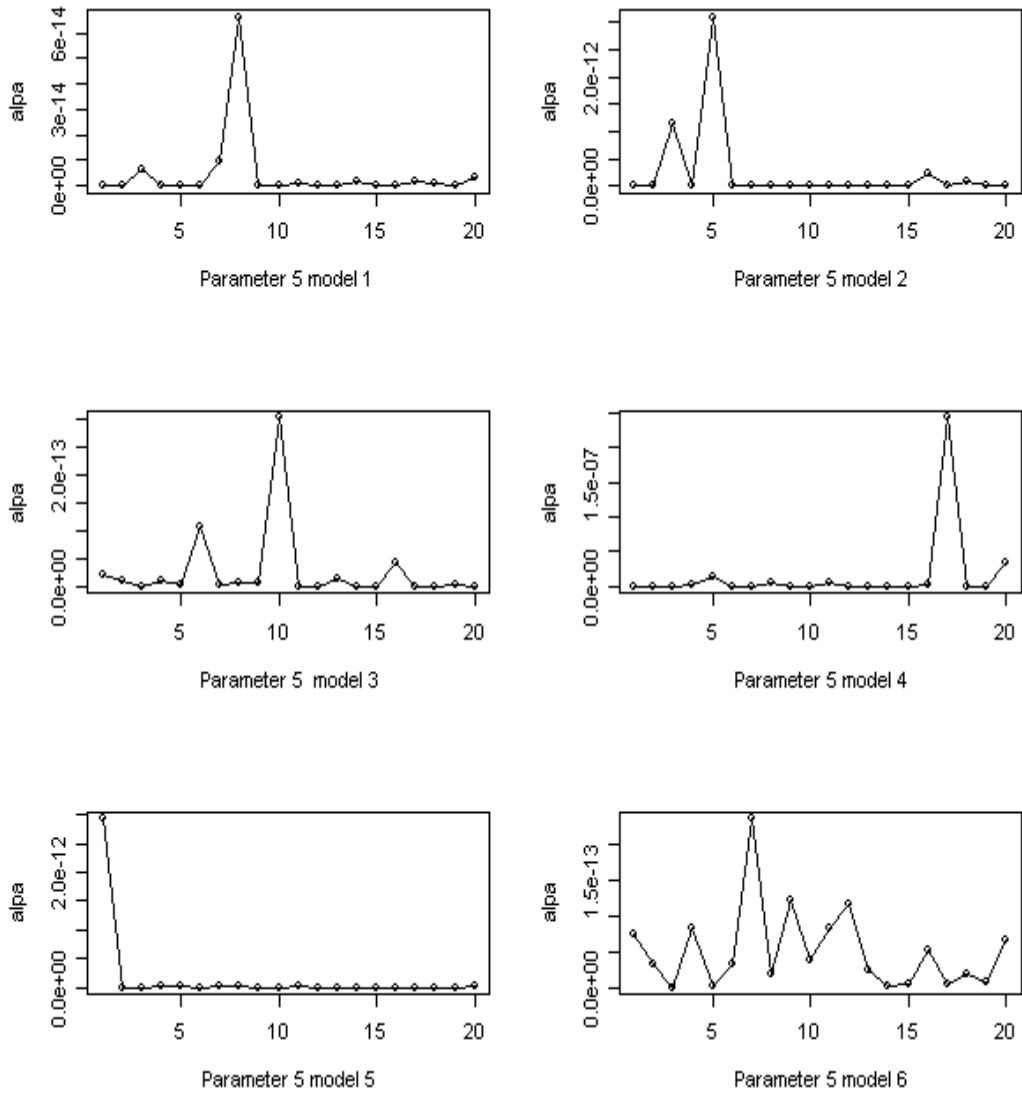
Gambar 1.8. Grafik kesalahan tipe I (alpha) pada $n=500$ replikasi 20 untuk parameter B02



Gambar 1.9. Grafik kesalahan tipe I (alpha) pada n=500 replikasi 20 untuk parameter B1



Gambar 1.10. Grafik kesalahan tipe I (alpha) pada n=500 replikasi 20 untuk parameter B2



Gambar 1.11. Grafik kesalahan tipe I (α) pada $n=500$ replikasi 20 untuk parameter γ

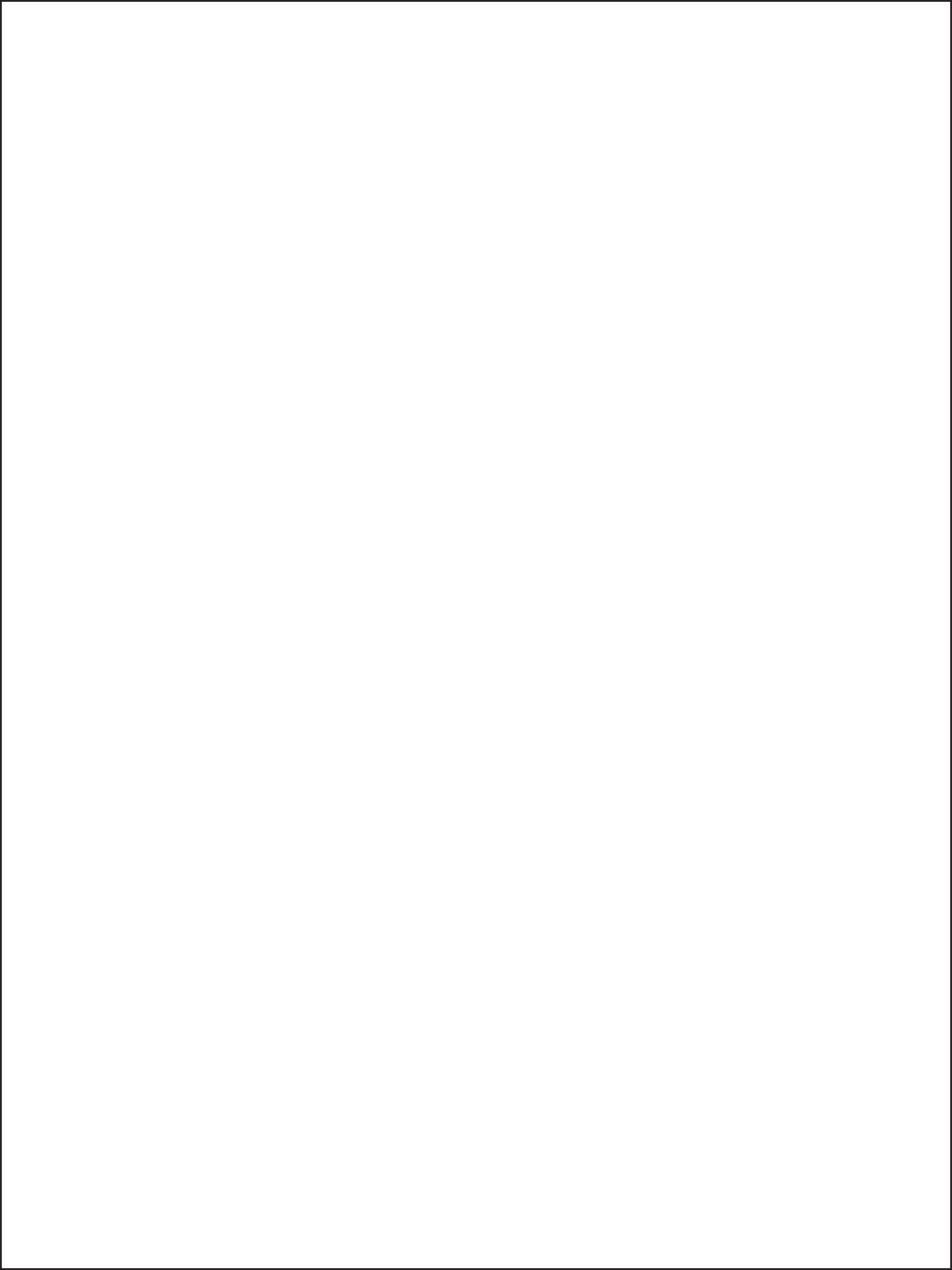
Berapa model lain yang telah dikembangkan untuk mengakomodasi adanya korelasi antar alternatif pilihan antara lain *Model Probit Multinomial* pendekatan klasik maupun bayesian. Ruud, dkk. (1996) mengembangkan strategi estimasi yang baru untuk *Model Multinomial Probit* (MNP) yaitu didasarkan pada metode *Simulated Moments*. Pendekatan Bayesian dilakukan oleh McCullach, dkk. (2000) dan Imey, dkk.

(2005).

Model lain yang telah dikembangkan dengan memasukkan unsur korelasi adalah model *Generalized extreme Value* (GEV). Model GEV merupakan pengembangan dari model logit yang disusun berdasarkan adanya korelasi antar alternatif pilihan yang tersedia. Jika semua korelasi di antara alternatif yang ada bernilai nol, maka GEV menjadi model *Logit* standar. Beberapa jenis model yang termasuk dalam GEV antara lain model *Generalized Nested Logit* (Wen dan Koppelman, 2001) dan *model Cross-Nested Logit* (Bhat, 2003).

Adanya korelasi antar komponen error masing-masing alternatif dalam model MNL mengakibatkan terdapat bias estimatornya. Dengan adanya bias ini, nilai Pseudo R^2 cenderung besar. Sehingga jika diperoleh nilai Pseudo R^2 mendekati satu, perlu dicurigai kemungkinan adanya korelasi. Untuk keperluan uji kecocokan model, harus hati-hati dalam menggunakan statistik Pseudo R^2 , apalagi pada kondisi nilai probabilitas antar alternatif relatif sama akan menghasilkan nilai Pseudo R^2 yang kecil.

Agar estimator masing-masing parameter dapat diperoleh dan dilakukan pengujian, untuk model MNL dengan 5 parameter disarankan menggunakan sampel sebesar 500 atau lebih. Jika diduga terdapat korelasi, perlu dipertimbangkan penggunaan model yang lain seperti model MNP maupun Model GEV.



Bab 2

Model Probit dan Model Logit pada Respons Biner

Model linear adalah model yang sederhana dalam estimasi parameter dan penggunaannya, tetapi tidak cocok untuk respons yang mempunyai nilai terbatas seperti respons berupa probabilitas (bernilai 0 s/d 1), respons persentase (0 s/d 100), respons nilai IPK (0 s/d 4) (Greene, 2003). Dalam model respon biner, akan dicari

$$P(y=1|x) = P(y=1|x_1, \dots, x_k) \quad (2.1)$$

di mana x menyatakan kumpulan variabel independen/eksplanatori. Sebagai contoh y adalah status bekerja (bekerja atau menganggur), x mungkin variabel karakteristik individu seperti pendidikan, usia, dll.

Dalam model nonlinear, misalkan didefinisikan fungsi $G(\cdot)$ untuk menjamin nilainya antara null dan 1, $0 \leq G(z) \leq 1$ untuk sebarang bilangan real z . Sehingga dapat disusun model

$$P(y=1|x) = G(\beta_0 + \beta_1 x_1 + \dots + \beta_k x_k) = G(\beta_0 + x\beta) \quad (2.2)$$

Untuk model *Logit*, didefinisikan $G(\cdot)$,

$$G(z) = \frac{\exp(z)}{[1 + \exp(z)]} = \Lambda(z) \quad (2.3)$$

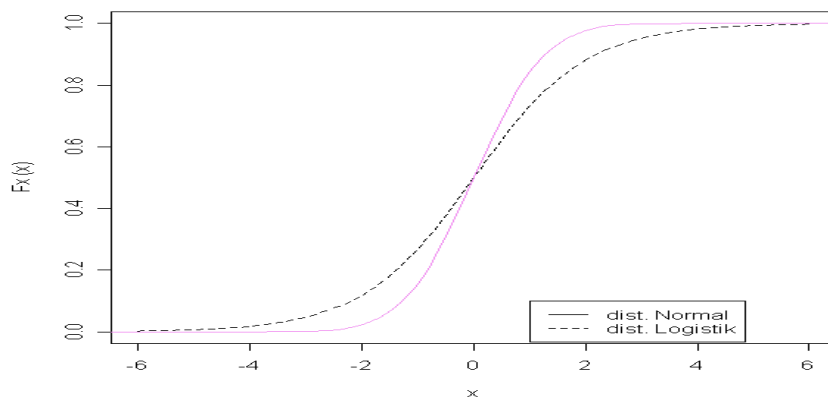
$G(z)$ adalah fungsi distribusi kumulatif untuk variabel *random* distribusi Logistik standar. Untuk model *Probit*, $G(\cdot)$ adalah fungsi distribusi Kumulatif Normal Standar yang dapat dituliskan sebagai

$$G(z) = \Phi(z) \equiv \int_{-\infty}^z \phi(v) dv \quad (2.4)$$

di mana $\phi(z)$ adalah

$$\phi(z) = (2\pi)^{-1/2} \exp(-z^2/2) \quad (2.5)$$

Fungsi $G(\cdot)$ pada model *Logit* dan model *Probit* merupakan kurva yang saling berimpit, seperti terlihat dalam gambar 2.1.



Gambar 2.1. Grafik distribusi kumulatif normal standar dan distribusi logistik

Selanjutnya akan dibahas spesifikasi model *Logit* dan model *Probit*, estimasi parameter dengan MLE, statistik uji, dan interpretasi estimator.

2.1 Spesifikasi model *probit* dan model *logit*

Model *Logit* dan model *Probit* dapat diturunkan dari model variabel laten yang memenuhi asumsi-asumsi pada model linear klasik. Misal y^* adalah variabel yang tidak terobservasi atau variabel laten, yang dihitung dari

$$y^* = \beta_0 + x\beta + e \text{ dan } y = 1[y^* > 0] \quad (2.6)$$

Di mana notasi $1(\cdot)$ merupakan fungsi indikator yang mempunyai nilai 1 jika kejadian memenuhi sifat dalam tanda kurung dan bernilai null jika tidak sesuai dengan sifat dalam tanda kurung. Kita mengasumsikan bahwa e adalah independen dengan x dan bahwa e mempunyai distribusi Logistik standar atau berdistribusi standar normal. Dari gambar 2.1. terlihat bahwa e berdistribusi simetrik dititik null, sehingga

$$1-G(z) = G(z)$$

Untuk semua bilangan real z . Dengan asumsi ini kita dapat menurunkan probabilitas respon untuk y :

$$P(y=1|x) = P(y^* > 0|x) = P(e > -(\beta_0 + x\beta)|x)$$

$$= 1 - G[-(\beta_0 + x\beta)] = G(\beta_0 + x\beta) \quad (2.7)$$

Persamaan ini tepat sama dengan persamaan (2).

Dalam banyak aplikasi model respon biner, tujuan utamanya adalah menjelaskan efek perubahan x terhadap probabilitas $P(y=1|x)$. Dengan formulasi pada variabel laten, kita dapat menghitung efek x terhadap variabel laten y^* . Sebagaimana yang akan kita lihat, untuk model *Logit* dan *Probit*, secara langsung efek masing-masing x_j pada

$$E(y^*|x) = \beta_0 + \beta_1 x$$

dan pada

$$E(y|x) = P(y=1|x) = G(\beta_0 + \beta_1 x)$$

adalah selalu sama.

Untuk mencari efek parsial variabel x yang kontinu pada $p(x) = P(y=1|x)$ dapat diperoleh dari derivatif parsial

$$\frac{\partial p(x)}{\partial x_j} = g(\beta_0 + x\beta)\beta_j \quad g(z) \equiv \frac{dG}{dz}(z) \quad (2.8)$$

$g(\cdot)$ adalah fungsi densitas. Efek x_j pada $p(x)$ akan mempunyai tanda yang sama dengan tanda dari β_j sebab $g(\cdot)$ selalu bernilai positif.

Dari persamaan (2.8), menunjukkan efek relatif sebarang dua nilai variabel x tidak tergantung pada x . Rasio efek parsial untuk x_j dan x_k adalah β_j/β_k .

Dalam kasus khusus, jika $g(\cdot)$ adalah simetrik pada titik null maka efek terbesar terjadi pada $\beta_0 + \beta_1 x = 0$.

Sebagai contoh, pada model *Probit*

$$g(z) = \phi(z)$$

$$g(0) = \frac{1}{\sqrt{2\pi}} \approx 0,4$$

Pada model Logit

$$g(z) = \frac{\exp(z)}{[1 + \exp(z)]^2}$$

sehingga $g(0) = 0,25$

Jika x_1 adalah variabel biner, maka efek parsial merupakan perubahan x_1 dari null ke satu pada kondisi nilai x yang lain tertentu, yaitu

$$\text{efek } x_1 = G(\beta_0 + \beta_1 + \beta_2 x_2 + \dots + \beta_k x_k) - G(\beta_0 + \beta_2 x_2 + \dots + \beta_k x_k) \quad (2.9)$$

Persamaan (2.9) ini tergantung pada nilai x_j yang lain. Misal y adalah status kerja dan x_1 adalah variabel *dummy* (seperti keikutsertaan dalam pelatihan kerja), maka persamaan (2.9) merupakan perubahan probabilitas dari menganggur ke probabilitas mendapat kerja pada x yang lain bernilai tertentu seperti (usia dan pendidikan, pengalaman).

Kita juga dapat menggunakan persamaan (2.9) untuk variabel diskrit yang lain misalnya jumlah anak, Jika x_k menyatakan variabel diskrit, maka efek probabilitas dari x_k dari c_k ke $(c_k + 1)$ yang disederhanakan menjadi efek

$$x_k = G(\beta_0 + \beta_{1x_1} + \beta_{2x_2} + \dots + \beta_{k(c_k+1)}) - G(\beta_0 + \beta_{1x_1} + \beta_{2x_2} + \dots + \beta_{kc_k}) \quad (2.10)$$

Untuk fungsi yang lain, misalkan

$$P(y = 1 | z) = G(\beta_0 + \beta_1 z_1 + \beta_2 z_1^2 + \beta_3 \log(z_2) + \beta_4 z_3) \quad (2.11)$$

efek parsial dari z_1 pada $P(y=1|z)$ adalah

$$\frac{\partial P(y = 1 | z)}{\partial z_1} = g(\beta_0 + x\beta)(\beta_1 + 2\beta_2 z_1) \quad (2.12)$$

efek parsial dari z_2 pada $P(y=1|z)$ adalah

$$\frac{\partial P(y = 1 | z)}{\partial z_2} = g(\beta_0 + x\beta)(\beta_3 z_2) \quad (2.13)$$

di mana,

$$x\beta = \beta_1 z_1 + \beta_2 z_1^2 + \beta_3 \log(z_2) + \beta_4 z_4$$

Model dengan interaksi antar variabel independen termasuk diskrit dan kontinu dapat diselesaikan dengan cara yang sama.

2.2 MLE dan uji hipotesis pada model *logit* dan *probit*

Pada model Linear, estimasi parameter dapat dilakukan dengan metode OLS (*Ordinary Least Squares*) atau bisa juga dengan WLS (*Weighted Least Squares*). Pada model nonlinear, model OLS maupun WLS tidak dapat digunakan. Metode estimasi parameter yang dapat digunakan adalah *Maximum Likelihood Estimation* (MLE) (Horowitz dan Savin, 2001). Untuk mendapatkan MLE, diperlukan fungsi densitas y_i pada nilai x tertentu.

$$f(y_i|x_i;\beta) = [G(x_i;\beta)]^{y_i}[1-G(x_i;\beta)]^{1-y_i} \text{ dengan } y_i=0,1 \quad (2.14)$$

Fungsi *Log Likelihood* untuk observasi ke- i adalah fungsi dari parameter dan data (x_i, y_i) , merupakan

$$l_i(\beta) = y_i \log [G(x_i;\beta)] + (1-y_i) \log [1-G(x_i;\beta)] \quad (2.15)$$

Karena $G(\cdot)$ bernilai antara null dan satu, maka $l_i(\beta)$ adalah *well-defined* untuk semua nilai β . *Log Likelihood* untuk sampel berukuran n adalah

$$L(\beta) = \sum_{i=1}^n l_i(\beta) \quad (2.16)$$

MLE untuk $\hat{\beta}$ yang dinotasikan dengan adalah nilai β yang memaksimalkan fungsi $L(\hat{\beta})$. Jika $G(\cdot)$ adalah fungsi densitas kumulatif logit standar maka adalah estimator logit. Jika $G(\cdot)$ adalah fungsi densitas kumulatif normal standar maka adalah estimator probit. MLE ini bersifat konsisten, asimtotis normal dan asimtotis efisien (Wooldridge, 1999).

Uji hipotesis dan interval konfidensi untuk parameter (*slope*) adalah (Koppelman, dkk., 2006):

a. Uji untuk masing-masing *slope*

$H_0 : \beta_j = \beta_j(0)$ didasarkan pada statistik Wald:

$$Z_0 = \frac{B_j - \beta_j^0}{\hat{SE}(B_j)} \quad (2.17)$$

b. Uji untuk beberapa slope

$H_0 : \beta_j = \dots = \beta_q = 0$ didasarkan pada statistik

$$X^2 = X^2_{\text{model 1}} - X^2_{\text{model 2}} \quad (2.18)$$

yang berdistribusi *Chi* Kuadrat dengan derajat bebasnya sama dengan selisih banyaknya parameter dari kedua model. X^2 adalah *deviance* yang mempunyai nilai $-2\log L$

Untuk menguji kecocokan model dapat digunakan statistik Pseudo R^2 yang identik dengan nilai R^2 (koefisien deterministik). (McFadden (1974))

$$\text{pseudo } R^2 = 1 - \frac{X_1^2}{X_0^2} \quad (2.19)$$

Jika model secara sempurna memprediksi nilai Y ($P_i = 1$ maka $y_i = 1$ dan jika $P_i = 0$ maka $y_i = 0$) maka $\log L = 0$ (atau nilai *devianceny*a nol). Sehingga nilai maksimum dari pseudo R^2 adalah satu. Pemasalahan dalam penggunaan pseudo R^2 ini adalah tidak adanya kaidah untuk menyatakan pada nilai berapa sedemikian hingga model dikatakan baik. Permasalahan kedua adalah peningkatan nilai pseudo R^2 pada penambahan variabel independen tidak dapat menjelaskan seberapa penting variabel tersebut (Koppelman, dkk., 2006). Karena *Goodness-of-fit* bukan ukuran yang penting secara statistik, sehingga lebih baik menggunakan nilai signifikan variabel *explanatory* (Maddala, 1983).

2.3 Interpretasi estimator

Estimasi koefisien, standar *error* dan nilai fungsi Log-Likelihood dari model Logit dan Probit dapat dilihat dalam *output* beberapa *software* statistik seperti Minitab, SPSS dan Splus. Koefisien memberikan tanda efek parsial untuk masing-masing x_j pada respons probabilitas dan signifikansi x_i dihitung pada tingkat signifikansi yang ditentukan. Ukuran kecocokan model biasanya dilaporkan dengan istilah *percent correctly predicted*, yang dihitung sebagai berikut. Untuk setiap i , kita menghitung estimasi probabilitas pada y_i yang mempunyai nilai satu, $G(\hat{\beta}_0 + x_i \hat{\beta})$. Jika $G(\hat{\beta}_0 + x_i \hat{\beta}) >$

0.5 , maka prediksi y_i adalah satu dan jika $G(\hat{\beta}_0 + x_i \hat{\beta}) \leq 0.5$, prediksi y_i adalah null.

Persentase dihitung dari banyaknya prediksi y_i yang sesuai dengan nilai aktual y_i . Ukuran ini sedikit bermanfaat, tetapi mungkin akan mempunyai nilai persentase

yang besar walaupun tanpa memperhatikan modelnya. Misalnya dari sampel berukuran 200, 180 observasi mempunyai nilai $y_i=0$ dan dari aturan prediksi diperoleh 150 nilai prediksi y adalah null. Walaupun tidak satupun prediksi y pada nilai satu yang tepat, kita masih memprediksikan 70% prediksi adalah benar. Oleh karena itu lebih baik menghitung presentasi pada nilai $y=0$ dan $y=1$.

Seringkali kita ingin mengestimasi efek x_j terhadap respons probabilitas, $P(y=1|x)$. Jika x_j adalah kontinu, maka

$\Delta \hat{P}(y=1|x) \approx [g(\hat{\beta}_0 + x\hat{\beta})\hat{\beta}_j] \Delta x_j$ untuk perubahan x_j yang kecil. Karena $g(\hat{\beta}_0 + x\hat{\beta})$ tergantung pada x , kita harus menghitungnya pada nilai x tertentu.

Seringkali rata-rata sampel dari x_j dipakai untuk mendapatkan nilai $g(\hat{\beta}_0 + x\hat{\beta})$.

Kadang dipakai nilai maksimum dan minimum untuk mendapatkan nilai $g(\hat{\beta}_0 + x\hat{\beta})$.

Sebagaimana yang disebutkan diawal, pada model *Probit*, $g(0) \approx 0,4$ dan untuk model *Logit*, $g(0) = 0,25$. Sehingga estimator koefisien dari model *Logit* sama dengan 1,6 (atau 0,4/0,25) kali koefisien model *Probit*, atau koefisien model *Probit* sama dengan 0,625 kali koefisien model *Logit*. Pada model Linear, $g(0) = 1$, sehingga

koefisien model Linear = 0,25 kali koefisien model *Logit* dan

koefisien model *Probit* = 2,5 kali koefisien model Linear (2.20)

Perbandingan yang lebih akurat adalah dengan mengalikan koefisien probit dengan

$$\phi(\hat{\beta}_0 + \bar{x}\hat{\beta}) \text{ dan mengalikan koefisien Logit dengan } \frac{\exp(\hat{\beta}_0 + \bar{x}\hat{\beta})}{[1 + \exp(\hat{\beta}_0 + \bar{x}\hat{\beta})]^2}.$$

Jika x_k adalah variabel biner, maka lebih mudah menggunakan nilai null atau satu. Jika x_k sebagai variabel diskrit, maka kita dapat mengestimasi perubahan probabilitas pada perubahan x_k dari c_k ke (c_k+1) dengan menggunakan

$$\text{efek } x_k = G(\hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \dots + \hat{\beta}_k (c_k+1)) - G(\hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \dots + \hat{\beta}_k c_k) \quad (2.21)$$

Khususnya jika x_k adalah variabel biner, maka $c_k = 0$.

2.4 Contoh kasus

Mroz (1987), yang telah melakukan survei mengenai partisipasi ibu rumah tangga dalam kerja. Dari 753 wanita dalam sampel terdapat 428 orang bekerja. Variabel y menyatakan apakah wanita yang disurvei bekerja ($y=1$) atau tidak bekerja ($y=0$). Diasumsikan variabel y dipengaruhi oleh *income* suami yang diukur dalam ribuan dolar (x_1), pendidikan dalam tahun (x_2), pengalaman kerja (x_3), usia (x_4), jumlah anak usia kurang dari 6 tahun (x_5), jumlah anak usia antara 6 th s/d 18 th (x_6). Data dapat diakses dalam *library Ecdat* di program R. Estimasi parameter dari model Linier, model *Probit* dan model logistik dengan menggunakan program R ditampilkan dalam tabel 2.1. Pada model Linear digunakan metode *Ordinary Least Square* (OLS) sedangkan pada model *Probit* dan *Logit* digunakan metode MLE.

Tabel 2.1. Estimasi parameter dari model linear, model *logit* dan model *probit*

| Variabel | Linear (OLS) | Logit (MLE) | Probit (MLE) |
|------------------|--------------------------|-------------------------|--------------------------|
| x_1 | 0.0171327 (6.17e-07) | 0.104271 (1.80e-06) | 0.0519502 (2.69e-05) |
| x_2 | 0.0187313 (0.0119) | 0.112243 (0.00887) | 0.0721611 (0.00428) |
| x_3 | 0.0379241 (2.73e-11) | 0.202105 (7.42e-10) | 0.1198423 (3.23e-10) |
| $(x_3)^2$ | -0.0005128 (0.0052) | -0.002760 (0.00828) | -0.0016026 (0.00892) |
| x_4 | -0.0172213 (2.12e-12) | -0.095847 (4.32e-11) | -0.0572580 (8.87e-12) |
| x_5 | -0.2543358 (4.97e-14) | -1.389406 (1.01e-11) | -0.8307527 (1.51e-12) |
| x_6 | 0.0130446 (0.3165)* | 0.055377 (0.46208)* | 0.0344627 (0.43490)* |
| <i>intersept</i> | 1.6418709 (2e-16) | 0.680194 (0.42888)* | 0.4382283 (0.38645)* |

| Variabel | Linear (OLS) | Logit (MLE) | Probit (MLE) |
|-----------------------|--------------------|---------------------|---------------------|
| Null deviance | 184.73 (df=752) | 1029.75 (df=752) | 1029.75 (df=752) |
| Residual deviance | 132.43 (df=745) | 784.75 (df=745) | 786.33 (df=745) |
| Log-likelihood | - | -392.3727 (df=8) | -393.1644 (df=8) |
| pseudo R ² | 0.2831 | 0.23797 | 0.236387 |

keterangan:

$$D_0 = -2\log L_0 = \text{Null deviance}$$

$$D_1 = -2\log L_1 = \text{Residual deviance}$$

$$\text{pseudo } R^2 = 1 - (D_1/D_0)$$

nilai dalam tanda kurung menyatakan tingkat signifikansi (galat tipe I)

*) : tidak signifikan pada $\alpha=0,05$

Ketiga model menghasilkan kesimpulan yang sama. Variabel x1, x2, x3 mempunyai koefisien negatif dan signifikan. Variabel (x3)², x4 dan x5 mempunyai koefisien negatif. Variabel x6 pada ketiga model tidak signifikan dan intersep pada model *Probit* maupun *Logit* juga tidak signifikan.

Sebagaimana yang telah dijelaskan sebelumnya, nilai koefisien dari ketiga model tidak bisa dibandingkan secara langsung, harus dilakukan penyetaraan (lihat persamaan 2.15).

Tabel 2.2. Estimasi parameter setelah distandardisasi

| Variabel | Linear (OLS) | Logit (MLE) | Probit (MLE) |
|----------|--------------|-------------|--------------|
| x1 | 0.0171327 | 0.026068 | 0.02078 |
| x2 | 0.0187313 | 0.028061 | 0.028864 |

| Variabel | Linear (OLS) | Logit (MLE) | Probit (MLE) |
|-------------------|--------------|-------------|--------------|
| x3 | 0.0379241 | 0.050526 | 0.047937 |
| (x3) ² | -0.0005128 | -0.00069 | -0.00064 |
| x4 | -0.0172213 | -0.02396 | -0.0229 |
| x5 | -0.2543358 | -0.34735 | -0.3323 |
| x6 | 0.0130446 | 0.013844 | 0.013785 |
| intersept | 1.6418709 | 0.170049 | 0.175291 |

Jika kita mengevaluasi fungsi densitas normal standar,

$$\phi(\hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \dots + \hat{\beta}_k x_k)$$

mempunyai rata-rata $0.2977221 \approx 0,3$. sesuai dengan persamaan (2.8), berarti untuk mengestimasi perubahan respons probabilitas pada perubahan satu satuan suatu variabel independen tertentu, kita mengalikan koefisien variabel yang bersangkutan dengan 0,3. sedangkan pada model *Logit*, kita mengalikannya dengan 0,1875.

Pada kasus ini, perbedaan terbesar antara model Linear dengan model *Logit* dan model *Probit* adalah model Linear mengandung suku konstanta. Sehingga efek parsialnyapun juga memuat suku konstanta. Misalkan kita ingin melihat pengaruh kepemilikan anak balita (x5) terhadap y, atau menghitung efek parsial dari variabel $x_5=0$ ke $x_5=1$ pada variabel yang lain tertentu. Misal variabel yang lain diambil nilai rata-ratanya, yaitu $x_1=9.331913$, $x_2=12.28685$, $x_3=10.63081$, $x_4=42.53785$ dan $x_6=1$. Karena x_5 berupa variabel diskrit maka efek parsial dapat menggunakan persamaan (2.10).

Tabel 2.3. Estimasi efek parsial variabel x5

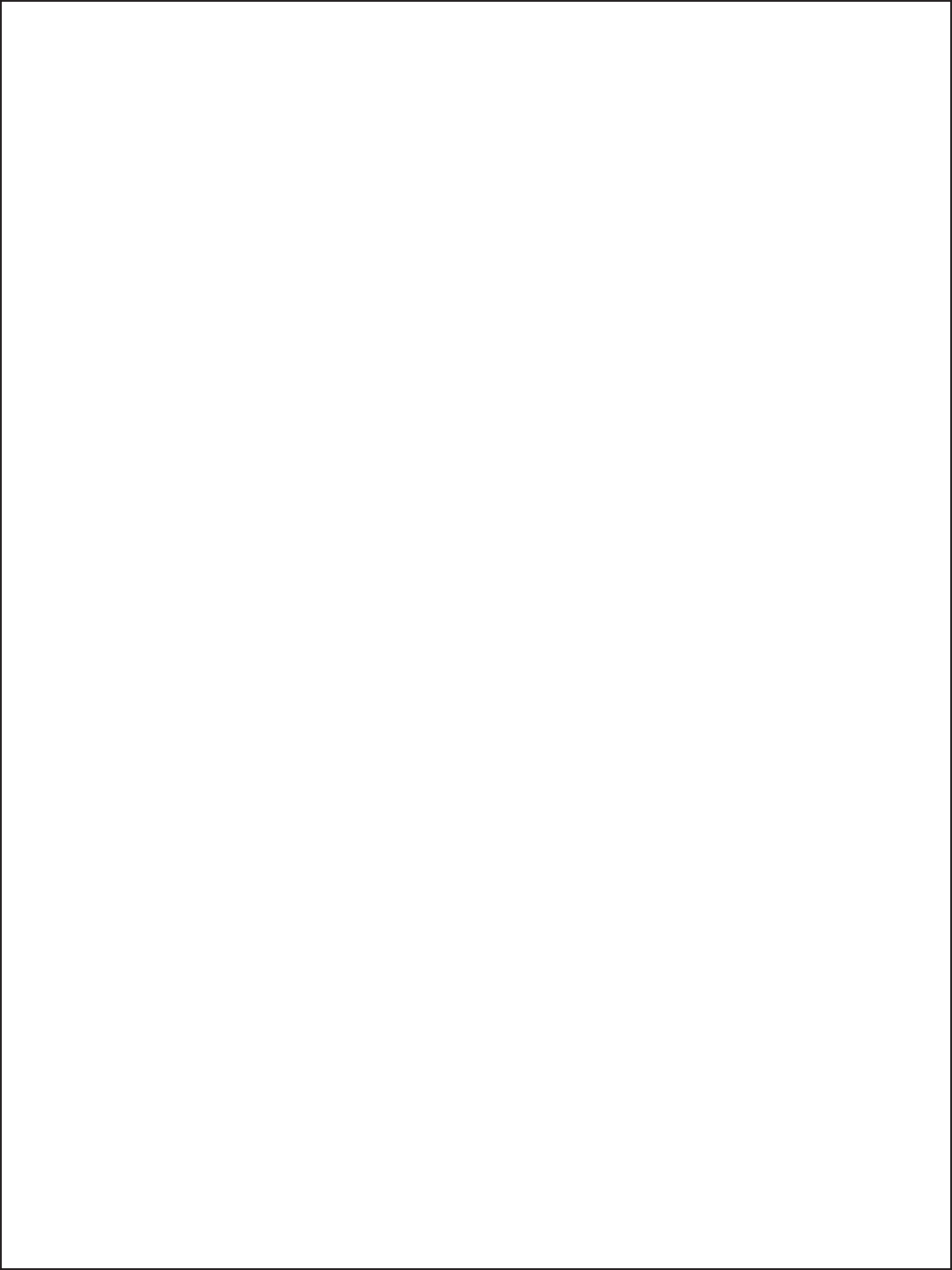
| Probabilitas | Model Linear | Logit | Probit |
|-------------------------------|--------------|-----------|------------|
| $P(x_5=0 x_t)$ | 1.657598 | 0.5416615 | 0.5251813 |
| $P(x_5=1 x_t)$ | 1.403262 | 0.2275192 | 0.2213652 |
| $P(x_5=1 x_t) - P(x_5=0 x_t)$ | -0.25434 | -0.314142 | -0.3038161 |
| $P(x_5=2 x_t)$ | 1.148927 | 0.0683842 | 0.0549833 |
| $P(x_5=2 x_t) - P(x_5=0 x_t)$ | -0.50867 | -0.473277 | -0.470198 |

Estimasi probabilitas pada model linear menghasilkan nilai $p > 1$ sehingga dapat disimpulkan bahwa model linear tidak cocok digunakan untuk respons terbatas. Estimasi



probabilitas pada model probit hasilnya tidak jauh berbeda dengan model *Logit*. Efek parsial, yaitu kepemilikan anak usia kurang dari 6 tahun, dari tidak mempunyai anak ke mempunyai satu anak akan mengurangi probabilitas wanita bekerja sekitar 0,3. Demikian juga untuk perubahan dari tidak mempunyai anak menjadi dua anak akan mengurangi probabilitas sebesar 0,47.

Model *Probit* dan *Logit* merupakan anggota keluarga dari *Generalized Linear Model* yang secara luas telah digunakan untuk pemodelan dengan variabel dependen berupa data biner. Model *Probit* didasarkan pada distribusi normal, sedangkan model *Logit* didasarkan pada distribusi Logistik. Estimator koefisien dari model *Logit* sama dengan 1,6 kali koefisien model *Probit*, atau koefisien model probit sama dengan 0,625 kali koefisien model *Logit*. Koefisien model linear setara dengan 0,25 kali koefisien model *Logit*. Dengan metode MLE dan berdasarkan teori sampel besar, diperoleh uji parameter dengan statistik Walds. Efek parsial dari sebuah variabel explanatori dapat dihitung dari selisih nilai distribusi kumulatifnya.



Bab 3 Model Probit

3.1 Pendahuluan

Probabilitas pembuat keputusan i memilih alternatif k dapat dinyatakan sebagai:

$$\begin{aligned}
 p_{ik} &= P(U_{ik} > U_{ij}) \quad \forall j \neq k \\
 &= P(\varepsilon_{ij} - \varepsilon_{ik} < V_{ik} - V_{ij}) \quad \forall j \neq k \\
 &= \int_{\varepsilon} I(\varepsilon_j - \varepsilon_k < V_k - V_j) f(\varepsilon_i) d\varepsilon_i \quad \forall j \neq k
 \end{aligned} \tag{3.1}$$

$I(\cdot)$ adalah fungsi indikator, yang bernilai 1 jika pernyataan dalam kurung benar dan bernilai 0 jika pernyataan salah. Selanjutnya dapat dipilih atau ditentukan densitas $f(\cdot)$ yang sesuai/tepat.

Beberapa model yang dapat digunakan dalam pemodelan pilihan diskrit (*Discrete Choice Models*) antara lain model *Multinomial Logistik* (MNL), model *Multinomial Probit* (MNP) dan model *Generalized Extreme Value* (GEV). Model MNL mengasumsikan bahwa komponen *error*-nya berdistribusi *extreme value* tipe I dan saling independen. Model MNP mengasumsikan bahwa komponen *error*-nya berdistribusi *multivariate* normal. Model MNL dan Model GEV mempunyai persamaan probabilitas dalam bentuk persamaan tertutup, tetapi mempunyai keterbatasan dalam mengakomodasi adanya variasi individu (Train, 2003). Model *Probit* secara analitik sangat menarik, tetapi sangat sulit melakukan estimasi parameternya (Davidson dan Russel, 1999). Kesulitan estimasi parameter karena memerlukan integral rangkap. Perhitungan integral dapat didekati dengan menggunakan simulasi (Hajivassiliou dan Ruud, 1994).

3.2 Model multinomial probit (MNP)

Pada model MNP, diasumsikan bahwa vektor $\varepsilon_i^t = (\varepsilon_{i1}, \dots, \varepsilon_{ij})$ berdistribusi multivariat normal dengan *mean* nol dan matrik kovariansi Σ . Densitas untuk ε_i adalah

$$\phi(\varepsilon_i) = \frac{1}{(2\pi)^{J/2} |\Sigma|^{1/2}} \exp\left[-\frac{1}{2} \varepsilon_i^t \Sigma^{-1} \varepsilon_i\right] \tag{3.2}$$

Probabilitas pembuat keputusan i memilih alternatif k adalah:

$$\begin{aligned}
p_{ik} &= P(V_{ik} + \varepsilon_{ik} > V_{ij} + \varepsilon_{ij}) \quad \forall j \neq k \\
&= \int I(V_{ik} + \varepsilon_{ik} > V_{ij} + \varepsilon_{ij}) \phi(\varepsilon_i) d\varepsilon_i \quad \forall j \neq k
\end{aligned} \tag{3.3}$$

dengan $I(\cdot)$ merupakan fungsi indikator dan integral terhadap semua nilai ε_i .

Probabilitas pilihan dapat dinyatakan sebagai:

$$p_k = \int_{\varepsilon_i \in B_k} \phi(\varepsilon_i) d\varepsilon_i \tag{3.4}$$

Dengan B_{ik} adalah himpunan *error* ε_i yang dihasilkan oleh pembuat keputusan karena memilih alternatif i .

$$B_{ik} = \{\varepsilon_i \mid V_{ik} + \varepsilon_{ik} > V_{ij} + \varepsilon_{ij}\} \quad \forall j \neq k$$

Penyajian probabilitas pada persamaan (3.5) merupakan integral berdimensi J atas *error* ε_{ij} , $j = 1, 2, \dots, J$. Karena hanya berbeda dalam bentuk utiliti, maka probabilitas pilihan dapat dinyatakan sebagai integral berdimensi $J-1$ atas selisih di antara *error*-nya. Misal diambil selisih terhadap alternatif k , maka dapat didefinisikan:

$$\tilde{U}_{ijk} = U_j - U_k,$$

$$\tilde{V}_{ijk} = V_j - V_k$$

$$\tilde{\varepsilon}_{ijk} = \varepsilon_j - \varepsilon_k.$$

Persamaan utilitinya menjadi:

$$\tilde{U}_{ijk} = \tilde{V}_{ijk} + \tilde{\varepsilon}_{ijk} \text{ untuk } j \neq k; \tilde{\varepsilon}_k^t = (\tilde{\varepsilon}_{i1}, \dots, \tilde{\varepsilon}_{ij}) \tag{3.5}$$

di mana tanda "... " adalah simbol semua kecuali k , sehingga $\tilde{\varepsilon}_k^t$ adalah matrik $(J-1) \times 1$ dan $\tilde{\varepsilon}_k^t \sim N(0, \tilde{\Sigma}_k)$. Matrik kovariansi $\tilde{\Sigma}_k$ dapat diturunkan dari Σ dan dapat dicari faktor Choleski, L_k sedemikian hingga $L_k L_k^t = \tilde{\Sigma}_k$.

$$L_k = \begin{pmatrix} c_1 & 0 & 0 & 0 & 0 & \dots & 0 \\ c_2 & c_2 & 0 & 0 & 0 & \dots & 0 \\ c_3 & c_3 & c_3 & 0 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ c_{J1} & c_{J2} & c_{J3} & \dots & \dots & c_{(J-1)1} & 0 \end{pmatrix}$$

di mana Persamaan utiliti dapat disusun sebagai:

$$\tilde{U}_{\dot{k}} = \tilde{V}_{\dot{k}} + L_k \eta_i$$

dengan $\eta_i^t = (\eta_{1i}, \dots, \eta_{(j-1)i})$ adalah vektor yang berdistribusi normal *standard* independen, $\eta_{ji} \sim N(0, 1) \quad \forall j$, $\tilde{U}_{\dot{k}}^t = (\tilde{U}_{i1k}, \dots, \tilde{U}_{iJk})$ dan $\tilde{V}_{\dot{k}}^t = (\tilde{V}_{i1k}, \dots, \tilde{V}_{iJk})$.

Model dapat dituliskan secara eksplisit:

$$\tilde{U}_{i1k} = \tilde{V}_{i1k} + c_1 \eta_{1i},$$

$$\tilde{U}_{i2k} = \tilde{V}_{i2k} + c_2 \eta_{1i} + c_2 \eta_{2i}$$

$$\tilde{U}_{i3k} = \tilde{V}_{i3k} + c_3 \eta_{1i} + c_3 \eta_{2i} + c_3 \eta_{3i}$$

.....

$$\tilde{U}_{iJk} = \tilde{V}_{iJk} + c_{J1} \eta_{1i} + c_{J2} \eta_{2i} + c_{J3} \eta_{3i} + \dots + c_J \eta_{Ji}$$

Probabilitas individu i memilih alternatif k adalah

$$\begin{aligned} p_{\dot{k}} &= P(\tilde{U}_{ijk} < 0) \quad \forall j \neq k \\ &= P(\tilde{U}_{i1k} < 0, \tilde{U}_{i2k} < 0, \dots, \tilde{U}_{iJk} < 0) \\ &= P\left(\eta_1 < \frac{-\tilde{V}_{i1k}}{c_1}\right) \cdot P\left(\eta_2 < \frac{-(\tilde{V}_{i2k} + c_2 \eta_{1i})}{c_2} \mid \eta_1 < \frac{-\tilde{V}_{i1k}}{c_1}\right) \cdot \\ &\quad P\left(\eta_3 < \frac{-(\tilde{V}_{i3k} + c_3 \eta_{1i} + c_3 \eta_{2i})}{c_3} \mid \eta_1 < \frac{-\tilde{V}_{i1k}}{c_1} \text{ dan } \eta_2 < \frac{-(\tilde{V}_{i2k} + c_2 \eta_{1i})}{c_2}\right) \dots \end{aligned}$$

$$P \left(\eta_{1i} < \frac{-(\tilde{V}_{ijk} + c_{J1}\eta_{1i} + \dots + c_{J(J-1)}\eta_{(J-1)i})}{c_3} \mid \eta_{1i} < \frac{-\tilde{V}_{ilk}}{c_1}, \dots, \dots, \eta_{(J-1)i} < \frac{-(\tilde{V}_{i(J-1)k} + c_{(J-1)1}\eta_{1i} + \dots + c_{(J-1)(J-2)}\eta_{(J-2)i})}{c_{(J-1)(J-1)}} \right) \quad (3.6)$$

Didefinisikan vektor $\tilde{\varepsilon}_k = (\tilde{\varepsilon}_{1k}, \dots, \tilde{\varepsilon}_{Jk})$ di mana “...” adalah notasi untuk semua alternatif kecuali k, sehingga $\tilde{\varepsilon}_k$ berdimensi J-1. Karena selisih dua distribusi normal adalah normal, maka densitas selisih error tersebut adalah:

$$\phi(\tilde{\varepsilon}_k) = \frac{1}{(2\pi)^{(J-1)/2} |\tilde{\Sigma}_k|^{1/2}} \exp\left[-\frac{1}{2} \tilde{\varepsilon}_k^t \tilde{\Sigma}_k^{-1} \tilde{\varepsilon}_k\right] \quad (3.7)$$

Di mana $\tilde{\Sigma}_k$ adalah *covarians* dari $\tilde{\varepsilon}_k$, yang diturunkan dari Σ . Selanjutnya, probabilitas pilihan dapat disajikan dalam selisih utiliti:

$$p_k = \int I(\tilde{V}_{ijk} + \tilde{\varepsilon}_{ijk} < 0) \phi(\tilde{\varepsilon}_k) d\tilde{\varepsilon}_k \quad \forall j \neq k \quad (3.8)$$

Yang merupakan integral berdimensi (J-1) atas semua nilai yang mungkin dari selisih *error*. Penyajian tersebut ekuivalen dengan:

$$p_k = \int_{\tilde{\varepsilon}_k \in \tilde{B}_k} \phi(\tilde{\varepsilon}_k) d\tilde{\varepsilon}_k \quad (3.9)$$

dengan $\tilde{B}_k = \{\tilde{\varepsilon}_k \mid \tilde{V}_{ijk} + \tilde{\varepsilon}_{ijk} < 0 \quad \forall j \neq k\}$ yang merupakan integral berdimensi (J-1) atas nilai selisih dalam \tilde{B}_k .

Untuk menghitung p_{ik} , memerlukan matrik kovarians $\tilde{\Sigma}_k$ dari selisih *error*. Matrik $\tilde{\Sigma}_k$

dapat diturunkan secara langsung dari Σ . Misal M_k adalah matrik identitas berdimensi (J-1) dan menambahkan kolom ke-i yang bernilai “-1”, sehingga M_k berdimensi (J-1)xJ. Misal J= 4 alternatif dan k=3.

$$M_k = \begin{pmatrix} 1 & 0 & -1 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & -1 & 1 \end{pmatrix}$$

Matrik ini dapat digunakan untuk mentransformasi Σ ke dalam $\tilde{\Sigma}_k$, yaitu:

$$\tilde{\Sigma}_k = M_k \Sigma M_k^t.$$

Matrik $\tilde{\Sigma}_k$ berdimensi (J-1)x(J-1), sementara Σ berdimensi JxJ. Misal terdapat tiga alternatif dengan *error* ($\epsilon_{i1}, \epsilon_{i2}, \epsilon_{i3}$) dengan *covarians*

$$\Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{12} & \sigma_{22} & \sigma_{23} \\ \sigma_{13} & \sigma_{23} & \sigma_{33} \end{pmatrix}$$

Andaikan diambil selisih terhadap alternatif 2. Nilai selisih *error* adalah $(\tilde{\epsilon}_{n12}, \tilde{\epsilon}_{n32})$

yang mempunyai *covarians*

$$\tilde{\Sigma}_2 = Cov \begin{pmatrix} \epsilon_{i1} - \epsilon_{i2} \\ \epsilon_{i3} - \epsilon_{i2} \end{pmatrix} = \begin{pmatrix} \sigma_{11} + \sigma_{22} - 2\sigma_{12} & \sigma_{13} + \sigma_{22} - \sigma_{12} - \sigma_{23} \\ \sigma_{13} + \sigma_{22} - \sigma_{12} - \sigma_{23} & \sigma_{33} + \sigma_{22} - 2\sigma_{23} \end{pmatrix}$$

Matrik kovarians ini dapat juga diturunkan dengan transformasi

$$\tilde{\Sigma}_2 = M_2 \Sigma M_2^t,$$

$$\begin{aligned} \tilde{\Sigma}_2 &= \begin{pmatrix} 1 & -1 & 0 \\ 0 & -1 & 1 \end{pmatrix} \begin{pmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{12} & \sigma_{22} & \sigma_{23} \\ \sigma_{13} & \sigma_{23} & \sigma_{33} \end{pmatrix} \begin{pmatrix} 1 & 0 \\ -1 & -1 \\ 0 & 1 \end{pmatrix} \\ &= \begin{pmatrix} \sigma_{11} - \sigma_{12} & \sigma_{12} - \sigma_{22} & \sigma_{13} - \sigma_{23} \\ \sigma_{12} - \sigma_{13} & \sigma_{22} - \sigma_{23} & \sigma_{23} - \sigma_{33} \end{pmatrix} \begin{pmatrix} 1 & 0 \\ -1 & -1 \\ 0 & 1 \end{pmatrix} \\ &= \begin{pmatrix} \sigma_{11} + \sigma_{22} - 2\sigma_{12} & \sigma_{13} + \sigma_{22} - \sigma_{12} - \sigma_{23} \\ \sigma_{13} + \sigma_{22} - \sigma_{12} - \sigma_{23} & \sigma_{33} + \sigma_{22} - 2\sigma_{23} \end{pmatrix} = \begin{pmatrix} \theta_1 & \theta_2 \\ \theta_2 & \theta_3 \end{pmatrix} \end{aligned}$$

Sebagaimana yang kita lihat, transformasi dengan M_i akan dipakai dalam mensimulasi probabilitas probit.

3.3 Identifikasi parameter

Penambahan konstanta pada setiap variabel utilitas untuk semua alternatif tidak merubah peringkat utiliti dari semua alternatif. Demikian juga mengalikan dengan bilangan positif terhadap utiliti pada semua alternatif tidak merubah peringkat utilitinya. Oleh karena itu dalam model probit normalisasi skala dan level perlu dilakukan.

Normalisasi model berhubungan dengan identifikasi parameter. Sebuah parameter teridentifikasi (*identified*) jika dapat diestimasi dan tidak teridentifikasi (*unidentified*) jika tidak terestimasi. Sebagai contoh parameter-*unidentified* adalah k dalam utiliti

$$U_{nj} = V_{nj} + k + \varepsilon_{nj}.$$

Karena k merupakan ukuran keseluruhan level utiliti (*overall level of utility*) sehingga peneliti tidak dapat melakukan inferensi k . Parameter yang tidak mempengaruhi probabilitas pilihan (perilaku keputusan) tidak dapat diestimasi. Dalam model yang tidak dinormalisasi, parameter-parameter yang nampak tidak dapat diidentifikasi, parameter tersebut berhubungan dengan *skala* dan *level* utiliti yang tidak berefek pada perilaku keputusan. Sedangkan pada model yang dinormalisasi, parameter tersebut tidak nampak. Kesulitan muncul karena tidak selalu jelas berhubungan dengan skala dan *level*. Dalam contoh di awal, bahwa parameter k adalah berhubungan dengan *level*. Bunch dan Kitamura (1989) telah menunjukkan aplikasi model *probit* dalam beberapa artikel yang dipublikasikan tidak dinormalisasi dan memuat parameter *unidentified*.

Berikut ini akan dibahas sebuah prosedur yang dapat selalu digunakan untuk menormalisasi model *probit* dan menjamin semua parameternya teridentifikasi. Akan dijelaskan prosedur dalam contoh model empat alternatif dan dapat digeneralisasi dengan mudah. Dalam pembahasan sebelumnya, utilitas disajikan sebagai:

$$U_{ij} = V_{ij} + \varepsilon_{ij}, j = 1, \dots, 4.$$

Vektor *errors*-nya adalah $\varepsilon_i = (\varepsilon_{i1}, \dots, \varepsilon_{i4})$ yang berdistribusi normal dengan *mean* nol dan matrik kovariansi

$$\Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} & \sigma_{14} \\ \cdot & \sigma_{22} & \sigma_{23} & \sigma_{24} \\ \cdot & \cdot & \sigma_{33} & \sigma_{34} \\ \cdot & \cdot & \cdot & \sigma_{44} \end{pmatrix}$$

Model untuk empat alternatif akan terdapat 10 nilai *varians-covarians*. Secara umum model dengan J alternatif akan mempunyai $J(J + 1)/2$ nilai *varian-covarians* yang berbeda dalam matrik *covarian*-nya. Dalam prosedur ini selalu mengambil nilai selisihnya terhadap alternatif pertama, karena akan lebih menyederhanakan analisis. Didefinisikan selisih *error* sebagai:

$$\tilde{\varepsilon}_{ijk} = \varepsilon_{ij} - \varepsilon_{i1} \text{ untuk } j = 2, 3, 4,$$

dan didefinisikan vektor selisih *error* sebagai $\tilde{\varepsilon}_{i1} = (\tilde{\varepsilon}_{i21}, \tilde{\varepsilon}_{i31}, \tilde{\varepsilon}_{i41})$. Indeks 1 dalam $\tilde{\varepsilon}_{i1}$ berarti selisih *error* terhadap alternatif pertama, bukan *error* untuk alternatif pertama.

Matrik *covarian* untuk vektor selisih *error* adalah

$$\tilde{\Sigma}_1 = \begin{pmatrix} \theta_{22} & \theta_{23} & \theta_{24} \\ \cdot & \theta_{33} & \theta_{34} \\ \cdot & \cdot & \theta_{44} \end{pmatrix}$$

Di mana θ berhubungan dengan σ yaitu:

$$\theta_{22} = \sigma_{22} + \sigma_{11} - 2\sigma_{12},$$

$$\theta_{23} = \sigma_{23} + \sigma_{11} - \sigma_{12} - \sigma_{13}$$

$$\theta_{24} = \sigma_{24} + \sigma_{11} - \sigma_{12} - \sigma_{14}$$

$$\theta_{33} = \sigma_{33} + \sigma_{11} - 2\sigma_{13},$$

$$\theta_{34} = \sigma_{34} + \sigma_{11} - \sigma_{13} - \sigma_{14}$$

$$\theta_{44} = \sigma_{44} + \sigma_{11} - 2\sigma_{14},$$

Dengan komputasi, matrik ini dapat diperoleh menggunakan tranformasi M_1 yang didefinisikan di atas yaitu

$$\tilde{\Sigma}_1 = M_1 \Sigma M_1^t$$

Selanjutnya normalisasi dengan skala *utility*, yaitu Variansi dari selisih *error* $\hat{\varepsilon}_{i2}$ ditransformasi menjadi satu, sehingga matrik kovariannya menjadi:

$$\tilde{\Sigma}_1^* = \begin{pmatrix} 1 & \theta_{23}^* & \theta_{24}^* \\ \cdot & \theta_{33}^* & \theta_{34}^* \\ \cdot & \cdot & \theta_{44}^* \end{pmatrix}$$

di mana:

$$\theta_{33}^* = \frac{\sigma_{33} + \sigma_{11} - 2\sigma_{13}}{\sigma_{22} + \sigma_{11} - 2\sigma_{12}}$$

$$\theta_{44}^* = \frac{\sigma_{44} + \sigma_{11} - 2\sigma_{14}}{\sigma_{22} + \sigma_{11} - 2\sigma_{12}}$$

$$\theta_{23}^* = \frac{\sigma_{23} + \sigma_{11} - \sigma_{12} - \sigma_{13}}{\sigma_{22} + \sigma_{11} - 2\sigma_{12}}$$

$$\theta_{24}^* = \frac{\sigma_{24} + \sigma_{11} - \sigma_{12} - \sigma_{14}}{\sigma_{22} + \sigma_{11} - 2\sigma_{12}}$$

$$\theta_{34}^* = \frac{\sigma_{34} + \sigma_{11} - \sigma_{13} - \sigma_{14}}{\sigma_{22} + \sigma_{11} - 2\sigma_{12}}$$

Kelima parameter dalam matrik kovarians ini merupakan parameter teridentifikasi. Jumlah ini jauh lebih kecil dibanding 10 elemen dalam matrik *covarians* sebelumnya. Karena hanya 5 parameter θ^* dan 10 parameter σ , maka tidak mungkin mendapatkan seluruh nilai σ dari nilai estimasi θ^* . Secara umum, sebuah model dengan J alternatif dan dalam matrik *covarians* terdapat $J(J+1)/2$ parameters. Setelah dinormalisasi, parameter dalam matrik kovariansi sebanyak $[(J-1)J/2] - 1$. Sehingga hanya terdapat $[(J-1)J/2] - 1$ parameters yang teridentifikasi.

Andaikan peneliti menetapkan struktur matrik *covarian*. Sehingga matrik *full covarian* untuk *error* mempunyai bentuk khusus (nilai elemennya tertentu), atau hubungan di antara elemen dalam matrik kovarian. Peneliti dapat membatasi bentuk matrik kovarian. Struktur tersebut dapat berbentuk bermacam macam, tergantung pada aplikasinya. Yai, dkk. (1997) mengestimasi model *probit* dari rute pilihan di mana kovariansi antara dua rute hanya tergantung pada jarak nya, struktur ini mereduksi jumlah parameter kovarian menjadi satu, yaitu kovariansi jarak. Bolduc, dkk. (1996) mengestimasi model pilihan fisik dari lokasi di mana kovariansi antar lokasi adalah sebuah fungsi kedekatannya dengan yang lain. Bolduc (1992) memberi nama struktur



"generalized autoregressive". Haaijer, dkk. (1998) memilih struktur *factor-analytic* yang muncul dari koefisien *random* variabel penjelas. Seringkali struktur yang ditetapkan akan cukup untuk menormalisasi model. Bunch dan Kitamura (1989) memberikan kasus di mana peneliti menetapkan struktur kovarian yang nampak telah dinormalisasi tetapi sesungguhnya belum dinormalisasi.

Prosedur yang telah disampaikan di awal akan digunakan pada pembatasan matrik kovarian yang cukup untuk menormalisasi model. Peneliti menetapkan Σ dengan elemen-elemen yang terbatas. Kemudian menggunakan prosedur untuk mendapatkan $\tilde{\Sigma}_1^*$ yang dinormalisasi untuk skala dan *level*. Kita tahu bahwa masing-masing elemen $\tilde{\Sigma}_1^*$ adalah teridentifikasi. Jika pembatasan elemen pada Σ dapat dihitung dari elemen elemen pada $\tilde{\Sigma}_1^*$ maka pembatasan tersebut cukup untuk menormalisasi. Dalam kasus ini, setiap parameter dalam Σ yang terbatas adalah teridentifikasi. Dengan lain kata jika elemen Σ tidak dapat dihitung dari elemen $\tilde{\Sigma}_1^*$ maka pembatasan tersebut tidak cukup untuk menormalisasi model dan parameter dalam Σ adalah tidak teridentifikasi. Sebagai gambaran, misal peneliti mengestimasi model empat alternatif dan diasumsikan bahwa matrik kovariansi *error* mempunyai bentuk:

$$\Sigma = \begin{pmatrix} 1+\rho & \rho & 0 & 0 \\ . & 1+\rho & 0 & 0 \\ . & . & 1+\rho & \rho \\ . & . & . & 1+\rho \end{pmatrix}$$

Matrik kovariansi ini mempunyai struktur bahwa terdapat korelasi antara *error* pertama dan ke dua, ke tiga dan ke empat. Pasangan korelasinya adalah $\rho/(1 + \rho)$. Catatan bahwa elemen diagonalnya $(1 + \rho)$, peneliti menjamin bahwa korelasinya bernilai antara -1 dan 1 untuk suatu nilai ρ . Bagaimana menormalisasi terhadap skala dan *level*? Kita akan gunakan prosedur dimuka. Pertama, diambil selisih terhadap alternatif pertama. Matrik Kovarians dari selisih *error* tersebut adalah

$$\tilde{\Sigma}_1 = \begin{pmatrix} \theta_{22} & \theta_{23} & \theta_{24} \\ & \theta_{33} & \theta_{34} \\ & & \theta_{44} \end{pmatrix}$$

dengan

$$\theta_{22} = 2, \theta_{33} = 2+2\rho, \theta_{44} = 2+2\rho, \theta_{34} = 12+2\rho, \theta_{23} = 1, \theta_{24} = 1.$$

Selanjutnya menormalisasi untuk skala dengan mentransformasi elemen (1,1) menjadi 1, diperoleh matrik *covarians* ternormalisasi

$$\tilde{\Sigma}_1^* = \begin{pmatrix} 1 & \theta_{23}^* & \theta_{24}^* \\ \dots & \theta_{33}^* & \theta_{34}^* \\ \dots & \dots & \theta_{44}^* \end{pmatrix}$$

dengan

$$\theta_{33}^* = 1 + \rho, \theta_{44}^* = 1 + \rho, \theta_{23}^* = 1/2, \theta_{24}^* = 1/2, \theta_{34}^* = 1/2 + \rho$$

Oleh karena itu

$$\theta_{33}^* = \theta_{44}^* = \theta_{34}^* + 1/2$$

dan parameter θ^* lainnya diberi nilai tertentu. Terdapat satu parameter dalam $\tilde{\Sigma}_1^*$

sebagaimana dalam Σ . Jika didefinisikan $\theta = 1 + \rho$ maka

$$\tilde{\Sigma}_1^* = \begin{pmatrix} 1 & \frac{1}{2} & \frac{1}{2} \\ \dots & \theta & \theta - \frac{1}{2} \\ \dots & \dots & \theta \end{pmatrix}$$

Parameter ρ dapat dihitung secara langsung dari θ . Parameter dalam matrik Σ dapat dihitung dari parameter-parameter dalam matrik $\tilde{\Sigma}_1^*$.

Selanjutnya prosedur normalisasi matrik kovariansi akan dilakukan pada kasus ini. Dengan adanya asumsi korelasi yang berbeda pada *error* pertama dengan *error* kedua dan korelasi *error* ke tiga dengan ke empat, matrik kovariansnya adalah:

$$\Sigma = \begin{pmatrix} 1 + \rho_1 & \rho_1 & 0 & 0 \\ \cdot & 1 + \rho_1 & 0 & 0 \\ \cdot & \cdot & 1 + \rho_2 & \rho_2 \\ \cdot & \cdot & \cdot & 1 + \rho_2 \end{pmatrix}$$

Korelasi antara *error* pertama dan ke dua adalah $\rho_1/(1 + \rho_1)$, dan korelasi antara *error* ke tiga dan ke-empat adalah $\rho_2/(1 + \rho_2)$. Selanjutnya dengan tranformasi skala diperoleh matrik kovariansi

$$\tilde{\Sigma}_1^* = \begin{pmatrix} 1 & \frac{1}{2} & \frac{1}{2} \\ \cdot & \theta & \theta - \frac{1}{2} \\ \cdot & \cdot & \theta \end{pmatrix}$$

Dengan $\theta = 1 + (\rho_1 + \rho_2)/2$. Nilai ρ_1 dan ρ_2 tidak dapat dihitung dari sebuah nilai θ . Model original tidak ternormalisasi untuk skala dan level, dan parameter ρ_1 dan ρ_2 menjadi tidak teridentifikasi.

Dalam model ternormalisasi, yang nampak adalah rata-rata ρ yaitu $(\rho_1 + \rho_2)/2$. Sehingga memungkinkan menghitung rata-rata ρ dari θ . Hal ini berarti bahwa rata-rata ρ adalah teridentifikasi, tetapi bukan nilai individualnya. Ketika $\rho_1 = \rho_2$, sebagaimana dalam contoh di awal, model adalah ternormalisasi karena masing-masing ρ adalah sama dengan rata-rata ρ . Asumsi $\rho_1 = \rho_2$ adalah tidak berbeda dengan asumsi bahwa $\rho_1 = 3\rho_2$, atau hubungan yang lain. Dengan demikian kita tahu bagaimana menjamin model probit ternormalisasi untuk skala dan level. Selanjutnya akan dilihat model probit dalam dalam mengakomodasi adanya variasi individu (*taste variation*)

3.4 Variasi individu

Model *Probit* cukup baik untuk menyusun model dengan koefisien *random*, yaitu koefisien adalah berdistribusi normal. Haaijer et al. (1998) juga telah mengaplikasikan model ini. Asumsi bahwa penyajian utiliti adalah *linear* dalam parameter dan koefisien bervariasi secara *random* atas pembuat keputusan. (sejauh ini diasumsikan *fixed*). Utilitinya adalah $U_{ij} = \beta_i^t x_{ij} + \varepsilon_{ij}$, dengan β_i adalah vektor koefisien untuk pembuat keputusan i yang merepresentasikan variasi individu. Andaikan β_i adalah berdistribusi normal dalam populasi dengan *mean* b dan kovariansi W : $\beta_i \sim N(b, W)$. Permasalahannya adalah bagaimana mengestimasi parameter b dan W .

Utiliti dapat dituliskan dengan β_i yang diuraikan ke dalam mean dan deviasi:

$$U_{ij} = b^t x_{ij} + \tilde{\beta}_i^t x_{ij} + \varepsilon_{ij}, \text{ di mana } \tilde{\beta}_i = \beta_i - b$$

Dua suku terakhir tersebut dalam utiliti adalah *random*, yang dapat dinyatakan dalam η_{ij} sehingga

$$U_{ij} = b^t x_{ij} + \eta_{ij}.$$

Kovariansi η_{ij} tergantung pada W dan x_{ij} , sehingga kovariansi berbeda di antara

pembuat keputusan (individu). Kovariansi η_{ij} dapat dijelaskan dengan mudah untuk model dua alternatif dengan satu variabel independen. Dalam kasus ini, utilitinya adalah

$$U_{i1} = \beta_i x_{i1} + \varepsilon_{i1},$$

$$U_{i2} = \beta_i x_{i2} + \varepsilon_{i2}.$$

Asumsi bahwa β_i adalah berdistribusi normal dengan *mean* b dan variansi σ_β .

Asumsi bahwa ε_{i1} dan ε_{i2} adalah berdistribusi normal identik dengan variansi σ_ε . Dalam contoh ini diasumsikan independen dan secara umum utiliti dapat ditulis sebagai

$$U_{i1} = b x_{i1} + \eta_{i1},$$

$$U_{i2} = b x_{i2} + \eta_{i2},$$

Di mana η_{i1} dan η_{i2} distribusi normal. Masing-masing mempunyai mean nul:

$$E(\eta_{ij}) = E(\tilde{\beta}_i x_{ij} + \varepsilon_{ij}) = 0.$$

Kovariansiya dihitung sebagai berikut. Variansi masing-masing adalah

$$V(\eta_{ij}) = V(\tilde{\beta}_i x_{ij} + \varepsilon_{ij}) = (x_{ij})^2 \sigma_\beta + \sigma_\varepsilon.$$

Kovariansinya adalah

$$\begin{aligned} \text{Cov}(\eta_{i1}, \eta_{i2}) &= E[(\tilde{\beta}_i x_{i1} + \varepsilon_{i1})(\tilde{\beta}_i x_{i2} + \varepsilon_{i2})] \\ &= E[\tilde{\beta}_i^2 x_{i1} x_{i2} + \varepsilon_{i1} \varepsilon_{i2} + \varepsilon_{i1} \tilde{\beta}_i x_{i2} + \varepsilon_{i2} \tilde{\beta}_i x_{i1}] = x_{i1} x_{i2} \sigma_\beta. \end{aligned}$$

Matrik kovariansinya adalah

$$\Sigma = \begin{pmatrix} x_{i1}^2 \sigma_\beta + \sigma_\varepsilon & x_{i1} x_{i2} \sigma_\beta \\ x_{i1} x_{i2} \sigma_\beta & x_{i2}^2 \sigma_\beta + \sigma_\varepsilon \end{pmatrix}$$

$$= \sigma_{\beta} \begin{pmatrix} x_{i1}^2 & x_{i1}x_{i2} \\ x_{i1}x_{i2} & x_{i2}^2 \end{pmatrix} + \sigma_{\varepsilon} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

Selanjutnya dengan normalisasi skala utiliti, yaitu yang memenuhi $\sigma_{\varepsilon} = 1$. Diperoleh

$$\Sigma = \sigma_{\beta} \begin{pmatrix} x_{i1}^2 & x_{i1}x_{i2} \\ x_{i1}x_{i2} & x_{i2}^2 \end{pmatrix} + \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

Nilai x_{n1} dan x_{n2} diobservasi oleh peneliti dan parameter b dan σ_{β} diestimasi. Generalisasi untuk lebih dari satu variabel independen dan lebih dari dua alternatif dapat dilakukan.

Dari pembahasan ini, dapat disimpulkan bahwa jika terdapat J alternatif (pilihan) maka untuk menghitung probabilitas masing-masing pilihan memerlukan integral rangkap $J-1$. Komponen matrik kovarian yang teidentifikasi maksimum sebanyak $[(J - 1)J/2] - 1$ parameters. Model *Probit* juga dapat mengakomodasi variasi individu yaitu parameter koefisien dinyatakan sebagai variabel *random* yang berdistribusi normal.

3.5 Estimasi parameter

Estimasi parameter pada model *Probit* dapat dilakukan menggunakan metode *SMLE*. Dalam metode ini melibatkan perhitungan secara simulasi.

3.5.1 Simulasi *monte carlo*

Metode *Monte Carlo* adalah algoritma komputasi untuk mensimulasikan berbagai perilaku sistem fisika dan matematika. Simulasi komputer merupakan cara yang ampuh untuk memahami sistem yang rumit dan dapat digunakan untuk menguji teori. Penggunaan nama *Monte Carlo*, yang dipopulerkan oleh para pioner bidang tersebut merupakan nama kasino terkemuka di Monako. Penggunaan ke-*random*-an dan sifat pengulangan proses mirip dengan aktivitas yang dilakukan pada sebuah kasino. Penggunaan metode *Monte Carlo* memerlukan sejumlah besar bilangan *random*, dan hal tersebut semakin mudah dengan perkembangan pembangkit bilangan *pseudo random*. Dalam kenyataannya tidak dapat diperoleh barisan bilangan yang benar-benar *random* oleh karena itu menggunakan barisan *pseudo-random* yang biasa disebut dengan Pseudo-Monte Carlo (PMC).

Salah satu permasalahan dalam DCM adalah menghitung probabilitas pilihan, π_{ij} yang merupakan integral berdimensi besar. Perhitungan π_{ij} ini dapat didekati dengan

metode Monte Carlo. Probabilitas sebuah pilihan (π_{ij}) merupakan nilai harapan dari sebuah statistik pada densitasnya, yaitu

$$\pi_{ij} = E[h(U)] = \int h(u) \cdot f(u) \cdot du \quad (3.10)$$

Dengan $U \sim f(u)$ dan $h(u)$ adalah sebuah fungsi/statistik tertentu, yang secara asimtotis ekuivalen dengan menghitung mean sampel $h(u)$. Pengambilan sampel *random* sebanyak R kali dari distribusi $f(u)$, misalkan u^r adalah nilai variabel *random* U pada pengambilan ke- r ($r=1,2,\dots,R$) dan

$$E_R[h(U)] = \frac{1}{R} \sum_{r=1}^R h(u^r)$$

Jika $E(h(U)) < \infty$ dan pengambilan bersifat independen dari distribusi U dan

$u^r \sim f(u)$ dengan $r=1,2,\dots,R$

dari sifat *strong law of large number* maka

$$\tilde{h}_R = \frac{1}{R} \sum_{r=1}^R h(u^r) \xrightarrow{a.s.} E[h(u)] = \int h(u) f(u) du$$

atau

$$P(\lim_{R \rightarrow \infty} \tilde{h}_R = E[h(u)]) = 1 \text{ dan } E[\tilde{h}_R] = E[h(u)]$$

Jika $\sigma_h^2 \equiv \text{Var}[h(u)] < \infty$, maka menggunakan Teorema limit pusat diperoleh

$$\text{Var}[\tilde{h}_R] = \frac{\sigma_h^2}{R} \rightarrow 0 \text{ untuk } R \rightarrow \infty$$

Dalam membangkitkan bilangan *random* U dari densitas $f(u)$, terdapat beberapa metode yang telah dikembangkan untuk model Probit. Beberapa metode tersebut disampaikan secara ringkas oleh Hajivassiliou, dkk. (1996), yaitu Crude Frequency Simulator, Normal Importance sampling simulator, Stern Decomposition Simulator, Geweke Hajivassiliou-Keane Simulator (GHK), Acceptance/Rejection Simulator, Gibbs sampler Simulator. GHK simulator adalah yang paling luas penggunaannya dan paling akurat dalam menghitung nilai p_{ij} pada model Probit (Geweke, dkk., 1994).

Untuk mereduksi banyaknya iterasi, telah dikembangkan metode pengambilan sampel secara sistematis, yang disebut dengan Metode Quasi Monte Carlo. Beberapa metode tersebut antara lain sistematis sampling, barisan Halton, Randomized Halton Draws, Scrambled Halton Draws. Pada model Mixed Logit, metode Barisan Halton lebih efisien dibanding metode PMC (Bhat (2001), Train (2000), Revelt dan Train (1998)).

3.5.2 Metode simulasi GHK

Untuk menghitung nilai peluang sebagaimana pada persamaan 3.6, menggunakan algoritma GHK simulator (Train, 2003). GHK simulator dapat dihitung menggunakan algoritma berikut ini,

GHK simulator dapat dihitung menggunakan algoritma berikut ini, (Train, 2003)

1. Menghitung

$$P\left(\eta_1 < \frac{-\tilde{V}_{i1k}}{c_1}\right) = \Phi\left(\frac{-\tilde{V}_{i1k}}{c_1}\right)$$

2. Mengambil sebuah nilai η_1 , yang diberi label η_1^r dari distribusi normal terpotong pada

$$\Phi\left(\frac{-\tilde{V}_{i1k}}{c_1}\right)$$

Pengambilan dapat dilakukan sebagai berikut:

- a. Mengambil dari distribusi standard uniform μ_1^r .
- b. Menghitung

$$\eta_1^r = \Phi^{-1}\left(\mu_1^r \Phi\left(\frac{-\tilde{V}_{i1k}}{c_1}\right)\right)$$

3. Menghitung:

$$P\left(\eta_2 < \frac{-(\tilde{V}_{i2k} + c_2 \eta_1)}{c_2} \mid \eta_1 = \eta_1^r\right) = \Phi\left(\frac{-(\tilde{V}_{i2k} + c_2 \eta_1^r)}{c_2}\right)$$

Mengambil sebuah nilai η_2 , diberi label η_2^r dari normal standar terpotong

$$-\frac{(\tilde{V}_{i2k} + c_2 \eta_1^r)}{c_2}$$

Pengambilan ini dapat dilakukan sebagai berikut:

a. Mengambil dari standard uniform μ_2^r .

b. Menghitung $\eta_2^r = \Phi^{-1}\left(\mu_2^r \Phi\left(-\frac{(\tilde{V}_{i2k} + c_2 \eta_1^r)}{c_2}\right)\right)$

4. Menghitung

$$P\left(\eta_3 < \frac{-(\tilde{V}_{i3k} + c_3 \eta_1 + c_3 \eta_2)}{c_3} \mid \eta_1 = \eta_1^r, \eta_2 = \eta_2^r\right) = \Phi\left(\frac{-(\tilde{V}_{i3k} + c_3 \eta_1^r + c_3 \eta_2^r)}{c_3}\right)$$

5. Dan seterusnya untuk semua alternatif kecuali k..

6. Probabilitas simulasi untuk pengambilan ke-r dari η_1, η_2, \dots adalah dihitung sebagai

$$\begin{aligned} \check{p}_k^r &= \Phi\left(\frac{-\tilde{V}_{i1k}}{c_1}\right) \cdot \Phi\left(\frac{-(\tilde{V}_{i2k} + c_2 \eta_1^r)}{c_2}\right) \cdot \Phi\left(\frac{-(\tilde{V}_{i3k} + c_3 \eta_1^r + c_3 \eta_2^r)}{c_3}\right) \dots \\ &\dots \Phi\left(\frac{-(\tilde{V}_{iJk} + c_{J1} \eta_1^r + c_{J2} \eta_2^r + \dots + c_{J(J-1)} \eta_{(J-1)}^r)}{c_J}\right) \end{aligned} \quad (3.11)$$

7. Mengulangi langkah 1-7 beberapa kali, untuk $r = 1, \dots, R$.

8. Probabilitas simulasinya adalah

$$\check{p}_k = \frac{1}{R} \sum_{r=1}^R \check{p}_k^r \quad (3.12)$$

Algoritma ini sama dengan Algoritma Genz, yaitu dengan batas bawahnya adalah - Inf.

$$\begin{aligned} p_k &= P(\tilde{U}_{ijk} < 0) = P(\infty < \tilde{U}_{ijk} < 0) \quad \forall j \neq k \\ &= P(\infty < \tilde{U}_{i1k} < 0, \infty < \tilde{U}_{i2k} < 0, \dots, \infty < \tilde{U}_{iJk} < 0) \end{aligned}$$

$$= P(\infty < \tilde{\varepsilon}_{i1k} < -\tilde{V}_{i1k}, \infty < \tilde{\varepsilon}_{i2k} < -\tilde{V}_{i2k}, \dots, \infty < \tilde{\varepsilon}_{iJk} < -\tilde{V}_{iJk})$$

Sehingga dalam algoritma Genz, diambil nilai $d_i = 0$ dan $f_i = e f_{i-1}$. Sehingga nilai Itsum sama dengan nilai \tilde{p}_k .

Misalkan Pengamatan dilakukan dengan mengambil model untuk tiga alternatif dengan memasukkan variabel atribut pembuat keputusan (X_i) dan variabel atribut masing masing alternatif (Z_{ij}). Variabel X_i biasa disebut variabel sosio ekonomik/geografi, misalnya penghasilan, jenis kelamin, asal daerah, jumlah anak. Sedangkan variabel Z_{ij} misalkan untuk pilihan penggunaan alat transportasi (Bus, mobil pribadi, sepeda motor) maka Z_{ij} dapat berupa waktu tempuh, biaya. Model utilitinya adalah

$$U_{ij} = X_i \beta_j + Z_{ij} \gamma + \varepsilon_{ij} \text{ untuk } i=1,2,\dots,n \text{ dan } j=1,2,3.$$

$$U_{i1} = \beta_{01} + X_i \beta_1 + Z_{i1} \gamma + \varepsilon_{i1},$$

$$U_{i2} = \beta_{02} + X_i \beta_2 + Z_{i2} \gamma + \varepsilon_{i2}$$

$$U_{i3} = \beta_{03} + X_i \beta_3 + Z_{i3} \gamma + \varepsilon_{i3}$$

Dengan mengambil alternatif pertama sebagai base line, maka model terestimasi yang merupakan transformasi selisih $U_{ij} - U_{i1}$ adalah

$$U_{i2}^* = (\beta_{02} - \beta_{01}) + X_i(\beta_2 - \beta_1) + (Z_{i2} - Z_{i1})\gamma + (\varepsilon_{i2} - \varepsilon_{i1})$$

$$U_{i3}^* = (\beta_{03} - \beta_{01}) + X_i(\beta_3 - \beta_1) + (Z_{i3} - Z_{i1})\gamma + (\varepsilon_{i3} - \varepsilon_{i1})$$

$$U_{i1}^* = 0$$

Dengan matrik kovariansi ($\text{cov}(\varepsilon_{i1}, \varepsilon_{i2}, \varepsilon_{i3})$) sama dengan:

$$\Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{12} & \sigma_{22} & \sigma_{23} \\ \sigma_{13} & \sigma_{23} & \sigma_{33} \end{pmatrix}$$

Algoritma GHK tersebut dapat dituliskan dalam program R sebagai berikut:

#Program mengitung nilai probailitas dengan metode GHK

GHK<-function(b)

```
{
  b01=b[1]; b02=b[2]; b03=b[3]; b1=b[4]; b2=b[5]; b3=b[6]
  g=b[7]; s11=b[8];s12=b[9];s13=b[10];s22=b[11]
  s23=b[12];s33=b[13]
  V21=(b02-b01) + ( b2-b1)*X+g*(Z[,2]-Z[,1])
  V31=(b03-b01) + (b3-b1)*X+g*(Z[,3]-Z[,1])
  V32=V31-V21
  S=matrix(c(s11,s12,s13,s12,s22,s23,s13,s23,s33),3,3)
  M1= matrix(c(-1,-1,1,0,0,1),2,3)
  M2=matrix(c(1,0,-1,-1,0,1),2,3)
  S1= M2%%S%%t(M2)
  S2=M2%%S%%t(M2)
  F1=matrix(,N,1); F2= matrix(,N,1);F3= matrix(,N,1)
  Prob=matrix(,N,3)
  for (i in 1:N)
  {
    F1= sadmvn(-lnf, c(-V21[i],-V31[i]), mean=rep(0,2),S1)
    F2= sadmvn(-lnf, c(V21[i],-V32[i]), mean=rep(0,2),S2)
    F3= 1 - F1-F2
    Prob[i,]=c(F1,F2,F3)
  }
  Prob
}
```

Misalkan diambil data dengan $N= 500$, $X_i \sim \text{NID}(0,1)$, $Z_{ij} \sim \text{NID}(0,1)$ dan $\epsilon_i \sim N(0,\Sigma)$

$$\Sigma = \begin{pmatrix} 1 & 0.2 & 0.7 \\ 0.2 & 1 & 0.7 \\ 0.7 & 0.7 & 1 \end{pmatrix}$$

Parameter $\beta_{01}=2$, $\beta_{02}=1$, $\beta_{03}=0.5$, $\beta_1=-2$, $\beta_2=-1$, $\beta_3=-1$ dan $\gamma=0.8$, maka nilai probabilitas masing-masing alternatif (pilihan) pada masing-masing individu dapat dihitung sebagai berikut:

```
> library(mnormt)
> X=rnorm(N,3); prop.pilihan= matrix(,N,3); Z=matrix(,N,3); V=matrix(,N,3)
> for (i in 1:3){Z[,i]=rnorm(N)}
> prop.pilihan=GHK(c(2,1,0.5,-2,-1,-1,0.8,1,0.2,0.7,1,0.7,1))
```

3.6 Simulasi GHK dalam MSLE

Dalam model Probit terdapat parameter dalam matrik faktor *Cholesky* dan parameter koefisien (β). Parameter tersebut disusun dalam satu *vector* $\theta=(\beta,c)$ dengan $\beta=(\alpha_1, \dots, \alpha_J, \beta_1, \dots, \beta_J, \gamma)$ dan $c=(c_{11}, c_{21}, \dots, c_{(J-1)(J-1)})$ yang dapat diestimasi menggunakan fungsi *log-likelihood simulated*

$$\begin{aligned} \text{simlog}L(\theta) &= \sum_{i=1}^n \sum_{j=1}^J y_{ij} \log \left(\frac{1}{R} \sum_{r=1}^R \tilde{\pi}_j^{(r)}(\theta) \right) \\ &= \sum_{i=1}^n \sum_{j=1}^J y_{ij} \log \left(\frac{1}{R} \sum_{r=1}^R \prod_{l=1}^J \Phi_{lir} \right) \end{aligned} \tag{3.13}$$

Penaksir θ yang dihitung menggunakan metode *Newton-Raphson* memerlukan derivatif pertama dan derivatif ke dua dari fungsi *log-likelihood simulated*. Untuk menghindari derivatif kedua dari fungsi *log-likelihood simulated*, dapat digunakan metode iterasi BHHH atau metode iterasi BFGS.

Nilai batas distribusi normal pada persamaan (3.11) dapat dinotasikan dengan:

$$\eta_{1i} < a_{1i} = \frac{-V_{i1,k}}{c_1}$$

$$\eta_{2i} < a_{2i} = \frac{-(V_{i2,k} + c_2 \eta_1^{(r)})}{c_2} \dots$$

$$\eta_{(J-1)i} < a_{(J-1)i} = \frac{-(V_{i,J,k} + c_{(J-1)1} \eta_1^{(r)} + c_{(J-1)2} \eta_2^{(r)} + \dots + c_{(J-1)(J-2)} \eta_{(J-2)}^{(r)})}{c_{(J-1)(J-1)}}$$

sehingga secara ringkas dapat dinyatakan dalam bentuk

$$a_i = \begin{cases} - \left[\sum_{h=1}^{l-1} \frac{c_h}{c_l} \eta_h + \frac{V_{l,k}}{c_l} \right] & l > 1 \\ - \frac{V_{i1,k}}{c_1} & l = 1 \end{cases} \quad (3.14)$$

Persamaan tersebut menjadi

$$\begin{aligned} \tilde{\pi}_k^{(r)} &= P(\eta_{1i}^{(r)} < a_{1i}^{(r)}) P(\eta_{2i}^{(r)} < a_{2i}^{(r)}) \dots P(\eta_{(J-1)i}^{(r)} < a_{(J-1)i}^{(r)}) \\ \tilde{\pi}_k^{(r)} &= \Phi(a_{1i}^{(r)}) \Phi(a_{2i}^{(r)}) \dots \Phi(a_{(J-1)i}^{(r)}) = \Phi_{1i}^{(r)} \cdot \Phi_{2i}^{(r)} \dots \Phi_{(J-1)i}^{(r)} = \prod_{l=1}^{J-1} \Phi_l^{(r)} \\ \tilde{\pi}_k &= \frac{1}{R} \sum_{r=1}^R \tilde{\pi}_k^{(r)} = \frac{1}{R} \sum_{r=1}^R \left(\prod_{l=1}^{J-1} \Phi_l^{(r)} \right) \end{aligned} \quad (3.15)$$

Boulduc (1999) telah menghitung derivatif pertama fungsi simulated log-likelihood,

$$= \sum_{i=1}^n \sum_{j=1}^J \frac{y_j}{\frac{1}{R} \sum_{r=1}^R \tilde{\pi}_j^{(r)}} \left(\frac{1}{R} \sum_{r=1}^R \frac{\partial}{\partial \theta} \left(\prod_{l=1}^{J-1} \Phi_l^{(r)} \right) \right) \quad (3.16)$$

Dari sifat bahwa

$$\frac{\partial}{\partial \theta} \prod_{j=1}^{J-1} \Phi_j = \prod_{j=1}^{J-1} \Phi_j \sum_{l=1}^{J-1} \frac{\partial \log \Phi_l}{\partial \theta}$$

maka

$$\frac{\partial}{\partial \boldsymbol{\theta}} \text{simlog}L(\boldsymbol{\theta}) = \sum_{i=1}^n \sum_{j=1}^J \frac{y_{ij}}{\frac{1}{R} \sum_{r=1}^R \tilde{\pi}_{ij}^{(r)}} \frac{1}{R} \sum_{r=1}^R \left(\left(\prod_{j=1}^{J-1} \Phi_j^{(r)} \right) \left(\sum_{l=1}^{J-1} \frac{\partial \log \Phi_l^{(r)}}{\partial \boldsymbol{\theta}} \right) \right) \quad (3.17)$$

$$\frac{\partial}{\partial \boldsymbol{\theta}} \text{simlog}L(\boldsymbol{\theta}) = \sum_{i=1}^n \sum_{j=1}^J \frac{y_{ij}}{\frac{1}{R} \sum_{r=1}^R \tilde{\pi}_{ij}^{(r)}} \left(\frac{1}{R} \sum_{r=1}^R \left(\tilde{\pi}_{ij}^{(r)} \sum_{l=1}^{J-1} \frac{\partial \log \Phi_l^{(r)}}{\partial \boldsymbol{\theta}} \right) \right)$$

Selanjutnya

$$\frac{\partial \log \Phi_h^{(r)}}{\partial \boldsymbol{\theta}} = \frac{\phi_h^{(r)}}{\Phi_h^{(r)}} \cdot \frac{\partial a_h^{(r)}}{\partial \boldsymbol{\theta}} \quad (3.18)$$

dengan $\phi_h^{(r)} = \phi(a_h^{(r)})$ yang merupakan fungsi densitas normal. Karena $a_h^{(r)}$ adalah fungsi dari $\eta_{hi,r}$

$$\eta_h^{(r)} = \Phi^{-1}(u_h^{(r)} \Phi(a_h^{(r)}))$$

maka

$$\Phi(\eta_h^{(r)}) = (u_h^{(r)} \Phi(a_h^{(r)}))$$

$$\Leftrightarrow \frac{\partial \Phi(\eta_h^{(r)})}{\partial \boldsymbol{\theta}} = \frac{\partial}{\partial \boldsymbol{\theta}} (u_h^{(r)} \Phi(a_h^{(r)}))$$

$$\Leftrightarrow \phi(\eta_h^{(r)}) \frac{\partial \eta_h^{(r)}}{\partial \boldsymbol{\theta}} = u_h^{(r)} \cdot \phi(a_h^{(r)}) \frac{\partial a_h^{(r)}}{\partial \boldsymbol{\theta}}$$

sehingga

$$\frac{\partial \eta_h^{(r)}}{\partial \boldsymbol{\theta}} = u_h^{(r)} \cdot \frac{\phi(a_h^{(r)})}{\phi(\eta_h^{(r)})} \cdot \frac{\partial a_h^{(r)}}{\partial \boldsymbol{\theta}} \quad (3.19)$$

dengan $u_h^{(r)}$ menyatakan suatu nilai yang diambil secara *random* dari distribusi uniform pada interval yang ditentukan. Selanjutnya persamaan (3.14) diderivatiskan terhadap parameter β adalah

$$\frac{\partial a_i^{(r)}}{\partial \beta} = \begin{cases} -\sum_{h=1}^{l-1} \frac{c_h}{c_l} \frac{\partial \eta_h^{(r)}}{\partial \beta} - \frac{1}{c_l} \frac{\partial V_{l,k}}{\partial \beta} & l > 1 \\ -\frac{1}{c_1} \frac{\partial V_{l,k}}{\partial \beta} & l = 1 \end{cases}$$

$$\frac{\partial a_i^{(r)}}{\partial \beta} = \begin{cases} -\left[\sum_{h=1}^{l-1} \frac{c_h}{c_l} u_h^{(r)} \cdot \frac{\phi(a_h^{(r)})}{\phi(\eta_h^{(r)})} \cdot \frac{\partial a_h^{(r)}}{\partial \beta} + \frac{1}{c_l} \frac{\partial V_{l,k}}{\partial \beta} \right] & l > 1 \\ -\frac{1}{c_1} \frac{\partial V_{l,k}}{\partial \beta} & l = 1 \end{cases} \quad (3.20)$$

Selanjutnya dari persamaan (3.14) diderivatifkan terhadap

$$\frac{\partial a_i^{(r)}}{\partial c_j} = 0 \text{ untuk } j > l.$$

Dari persamaan (3.18), diperoleh

$$\frac{\partial \mathbf{h} \Phi_i^{(r)}}{\partial c_j} = \begin{cases} \frac{\phi_i^{(r)}}{\Phi_i^{(r)}} \cdot \frac{\partial a_i^{(r)}}{\partial c_j} & \text{untuk } k \leq j \leq l \\ 0 & \text{untuk } j > l \end{cases} \quad (3.21)$$

dan berdasarkan persamaan (3.19) diperoleh

$$\frac{\partial \eta_h^{(r)}}{\partial c_j} = u_h^{(r)} \cdot \frac{\phi(a_h^{(r)})}{\phi(\eta_h^{(r)})} \cdot \frac{\partial a_h^{(r)}}{\partial c_j}$$

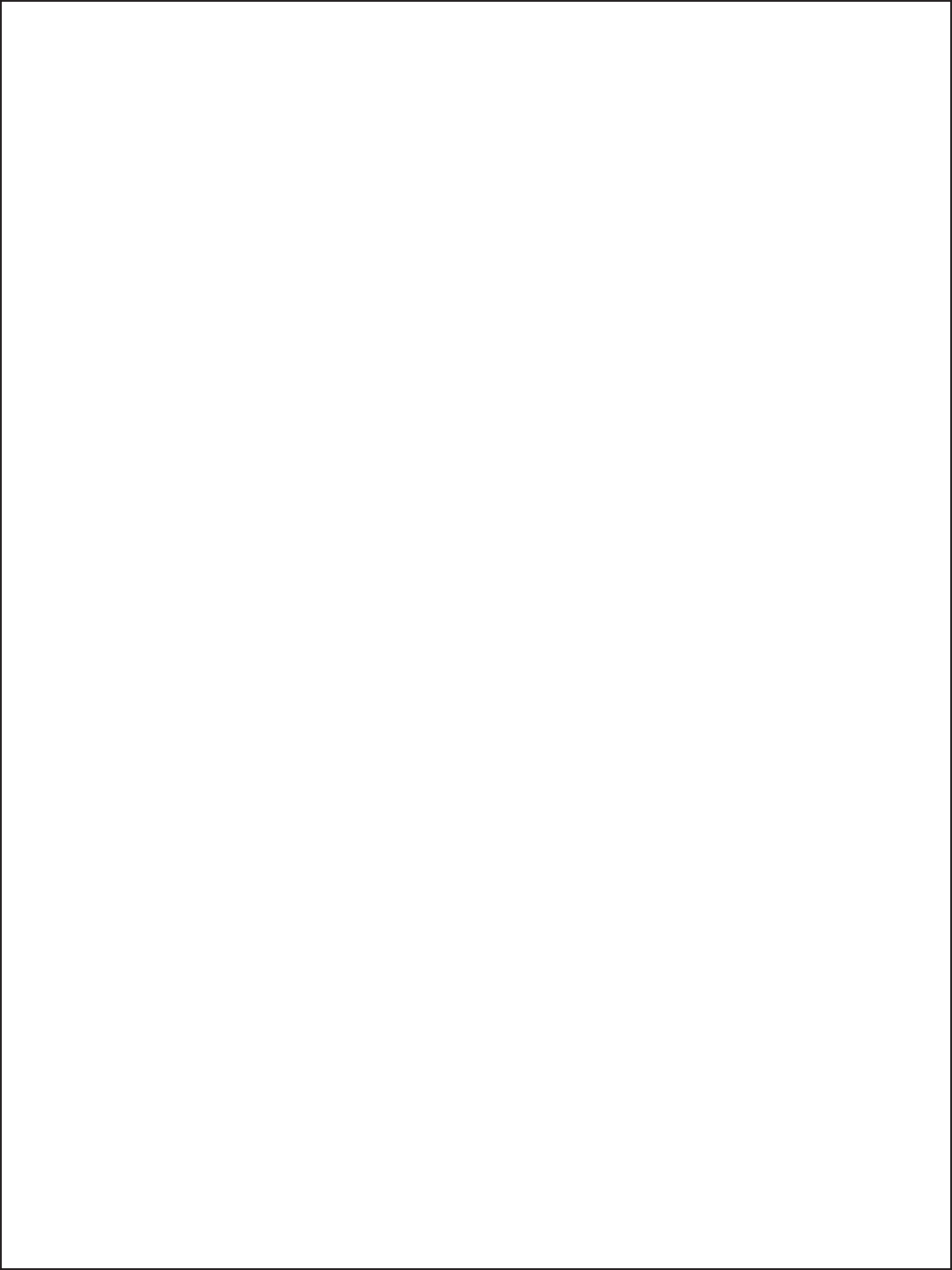
sehingga

$$\frac{\partial a_i^{(r)}}{\partial c_j} = \begin{cases} -\sum_{h=1}^{l-1} u_h^{(r)} \frac{c_h}{c_l} \frac{\phi(a_h^{(r)})}{\phi(\eta_h^{(r)})} \frac{\partial a_h^{(r)}}{\partial c_j} & \text{untuk } k \leq j < l \\ -\frac{\eta_k^{(r)}}{c_l} & \text{untuk } k < j = l \\ -\frac{a_l^{(r)}}{c_l} & \text{untuk } k = j = l \end{cases} \quad (3.22)$$

Jadi,

$$\frac{\partial}{\partial \boldsymbol{\theta}} \text{simlog}L(\boldsymbol{\theta}) = \sum_{i=1}^n \sum_{j=1}^J \frac{y_j}{\frac{1}{R} \sum_{r=1}^R \tilde{\pi}_j^{(r)}} \left(\frac{1}{R} \sum_{r=1}^R \left(\tilde{\pi}_j^{(r)} \sum_{l=1}^{J-1} \frac{\phi_l^{(r)}}{\Phi_l^{(r)}} \cdot \frac{a_l^{(r)}}{\partial \boldsymbol{\theta}} \right) \right) \quad (3.23)$$

dengan $\frac{\partial a_i^{(r)}}{\partial \boldsymbol{\beta}}$ dan $\frac{\partial a_i^{(r)}}{\partial c_k}$ masing-masing seperti pada persamaan (3.20) dan (3.22)



Bab 4

Model Mixed Logit

4.1 Pendahuluan

Tujuan dari analisis DCM adalah menghitung probabilitas responden dalam memilih sebuah pilihan. Persamaan probabilitas pada model *Logit* berupa persamaan tertutup yang perhitungannya relatif sederhana dibandingkan persamaan probabilitas dalam model *Probit* dan model *Mixed Logit* yang berbentuk persamaan terbuka dengan melibatkan integral rangkap. Dalam menghitung proporsi masing-masing pilihan, khususnya pada model *Probit* dan model *Mixed Logit*, simulasi memainkan peranan penting.

Model *Probit* maupun *Logit* untuk respon *Multinomial* telah menjadi perhatian dan banyak digunakan dalam berbagai bidang seperti dalam bidang transportasi, marketing, psikologi, dan survei (Rodriguez (2001), Agresti (1990), Ruud (1996)). Model *Multinomial Logit* lebih disenangi karena menghasilkan persamaan probabilitas dalam bentuk persamaan tertutup, sedangkan model *Multinomial Probit* (MNP) menghasilkan persamaan probabilitas dalam bentuk persamaan terbuka dengan melibatkan integral rangkap.

McFadden dan Train (2000) telah menyampaikan model *Mixed Multinomial Logit* yang disingkat dengan *Mixed Logit*. Model dapat diturunkan menjadi model *Logit* maupun model *Probit*. Beberapa peneliti memberi nama *Error Component* model atau *Logit Kernel Probit* model, ada juga yang menyebut dengan *Logit Kernel Mixed Logit* telah sangat populer dalam beberapa literatur dan diaplikasikan dalam berbagai bidang, seperti bidang transportasi, marketing, energi, perumahan dan lain-lain (Walker, dkk., 2004). Walker, dkk. (2004) menyebutkan bahwa model *Mixed Logit* merupakan "*model of the future.*"

Bhat (2001) memberikan hasil dari penelitiannya mengenai simulasi Monte Carlo untuk estimasi pada *Mixed Logit* yang menggunakan barisan pseudo-random standar dibandingkan dengan barisan Halton. Barisan Halton lebih efisien dibanding barisan pseudo-random standar. Demikian juga Train (2000), Revelt dan Train (1998) memberikan hasil yang sama. MNL memerlukan asumsi *Independence from irrelevant alternatives* (IIA). Pelanggaran asumsi ini mengakibatkan estimator yang diperoleh menjadi bias. Semakin besar dependensi antar alternatif mengakibatkan bias estimatornya semakin besar (Nugraha, dkk., 2007). Sementara itu, MNP secara teori

dapat mengakomodasi adanya korelasi antar pilihan tetapi dalam implementasinya masih menyisakan masalah dalam proses komputasinya. Dalam bab ini dibahas model *Mixed Logit* dalam menangani masalah dependensi antar alternatif. Model diaplikasikan pada data simulasi pada beberapa tingkat dependensi (korelasi) antar alternatif. Proses komputasi menggunakan program R.2.8.1.

4.2 Model *mixed logit*

Mixed Logit adalah model yang sangat fleksibel yang dapat didekati oleh banyak model *random* utilitas (McFadden dan Train, 2000). Dari persamaan (1.2), ditambahkan variabel *random* δ_j mempunyai densitas $f(\delta_j)$. Model utilitasnya dapat disajikan dalam bentuk

$$U_{ij} = V_{ij} + \delta_j + \varepsilon_{ij} \quad (4.1)$$

Pada umumnya digunakan asumsi $f(\gamma_j)$ berdistribusi normal dengan mean nol dan variansi 1.

$$\delta_j \sim N(0,1) \text{ dan } \delta = (\delta_1, \dots, \delta_j) \sim N(0, \Sigma)$$

Diasumsikan ε_{ij} berdistribusi extreme value dan independen terhadap δ_j .

Model *Mixed Logit* merupakan integral *Logit* standar terhadap densitas δ . Probabilitas responden i memilih alternatif k dapat dirumuskan menjadi

$$\pi_{ik} = \int g_{ik}(\delta) f(\delta) d\delta \quad (4.2)$$

$g_{ik}(\delta)$ adalah probabilitas logit yang dapat dituliskan sebagai berikut:

$$g_{ik}(\delta) = \frac{\exp[V_{ik} + \delta_k]}{\sum_{j=1}^J \exp[V_j + \delta_j]} \quad (4.3)$$

Probabilitas dalam *Mixed Logit* merupakan pembobotan secara rata-rata terhadap logit dengan menggunakan pembobot fungsi densitas $f(\delta)$. *Mixed Logit* adalah campuran antara fungsi logit dan fungsi densitas $f(\delta)$. Nilai probabilitas pada persamaan (4.2) dapat didekati menggunakan simulasi. Langkah-langkah simulasinya sebagai berikut (Train, 2003):

1. Mengambil sebuah nilai δ dari densitas $f(\delta)$ dan diberi label $\delta^{(r)}$. Pada pengambilan pertama, $r=1$.
2. Menghitung probabilitas logit $g_{ik}(\delta^{(r)})$ pada persamaan (4.2).
3. Mengulangi langkah 1 dan 2 sebanyak R dan menghitung rata-rata

$$\tilde{\pi}_k = \frac{1}{R} \sum_{r=1}^R g_k(\delta^{(r)}) \quad (4.4)$$

Train (2000) dan Bhat (2001) telah melakukan pengujian mengenai metode simulasi untuk model *Mixed Logit*. Mereka menyatakan bahwa pengambilan sampel secara sistematis menggunakan metode *Halton* lebih efisien dibanding dengan pengambilan secara *random*. $R = 100$ pada metode *Halton* sebanding dengan $R = 1000$ pada pengambilan secara *random*.

Parameter yang akan diestimasi dalam model *Mixed Logit* adalah *mean* dan *variansi* dalam densitas $f(\delta)$ dan parameter dalam V_{ij} .

4.3 Metode *maximum likelihood estimators* (MLE)

Misalkan Y_1, Y_2, \dots, Y_n adalah variabel *random* yang mempunyai densitas gabungan.

$$f(y; \theta) = f(y_1, y_2, \dots, y_n; \theta_1, \dots, \theta_j) \quad (4.5)$$

Fungsi ini tergantung pada parameter $\theta = (\theta_1, \dots, \theta_j)$. Jika Y_i saling independen maka

$$f(y; \theta) = \prod_{i=1}^n f(y_i, \theta) \quad (4.6)$$

Fungsi *likelihood* $L(\theta; y)$, secara aljabar sama dengan $f(y; \theta)$ yang merupakan fungsi dari θ untuk suatu nilai y (data sampel).

$$L = L(\theta; y) = f(y; \theta) \quad (4.7)$$

Misalkan Ω merupakan himpunan semua nilai yang mungkin untuk vektor parameter θ (Ω disebut juga ruang parameter). Greene (2005), mendefinisikan MLE untuk θ , yang dinotasikan dengan $\hat{\theta}_{MLE}$ adalah nilai θ yang memaksimalkan fungsi *likelihood* $L(\theta; y)$ pada data y

$$\hat{\theta}_{MLE} = \arg \max_{\theta} L(\theta; y) \quad (4.8)$$

4.4 Barisan Halton

Barisan Halton dibangkitkan dari sebuah bilangan yang ditentukan, biasanya bilangan prima. Secara umum barisan dari bilangan prima k dibuat secara iteratif, dengan barisan pada iterasi ke- $(t+1)$ adalah

$$s_{(t+1)} = \{s_t, s_t + 1/k^t, s_t + 2/k^t, \dots, s_t + (k-1)/k^t\}$$

Barisan Halton didefinisikan pada unit interval $[0,1]$. Ketika pengambilan sampel menggunakan *Barisan Halton*, maka dibuat barisan berdasar bilangan prima yang dipilih. Panjang barisan disesuaikan dengan banyaknya observasi. Untuk menghilangkan korelasi antar observasi, biasanya 10 bilangan pertama dalam barisan dibuang. Nilai pada *Barisan Halton* merupakan densitas kumulatif dari variabel *random* yang ingin dibangkitkan. Sehingga variabel *random* yang dibangkitkan didapatkan dengan cara mencari *invers* fungsi densitas kumulatif yang mempunyai nilai sesuai dengan *Barisan Haltonnya*.

4.5 Studi simulasi

Untuk mendeteksi pengaruh korelasi antar alternatif, akan dibangkitkan data Multinomial untuk $J=3$ dan $n=1000$. Diasumsikan terdapat korelasi antara alternatif pertama dan alternatif kedua. Sebagai contoh permasalahan moda transportasi. Terdapat tiga alternatif yaitu kendaraan pribadi, taksi dan angkutan umum. Alternatif taksi kemungkinan besar berkorelasi dengan kendaraan pribadi. Artinya bahwa pada orang yang biasanya naik kendaraan pribadi, jika tidak ada kendaraan pribadi akan memilih naik taksi, begitu juga sebaliknya.

Selanjutnya untuk mengestimasi parameter korelasi akan digunakan dua model, yaitu model I dan Model II. Utilitas pada Model I adalah

$$U_{i1} = V_{i1} + \delta_i + \varepsilon_{i1},$$

$$U_{i2} = V_{i2} + \delta_i + \varepsilon_{i2}, \tag{4.9}$$

$$U_{i3} = V_{i3} + \varepsilon_{i3}$$

$\varepsilon_{ij} \sim$ Extreme value tipe I, $\delta_i \sim N(0, \sigma_\delta^2)$. ε_{ij} dan δ_i saling independen untuk semua j dan $i. i=1, \dots, 1000$ dan $j=1,2,3$.

Utilitas pada Model II adalah

$$\begin{aligned}
 U_{i1} &= V_{i1} + \delta_{i1} + \varepsilon_{i1}, \\
 U_{i2} &= V_{i2} + \delta_{i2} + \varepsilon_{i2}, \\
 U_{i3} &= V_{i3} + \varepsilon_{i3}
 \end{aligned}
 \tag{4.10}$$

$\varepsilon_{it} \sim$ Extreme value tipe I, $\delta_{i1} \sim N(0,1)$, $\delta_{i2} \sim N(0,1)$ $\delta = (\delta_1, \delta_2) \sim N(0, \Sigma)$.

$$\Sigma = \begin{pmatrix} 1 & \sigma_{\delta} \\ \sigma_{\delta} & 1 \end{pmatrix} \text{ sehingga } -1 < \sigma_{\delta} < 1$$

Diasumsikan

$$V_{ij} = \alpha_j + \beta_j X_i + \gamma Z_{ij},$$

X_i merupakan variabel karakteristik individu dan Z_{ij} merupakan variabel karakteristik pilihan. Karena diasumsikan alternatif 3 sebagai *baseline* maka $\alpha_3 = \beta_3 = 0$. Data dibangkitkan pada parameter:

| | | | | |
|-----------------|----------------|-----------------|------------------|--------------|
| $\alpha_1 = -1$ | $\alpha_2 = 1$ | $\beta_1 = 0.5$ | $\beta_2 = -0.5$ | $\gamma = 1$ |
|-----------------|----------------|-----------------|------------------|--------------|

dan pada beberapa nilai $\sigma_{\delta}^2, \sigma_{12}$.

Berdasarkan data simulasi ini, disusun model menggunakan model *MNL* dan model *Mixed*. Selanjutnya estimator yang diperoleh pada beberapa nilai σ_{δ}^2 dan σ_{12} dari model *Mixed Logit* dibandingkan dengan model *MNL*.

4.5.1 Menghitung korelasi utilitas

Kovariansi antar pilihan pada persamaan (4.9) adalah:

$$Cov(U_i) = \begin{pmatrix} \sigma_{\delta}^2 + \frac{\pi^2}{6} & \sigma_{\delta} & 0 \\ \sigma_{\delta} & \sigma_{\delta}^2 + \frac{\pi^2}{6} & 0 \\ 0 & 0 & \frac{\pi^2}{6} \end{pmatrix}
 \tag{4.11}$$

Sehingga korelasi antara pilihan 1 dan pilihan 2 adalah:

$$\rho = \text{Cor}(U_{i1}, U_{i2}) = \frac{\sigma_{\delta}^2}{\sigma_{\delta}^2 + \frac{\pi^2}{6}} \geq 0 \text{ dan } \sigma_{\delta}^2 = \frac{\text{Cor}(U_{i1}, U_{i2}) \frac{\pi^2}{6}}{1 - \text{Cor}(U_{i1}, U_{i2})} \quad (4.12)$$

Persamaan 4.12 merupakan hubungan antara σ_{δ}^2 terhadap nilai korelasi. Data dibangkitkan pada beberapa nilai σ_{δ}^2 dan konversinya ke nilai korelasi yang disajikan dalam Tabel 4.1.

Tabel 4.1 Konversi variansi dan korelasi

| | | | | | | | |
|---------------------|----------|----------|----------|----------|----------|----------|----------|
| σ_{δ}^2 | 0.1 | 0.2 | 0.3 | 0.5 | 1 | 2 | 3 |
| ρ | 0.057364 | 0.108503 | 0.154379 | 0.233289 | 0.37832 | 0.548958 | 0.646097 |
| σ_{δ}^2 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| ρ | 0.708809 | 0.752642 | 0.785005 | 0.809879 | 0.829594 | 0.845605 | 0.858865 |
| σ_{δ}^2 | 11 | 12 | 13 | 14 | 15 | | |
| ρ | 0.870028 | 0.879555 | 0.88778 | 0.894954 | 0.901265 | | |

Kovariansi antar pilihan pada persamaan (4.10) adalah

$$\text{Cov}(U_i) = \begin{pmatrix} 1 + \frac{\pi^2}{6} & \sigma_2 & 0 \\ \sigma_2 & 1 + \frac{\pi^2}{6} & 0 \\ 0 & 0 & \frac{\pi^2}{6} \end{pmatrix} \quad (4.13)$$

sehingga korelasi antara pilihan 1 dan pilihan 2 adalah

$$\rho = \text{Cor}(U_{i1}, U_{i2}) = \frac{\sigma_2}{1 + \frac{\pi^2}{6}} \geq 0 \text{ dan } \sigma_2 = \text{Cor}(U_{i1}, U_{i2}) \cdot \left(1 + \frac{\pi^2}{6}\right)$$

Karena $-1 < \sigma_{\rho} < 1$ maka nilai $Cor(U_{i1}, U_{i2})$ yang terestimasi adalah

$$-\frac{1}{1 + \frac{\pi^2}{6}} < Cor(U_{i1}, U_{i2}) < \frac{1}{1 + \frac{\pi^2}{6}}$$

$$-0.3783 < Cor(U_{i1}, U_{i2}) < 0.37832 \quad (4.14)$$

Berdasarkan persamaan 4.14, model II tidak dapat digunakan untuk mengestimasi korelasi di luar interval tersebut. Oleh karena itu untuk mengestimasi parameter akan digunakan model utilitas dari persamaan 4.9.

4.5.2 Program estimasi parameter

Paket Program R.2.8.1 digunakan untuk membangkitkan data maupun untuk mengestimasi parameter. Fungsi *log-likelihood* untuk model MNL maupun model *Mixed Logit* dalam program R adalah:

```
MLE.logit<-function(a)
{
b01=a[1];b02=a[2];b1=a[3];b2=a[4];g=a[5]
p1=sum(Y1*(b01+X*b1+Z1*g) - Y1*log(exp(b01+X*b1+Z1*g)
)+exp(b02+X*b2+Z2*g)
+ exp( Z3*g )))
p2=sum(Y2*(b02+X*b2+Z2*g)- Y2*log(exp(b01+X*b1+Z1*g)
)+exp(b02+X*b2+Z2*g)
+ exp( Z3*g )))
p3=sum(Y3*(Z3*g) - Y3*log(exp(b01+X*b1+Z1*g)
)+exp(b02+X*b2+Z2*g)
+ exp(Z3*g )))
lg= p1+p2+p3
lg
}
MLE.mixed<-function (b)
{
b0=b[1:2];b1=b[3:4];c=b[5];bs=abs(b[6]);
g01= exp(sqrt(bs)*t(e))%x%exp(b0[1]+b1[1]*X+c*Z[,1])
g02= exp(sqrt(bs)*t(e))%x%exp(b0[2]+b1[2]*X+c*Z[,2])
g03= exp(0*t(e))%x%exp(c*Z[,3])
y1=as.vector(Y1); y2=as.vector(Y2); y3=as.vector(Y3)
```

```

g1=(g01/(g01+g02+g03))%*%ls/(n.ht)
g2=(g02/(g01+g02+g03))%*%ls/(n.ht)
g3=(g03/(g01+g02+g03))%*%ls/(n.ht)
g= (g1^y1)*(g2^y2)*(g3^y3)
-sum(log(g))
}

```

Beberapa *library* dalam program R yang dipakai adalah

- library(randtoolbox)*: digunakan untuk membangkitkan *Barisan Halton* yang digunakan dalam menghitung integral densitas normal.
- library(micEcon)*: digunakan untuk menghitung MLE pada model MNL.
- library(mnormt)*: digunakan untuk membangkitkan data multivariat normal.
- library(adapt)*: digunakan untuk mencari titik maksimum dari fungsi *log-likelihood* pada model *Mixed Logit*.

4.5.3 Hasil simulasi

Hasil estimasi parameter korelasi disajikan dalam Tabel 4.2 sedangkan estimasi parameter regresi disajikan dalam Tabel 4.3 dan Gambar 4.1 s/d 4.5.

Tabel 4.2 Hasil estimasi korelasi

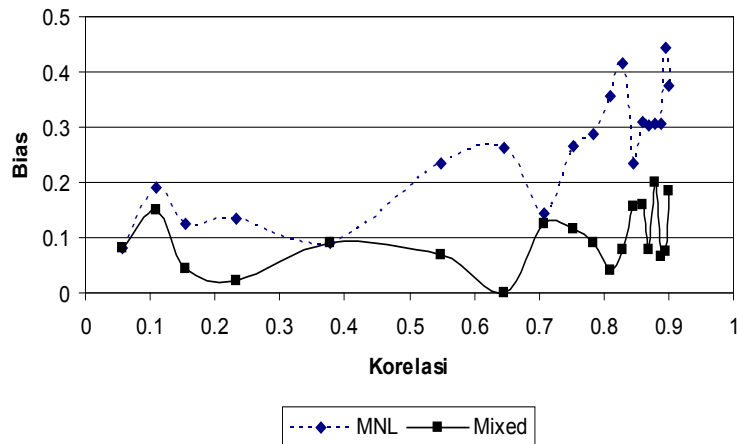
| σ_s^2 | ρ | $\hat{\sigma}_s^2$ | $\hat{\rho}$ |
|--------------|----------|--------------------|--------------|
| 0.1 | 0.057364 | 1.085E-05 | 6.6012E-06 |
| 0.2 | 0.108503 | 0.3210743 | 0.16345141 |
| 0.3 | 0.154379 | 0.7283749 | 0.30711846 |
| 0.5 | 0.233289 | 1.0737973 | 0.39520501 |
| 1 | 0.37832 | 1.0956326 | 0.400027 |
| 2 | 0.548958 | 2.0354631 | 0.55330596 |
| 3 | 0.646097 | 3.3105224 | 0.66828086 |
| 4 | 0.708809 | 4.1270951 | 0.71522294 |
| 5 | 0.752642 | 5.659063 | 0.77496679 |
| 6 | 0.785005 | 5.9036829 | 0.78226081 |
| 7 | 0.809879 | 6.1002293 | 0.7877875 |
| 8 | 0.829594 | 7.7276937 | 0.82464266 |
| 9 | 0.845605 | 8.3275278 | 0.835192 |
| 10 | 0.858865 | 8.6954989 | 0.84105775 |

| | | | |
|----|----------|------------|------------|
| 11 | 0.870028 | 9.5514109 | 0.85321 |
| 12 | 0.879555 | 9.8975351 | 0.85761243 |
| 13 | 0.88778 | 10.3296361 | 0.86275119 |
| 14 | 0.894954 | 13.918564 | 0.89440403 |
| 15 | 0.901265 | 18.2909094 | 0.91756536 |

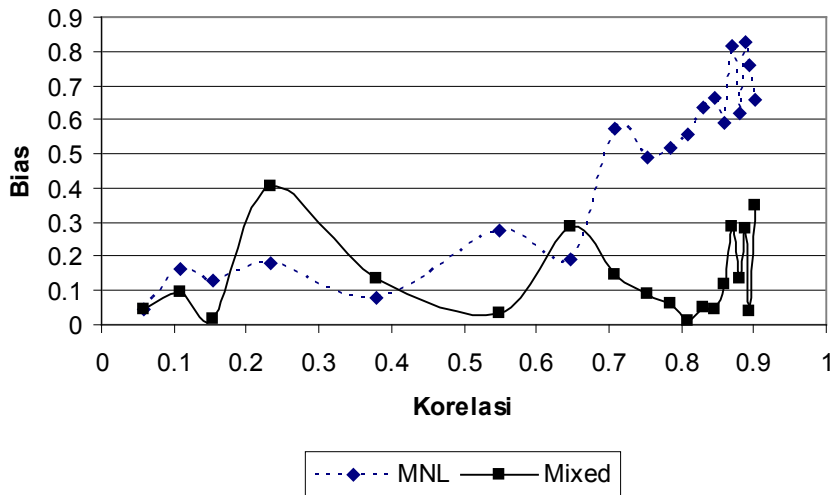
Dari Tabel 4.2 dapat terlihat bahwa model Mixed Logit dapat mengestimasi parameter korelasi.

Selanjutnya, bias untuk masing-masing parameter ($\alpha_1, \alpha_2, \beta_1, \beta_2, \gamma$) dapat dilihat pada Tabel 4.3. serat Gambar 4.1 s/d Gambar 4.5. Dari gambar tersebut dapat diperoleh beberapa kesimpulan

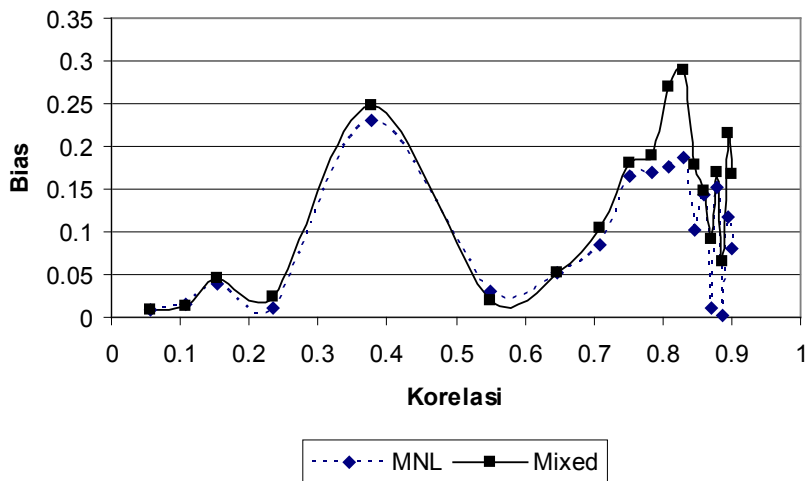
1. Pada umumnya bias yang dihasilkan untuk model MNL lebih besar dibanding bias pada model *Mixed Logit*.
2. Untuk parameter intersep (yaitu: α_1, α_2), pada korelasi tinggi (lebih dari 0.7) MNL menghasilkan bias yang lebih besar dari pada model *Mixed Logit*.
3. Untuk parameter koefisien X (yaitu: β_1, β_2), bias yang dari model MNL relatif sama dengan model *Mixed Logit*.
4. Untuk parameter koefisien Z (yaitu γ), pada korelasi tinggi (lebih dari 0.5) MNL menghasilkan bias yang lebih besar dari pada model *Mixed Logit*.



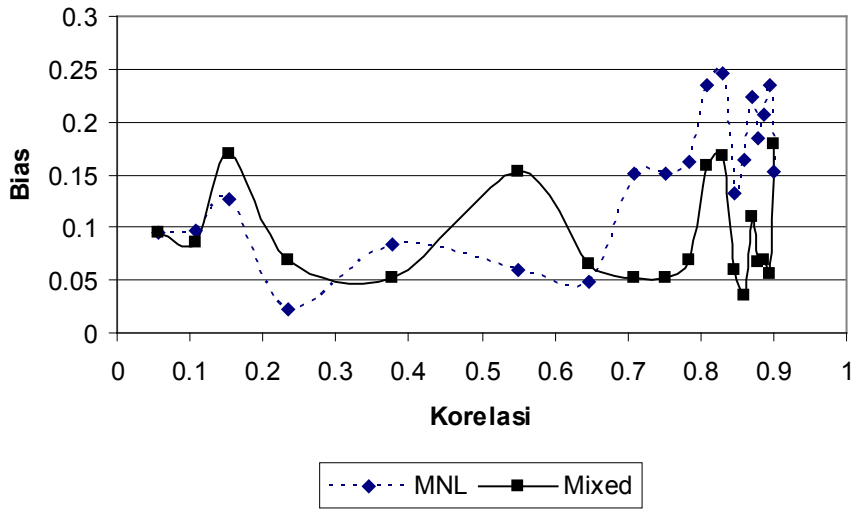
Gambar 4.1 Bias estimator parameter α_1 .



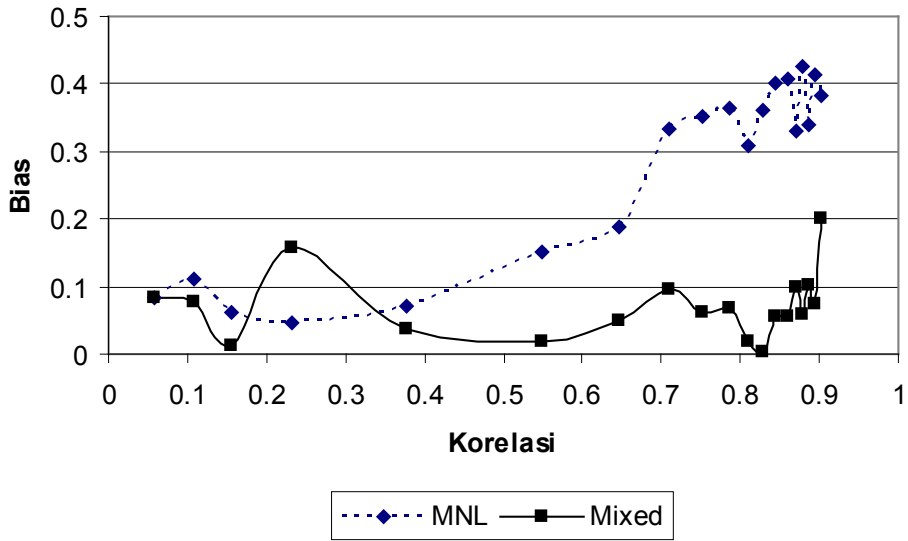
Gambar 4.2 Bias estimator parameter α_2 .



Gambar 4.3 Bias estimator parameter β_1 .



Gambar 4.4 Bias estimator parameter β_2 .



Gambar 4.5 Bias estimator parameter γ .

Tabel 4.3 Hasil estimasi parameter regresi

| σ_{δ}^2 | Model | Parameter | | | | |
|---------------------|-----------|---------------|--------------|---------------|------------------|------------|
| | | $\alpha_1=-1$ | $\alpha_2=1$ | $\beta_1=0.5$ | $\beta_2 = -0.5$ | $\gamma=1$ |
| 0.1 | Logit | -1.0820397 | 0.9570063 | 0.5085003 | -0.4049338 | 0.9157006 |
| | Mix-Logit | -1.0820500 | 0.9569940 | 0.5085772 | -0.4049219 | 0.9157116 |
| 0.2 | Logit | -1.1891563 | 0.8377265 | 0.5144175 | -0.4031753 | 0.8900812 |
| | Mix-Logit | -1.1485085 | 0.9065620 | 0.5126635 | -0.4139295 | 0.9223453 |
| 0.3 | Logit | -1.1263546 | 0.8719698 | 0.4604642 | -0.6273847 | 0.9394594 |
| | Mix-Logit | -1.0429815 | 1.0155655 | 0.4550223 | -0.6700660 | 1.0129981 |
| 0.5 | Logit | -1.1331315 | 1.1809890 | 0.5115591 | -0.5218364 | 1.0460728 |
| | Mix-Logit | -1.0209141 | 1.4069878 | 0.5232962 | -0.5684692 | 1.1588799 |
| 1 | Logit | -1.0910121 | 0.9232880 | 0.7293906 | -0.4161242 | 0.9296585 |
| | Mix-Logit | -1.0910121 | 1.1353897 | 0.7482258 | -0.4485884 | 1.0380471 |
| 2 | Logit | -1.2340452 | 0.7217710 | 0.5295068 | -0.5595222 | 0.8490157 |
| | Mix-Logit | -1.0683104 | 1.0311304 | 0.5205118 | -0.6526206 | 1.0197001 |
| 3 | Logit | -1.2631787 | 0.8102169 | 0.5518312 | -0.4519676 | 0.8109018 |
| | Mix-Logit | -0.9989709 | 1.2886179 | 0.5511986 | -0.5643533 | 1.0507553 |
| 4 | Logit | -1.1450981 | 0.4277022 | 0.4145928 | -0.3496570 | 0.6667977 |
| | Mix-Logit | -0.8751249 | 0.8517508 | 0.3952151 | -0.4471058 | 0.9045478 |
| 5 | Logit | -1.2657586 | 0.5095037 | 0.6658029 | -0.3498746 | 0.6481598 |
| | Mix-Logit | -0.8853560 | 1.0890588 | 0.6799238 | -0.4472538 | 0.9379534 |
| 6 | Logit | -1.2870457 | 0.4817746 | 0.6701468 | -0.3375699 | 0.6366787 |
| | Mix-Logit | -0.9082190 | 1.0615113 | 0.6887330 | -0.4302786 | 0.9316741 |
| 7 | Logit | -1.3546923 | 0.4441560 | 0.6758602 | -0.2650508 | 0.6899177 |
| | Mix-Logit | -1.0396803 | 0.9896029 | 0.7699706 | -0.3407205 | 1.0195644 |
| 8 | Logit | -1.4147482 | 0.3654857 | 0.6869802 | -0.2539744 | 0.6401623 |
| | Mix-Logit | -1.0778791 | 0.9499069 | 0.7901127 | -0.3323799 | 1.0044416 |
| 9 | Logit | -1.2350797 | 0.3339042 | 0.3979181 | -0.3673812 | 0.5974748 |
| | Mix-Logit | -0.8427982 | 0.9562321 | 0.3215740 | -0.5592697 | 0.9444958 |
| 10 | Logit | -1.3082698 | 0.4102022 | 0.6433415 | -0.3366008 | 0.5924132 |
| | Mix-Logit | -0.8408041 | 1.1169468 | 0.6475197 | -0.4638984 | 0.9451848 |
| 11 | Logit | -1.3042875 | 0.1833946 | 0.5112466 | -0.2755903 | 0.6705004 |
| | Mix-Logit | -1.0766507 | 0.7143608 | 0.5911738 | -0.3893042 | 1.0980798 |



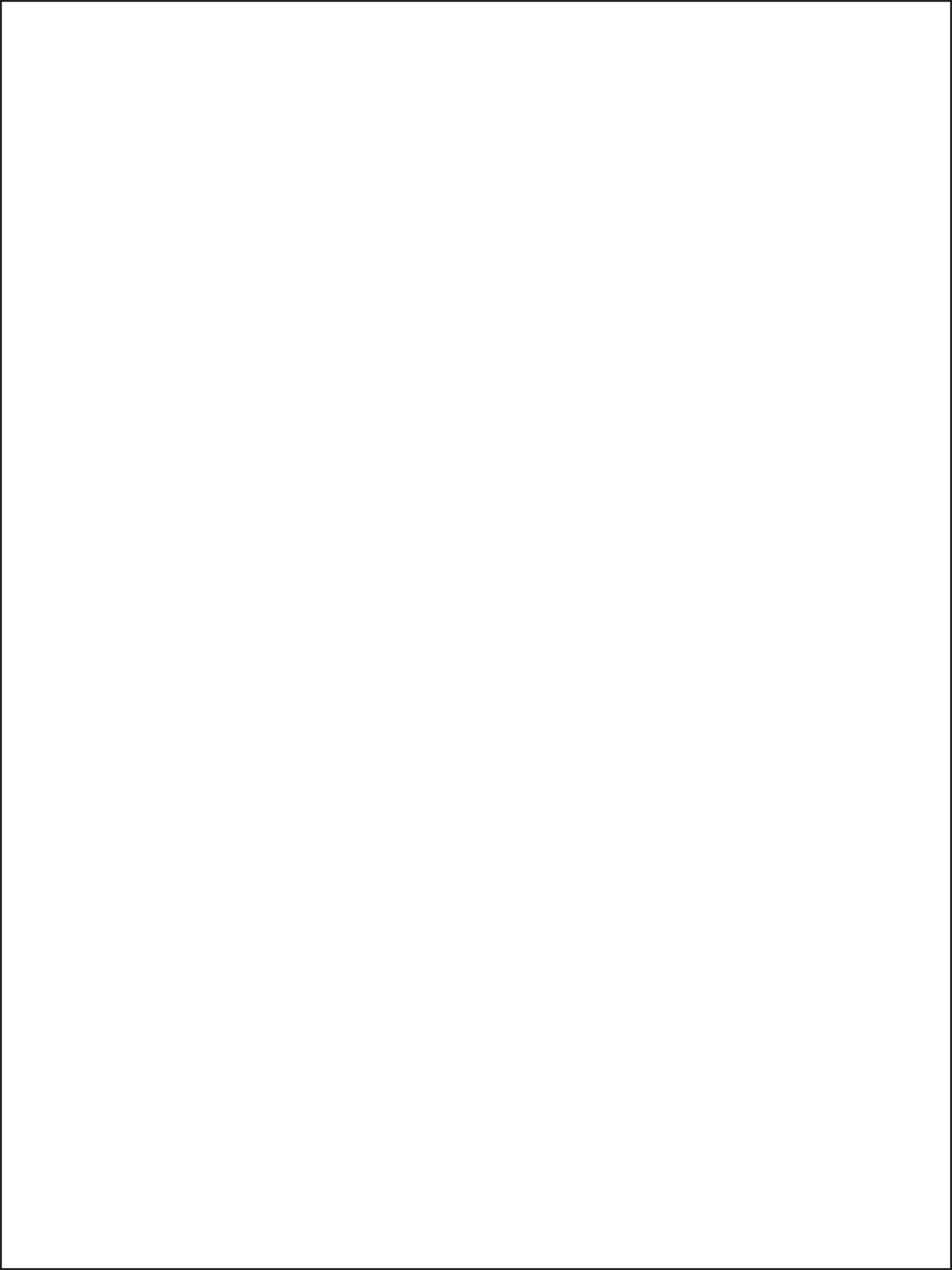
| | | | | | | |
|----|-----------|------------|-----------|-----------|------------|-----------|
| 12 | Logit | -1.3072919 | 0.3837021 | 0.6531657 | -0.3160782 | 0.5738949 |
| | Mix-Logit | -0.8012720 | 1.1337395 | 0.6698818 | -0.4337996 | 0.9406709 |
| 13 | Logit | -1.3058442 | 0.1704799 | 0.5016896 | -0.2939947 | 0.6591141 |
| | Mix-Logit | -1.0667426 | 0.7203540 | 0.5653857 | -0.4308860 | 1.1007823 |
| 14 | Logit | -1.4427927 | 0.2428831 | 0.3819696 | -0.2654855 | 0.5866355 |
| | Mix-Logit | -1.0758214 | 0.9594194 | 0.2856160 | -0.4434660 | 1.0734814 |
| 15 | Logit | -1.3749713 | 0.3419744 | 0.4191620 | -0.3464587 | 0.6159980 |
| | Mix-Logit | -0.8151681 | 1.3491172 | 0.3329803 | -0.6779866 | 1.1997187 |

Berdasarkan simulasi data untuk sampel terbatas ($n=1000$) dan jumlah alternatif 3 dapat disimpulkan bahwa

1. Model *Mixed Logit* dapat mengestimasi parameter korelasi dengan baik (nilai bias kecil).
2. Model *Mixed Logit* lebih baik dibanding model MNL, khususnya ketika terdapat korelasi antar alternatif.

Selanjutnya dari hasil penelitian ini dapat disarankan

- a. Bagi para praktisi yang akan menggunakan pemodelan respon diskrit sebaiknya (Jika diduga ada korelasi antar alternatif, sebaiknya menggunakan model *Mixed Logit*. Jika diasumsikan tidak ada korelasi antar alternatif, maka dapat digunakan model MNL.
- b. Bagi para Statistisi (peneliti Statistika) dapat mengembangkan metode komputasi, mengingat proses perhitungan dalam model *Mixed Logit* lebih lama dibanding model MNL. Di samping itu dapat dikembangkan metode *Estimasi* selain MLE.



Bab 5

Sifat-Sifat Penaksir Maximum *Likelihood* dan *Simulated MLE* pada Respon Multinomial

Misalkan Y_1, Y_2, \dots, Y_n adalah variabel *random* yang mempunyai densitas gabungan

$$f(y; \theta) = f(y_1, y_2, \dots, y_n; \theta_1, \dots, \theta_j)$$

Fungsi ini tergantung pada parameter $\theta = (\theta_1, \dots, \theta_j)$. Jika Y_i saling independen maka

$$f(y; \theta) = \prod_{i=1}^n f(y_i, \theta) \quad (5.1)$$

Fungsi *Likelihood* $L(\theta; y)$ adalah sama dengan $f(y; \theta)$ yang dipandang sebagai fungsi dari θ untuk suatu nilai y (Lehmann, 1998),

$$L(\theta; y) = f(y; \theta).$$

Oleh karena y_1, \dots, y_n saling independen, maka dapat dinyatakan sebagai fungsi *Log-likelihood*

$$\log L(\theta; y) = LL(\theta; y) = \sum_{i=1}^n \log f(y_i; \theta) \quad (5.2)$$

5.1 Penaksir *maximum likelihood*

Misalkan Ω merupakan himpunan semua nilai yang mungkin untuk vektor parameter θ (Ω disebut juga ruang parameter). MLE untuk θ , yang dinotasikan dengan $\hat{\theta}_{MLE}$ adalah nilai θ yang memaksimumkan fungsi *Likelihood* $L(\theta; y)$ pada data y ,

$$L(\hat{\theta}_{MLE}; y) = \max[L(\theta; y) \mid \theta \in \Omega].$$

atau cukup dituliskan dengan

$$\hat{\theta}_{MLE} = \arg \max_{\theta} L(\theta; y) \quad (5.3)$$

sehingga

$$L(\hat{\theta}; y) \geq L(\theta; y) \text{ untuk } \forall \theta \in \Omega$$

Karena fungsi logaritma adalah monoton, maka

$$\log L(\hat{\theta}; y) \geq \log L(\theta; y) \text{ untuk } \forall \theta \in \Omega$$

Untuk mendapatkan $\hat{\theta}$, diperoleh dengan cara mendifferensialkan fungsi log-likelihoodnya terhadap setiap elemen $\theta_j \in \theta$ dan menyelesaikan secara simultan persamaan *Likelihood*. Derivatif pertama fungsi *Log-likelihood* disebut vektor *score*,

$$\mathbf{g}(\theta) = \frac{\partial \log L(\theta; \mathbf{y})}{\partial \theta} = \sum_{i=1}^n \frac{\partial \log f(y_i; \theta)}{\partial \theta} = \sum_{i=1}^n \mathbf{g}_i(\theta) \quad (5.4)$$

$\hat{\theta}_{MLE}$ merupakan penyelesaian dari persamaan

$$\mathbf{g}(\theta) = \mathbf{0}$$

Turunan pertama fungsi *Log-likelihood* terhadap masing-masing parameter dapat dinyatakan sebagai

$$\frac{\partial \log L(\theta; y)}{\partial \theta_j} = 0 \text{ untuk } j=1,2,\dots,J \quad (5.5)$$

Matrik Hessian (\mathbf{H}), yaitu matrik yang elemen-elemennya merupakan derivatif ke dua dari fungsi log-likelihood adalah matrik definit negatif.

$$\mathbf{H}(\theta) = \frac{\partial^2 \log L(\theta; \mathbf{y})}{\partial \theta' \partial \theta} = \sum_{i=1}^n \frac{\partial^2 \log f(y_i; \theta)}{\partial \theta' \partial \theta} = \sum_{i=1}^n \mathbf{H}_i(\theta) \quad (5.6)$$

Elemen ke (j,k) matriks Hessian adalah

$$\frac{\partial^2 \log L(\theta; \mathbf{y})}{\partial \theta_j \partial \theta_k} \text{ untuk } j,k=1,2,\dots,J.$$

Selanjutnya $\mathbf{Inf}(\theta) = E[-\mathbf{H}(\theta)]$ dinamakan informasi *Fisher*. Jika terdapat beberapa nilai θ yang merupakan titik maksimum lokal untuk $\log L(\theta|\mathbf{y})$, maka $\hat{\theta}_{MLE}$ merupakan nilai θ yang mempunyai nilai $\log L(\theta|\mathbf{y})$ tertinggi di antara titik maksimum lokal tersebut.

Definisi 5.1. *Estimable* (Greene, 2003)

Vektor parameter θ dikatakan teridentifikasi atau *estimable* jika untuk suatu parameter $\theta^* \neq \theta$ pada data y maka $L(\theta^*|y) \neq L(\theta|y)$.

MLE merupakan penaksir yang menarik pada sampel besar (sifat asimtotis). Sifat MLE yang lain adalah konsisten dan asimtotis efisien.

Definisi 5.2. Konsisten (Bain dan Engelhardt, 1992)

Penaksir $\hat{\theta}_n$ yang merupakan barisan penaksir untuk parameter θ disebut penaksir konsisten jika untuk semua $\epsilon > 0$

$$\lim_{n \rightarrow \infty} P(\|\hat{\theta}_n - \theta\| < \epsilon) = 1.$$

Definisi 5.3. Efisien secara asimtotis (Lehmann dan Cassela, 1998)

Misalkan sebuah penaksir $\sqrt{n}(\hat{\theta} - \theta)$ berdistribusi normal asimtotis dengan variansi $u(\theta)$. Jika $u(\theta) = [I_n(\theta)]^{-1}$ maka $\hat{\theta}$ merupakan penaksir yang efisien secara asimtotis.

Misalkan Y sebagai variabel *random* dengan fungsi densitasnya (fungsi peluang) $f(\cdot, \theta)$, $\theta \in \Omega \subseteq \mathbb{R}^p$ (real), yang dimaksud syarat-syarat keraturan (*regularity conditions*) adalah $f(\mathbf{y}; \theta) > 0$ dan y independen terhadap θ

$$\frac{\partial f(\mathbf{y}; \theta)}{\partial \theta} \text{ ada untuk setiap } \theta \in \Omega \text{ dan } \mathbf{y}$$

Ω adalah interval terbuka dalam \mathbb{R}^p

$I_n(\theta)$ merupakan matrik definit positif untuk setiap $\theta \in \Omega$

$$\frac{\partial}{\partial \theta} \int \dots \int f(\mathbf{y} | \theta) dy_1 \dots dy_n \text{ dan } \frac{\partial}{\partial \theta} \int \dots \int h(\mathbf{y}) f(\mathbf{y} | \theta) dy_1 \dots dy_n \text{ ada untuk } \forall \theta \in \Omega \text{ dan } h(\mathbf{y}).$$

Syarat-syarat ini menjamin bahwa pengintegralan dan pendiferensialan dapat dipertukarkan, derivatif ada dan dapat diintegalkan.

Definisi 5.4. Konvergen dalam peluang (Dudewics dan Mishra, 1988)

Z_n dikatakan konvergen ke Z dalam peluang ($Z_n \xrightarrow{p} Z$) jika untuk setiap $\epsilon > 0$,

$$P(\lim_{n \rightarrow \infty} \|Z_n - Z\| \geq \epsilon) = 0 \text{ atau } P(\lim_{n \rightarrow \infty} \|Z_n - Z\| < \epsilon) = 1$$

yang disebut pula konvergen stokastik, atau konvergen dalam ukuran, atau konvergen lemah.

Teorema Khinchine (Dudewics dan Mishra, 1988)

Jika Y_1, \dots, Y_n adalah variabel *random* yang saling independen dan berdistribusi identik dan $E(Y_i)$ ada (misal $E(Y_i) = a$) maka

$$\frac{Y_1 + \dots + Y_n}{n} \xrightarrow{p} a.$$

Definisi 5.5. Konvergen dalam distribusi (Dudewics dan Mishra, 1988)

Z_n dikatakan konvergen ke Z dalam distribusi ($Z_n \xrightarrow{d} Z$) jika,

$$\lim_{n \rightarrow \infty} F_{Z_n}(t) = F_Z(t)$$

pada setiap titik t yang $F_Z(\cdot)$ kontinu. Dapat dikatakan $F_n(t) \cong F_{Z_n}(t)$ konvergen lemah ke $F(t) \cong F_Z(t)$.

Teorema limit pusat (Dudewics dan Mishra, 1988)

Suatu barisan distribusi gabungan variabel *random* Y_1, \dots, Y_n dengan mean dan variansi berhingga dikatakan memenuhi teorema limit pusat jika dan hanya jika Z_n , yang didefinisikan sebagai

$$Z_n = \frac{S_n - E(S_n)}{\sqrt{Var(S_n)}} \text{ dengan } S_n = Y_1 + \dots + Y_n$$

adalah konvergen dalam distribusi ke suatu variabel *random* berdistribusi normal dengan mean 0 dan variansi 1.

Jika Y_1, \dots, Y_n variabel *random* saling independen yang berdistribusi identik dengan $E(Y_i) = \mu$ dan $Var(Y_i) = \sigma^2 > 0$ (keduanya berhingga). Untuk semua $z, -\infty < z < \infty$ dan $n \rightarrow \infty$ maka

$$P\left[\frac{(Y_1 - \mu) + \dots + (Y_n - \mu)}{\sigma\sqrt{n}} < z\right] \rightarrow \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z \exp\left(-\frac{1}{2}y^2\right) dy.$$

Teorema Limit Pusat juga disebut teorema *Lindberg-Levy*, yang dapat juga dinyatakan sebagai

$$\frac{\sqrt{n}(\bar{Y}_n - \mu)}{\sigma} \xrightarrow{d} N(0,1) \cdot W$$

Teorema multivariate central limit theorem (Anderson, 1984)

Jika vektor *random* Y_1, \dots, Y_n masing-masing berukuran $1 \times p$ ($Y_i = (Y_{i1}, \dots, Y_{ip})$) adalah saling independen dan berdistribusi identik dengan mean $E(Y_i) = \mu$ dan matrik kovariansi $E(Y_i - \mu)'(Y_i - \mu) = \Sigma$ maka

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n (Y_i - \mu) \xrightarrow{d} N(\mathbf{0}, \Sigma)$$

Teorema sifat konsisten (Pawitan, 2001)

Jika Y_1, \dots, Y_n variabel *random* saling independen yang berdistribusi identik dengan densitas $f(\mathbf{Y}_i, \theta)$ dan $f(\mathbf{Y}_i, \theta)$ merupakan fungsi kontinu dari θ , maka $\hat{\theta}_n$ yang merupakan penaksir MLE dari θ bersifat konsisten, yaitu

$$\lim_{n \rightarrow \infty} P(\|\hat{\theta}_n - \theta\| < \varepsilon) = 1$$

Teorema sifat normal asimtotis (Pawitan, 2001)

Jika Y_1, \dots, Y_n variabel *random* saling independen yang berdistribusi identik dengan densitas $f(\mathbf{Y}_i, \theta)$ dan diasumsikan bahwa $\hat{\theta}_n$ yang merupakan penaksir MLE dari θ bersifat konsisten, maka di bawah sifat keteraturan

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{d} N[0, (\mathbf{Inf}(\theta_0))^{-1}] \text{ dengan } \mathbf{Inf}(\theta_0) = -E_0 \left[\frac{\partial^2 \log L(\theta; \mathbf{y})}{\partial \theta_0 \partial \theta_0'} \right]$$

Definisi 5.6. Penaksir teratur

Misalkan Y_1, \dots, Y_n adalah variabel *random* berdistribusi identik independen dengan densitas $f(Y|\theta)$ dan $t(Y) = t(Y_1, \dots, Y_n)$ adalah sebarang penaksir dengan $E\{t(Y)\}$ terdiferensial dalam θ dan densitas bersama $f(\underline{y}|\theta) = f(y_1, \dots, y_n|\theta)$ memenuhi:

$$\frac{\partial}{\partial \theta} \int \dots \int h(\underline{y}) f(\underline{y}|\theta) d\underline{y} = \int \dots \int h(\underline{y}) \frac{\partial}{\partial \theta} f(\underline{y}|\theta) d\underline{y}$$

untuk setiap $h(\underline{y})$ dengan $E|h(\underline{y})| < \infty$, maka $t(Y)$ disebut penaksir teratur.

Pada setiap kasus penaksir teratur, andaikan Y_1, \dots, Y_n variabel *random* bebas masing-masing mempunyai fungsi densitas (fungsi $f(y|\theta)$). Andaikan $t_n(Y_1, \dots, Y_n)$ suatu penaksir takbias dari θ , maka $\text{Var}(t_n) \geq \frac{1}{nI(\theta)}$

Selanjutnya batas bawah ini dapat digunakan untuk menilai penaksir-penaksir tak bias dengan membandingkan variansinya terhadap batas bawah tersebut.

Teorema *cramer-rao lower bound*

Dengan asumsi densitas y_i memenuhi sifat keteraturan, varians asimtotik estimator yang CAN dari parameter θ_0 tidak lebih besar dari

$$[I(\theta_0)]^{-1} = \left(-E_0 \left[\frac{\partial^2 \ln L(\theta_0)}{\partial \theta_0 \partial \theta_0'} \right] \right)^{-1} = \left(E_0 \left[\left(\frac{\partial \ln L(\theta_0)}{\partial \theta_0} \right) \left(\frac{\partial \ln L(\theta_0)}{\partial \theta_0} \right)' \right] \right)^{-1}$$

Akan dibuktikan dulu dua suku terakhir. Bukti berdasarkan derivatif setiap elemennya, Apakah

$$E \left[\frac{\partial \ln L}{\partial \theta_j} \frac{\partial \ln L}{\partial \theta_k} \right] = E \left[- \frac{\partial^2 \ln L}{\partial \theta_j \partial \theta_k} \right]$$

Bukti:

Untuk suatu θ_j dan θ_k , maka

$$\int L(x; \theta) dx = 1 \text{ maka}$$

$$\frac{\partial}{\partial \theta_j} \int L(x; \theta) dx = \int \frac{\partial}{\partial \theta_j} L(x; \theta) dx = 0$$

$$\Leftrightarrow \int \frac{\partial \log L(x; \theta)}{\partial \theta_j} L(x; \theta) dx = 0$$

sehingga

$$\frac{\partial}{\partial \theta_k} \int \frac{\partial \log L(x; \theta)}{\partial \theta_j} L(x; \theta) dx = 0$$

$$\Leftrightarrow \int \frac{\partial^2 \log L(x; \theta)}{\partial \theta_j \partial \theta_k} L(x; \theta) dx + \int \frac{\partial \log L(x; \theta)}{\partial \theta_j} \frac{\partial L(x; \theta)}{\partial \theta_k} dx = 0$$

$$\Leftrightarrow \int \frac{\partial^2 \log L(x; \theta)}{\partial \theta_j \partial \theta_k} L(x; \theta) dx + \int \frac{\partial \log L(x; \theta)}{\partial \theta_j} \frac{\partial \log L(x; \theta)}{\partial \theta_k} L(x; \theta) dx = 0$$

$$\Leftrightarrow \int \frac{\partial \log L(x; \theta)}{\partial \theta_j} \frac{\partial \log L(x; \theta)}{\partial \theta_k} L(x; \theta) dx = - \int \frac{\partial^2 \log L(x; \theta)}{\partial \theta_j \partial \theta_k} L(x; \theta) dx$$

jadi,

$$E = \left[\frac{\partial \ln L}{\partial \theta_j} \frac{\partial \ln L}{\partial \theta_k} \right] = E \left[- \frac{\partial^2 \ln L}{\partial \theta_j \partial \theta_k} \right]$$

(terbukti)

Diketahui juga bahwa

$$\left[\frac{\partial \log L(x; \theta)}{\partial \theta_j} \right] = \int \frac{\partial \log L(x; \theta)}{\partial \theta_j} L(x; \theta) dx = 0 \quad \text{untuk setiap } j$$

Maka

$$\left[\frac{\partial \log L(x; \theta)}{\partial \theta} \right] = \int \frac{\partial \log L(x; \theta)}{\partial \theta} L(x; \theta) dx = 0$$

Misal $T = t(Y)$ adalah penaksir unbiased dari θ , maka

$$E \left[T \cdot \frac{\partial \log L(x; \theta)}{\partial \theta} \right] = \int T \cdot \frac{\partial \log L(x; \theta)}{\partial \theta} L(x; \theta) dx$$

$$\Leftrightarrow \int T \cdot \frac{\partial L(x; \theta)}{\partial \theta} dx = \frac{\partial}{\partial \theta} \int T L(x; \theta) dx$$

$$\Leftrightarrow \frac{\partial}{\partial \theta} E(T) = \frac{\partial \theta}{\partial \theta} = 1$$

Dan

$$\text{Cov} \left[T, \frac{\partial \log L(x; \theta)}{\partial \theta} \right]^2 \leq \text{Var}(T) \cdot \text{Var} \left[\frac{\partial \log L(x; \theta)}{\partial \theta} \right]$$

$$E \left[T, \frac{\partial \log L(x; \theta)}{\partial \theta} \right]^2 \leq \text{Var}(T) \cdot E \left[\left(\frac{\partial \log L(x; \theta)}{\partial \theta} \right) \left(\frac{\partial \log L(x; \theta)}{\partial \theta} \right)^t \right]$$

$$\left[E \left[\left(\frac{\partial \log L(x; \theta)}{\partial \theta} \right) \left(\frac{\partial \log L(x; \theta)}{\partial \theta} \right)^t \right] \right]^{-1} \leq \text{Var}(T) \quad \text{(terbukti)}$$

Bukti sifat konsisten

Untuk membuktikan sifat konsisten, akan digunakan teorema berikut:

Teorema Likelihood Inequality

$$E_0 \left[\frac{1}{n} \ln L(\theta) \right] < E_0 \left[\frac{1}{n} \ln L(\theta_0) \right] \text{ untuk } \theta \neq \theta_0 \text{ (termasuk } \hat{\theta} \text{)}$$

Bukti:

Misalkan $\hat{\theta}$ adalah MLE dari θ , maka untuk $\theta \neq \hat{\theta}$

$$h L(\hat{\theta}) \geq h L(\theta)$$

Ambil variabel *random* $L(\theta)/L(\theta_0)$. Karena fungsi log adalah konkaf maka dengan teorema *Jensen's Inequality* diperoleh

$$E_0 \left[h \frac{L(\theta)}{L(\theta_0)} \right] < h E_0 \left[\frac{L(\theta)}{L(\theta_0)} \right] \text{ dan } E_0 \left[\frac{L(\theta)}{L(\theta_0)} \right] = \int \left[\frac{L(\theta)}{L(\theta_0)} \right] L(\theta_0) dy = 1$$

Sehingga

$$E_0 \left[h \frac{L(\theta)}{L(\theta_0)} \right] = E_0 [h L(\theta)] - E_0 [h L(\theta_0)] < 0$$

$$E_0 \left[\frac{1}{n} \ln L(\theta) \right] - E_0 \left[\frac{1}{n} \ln L(\theta_0) \right] < 0. \quad \text{(terbukti)}$$

Pengali $\frac{1}{n}$ untuk menyatakan rata-rata.

Teorema Likelihood Inequality dengan kata lain menyatakan bahwa nilai harapan dari *log-likelihood* akan mencapai nilai maksimum pada nilai parameter yang sesungguhnya.

Untuk semua θ (termasuk $\hat{\theta}$), dan sampel *random* berukuran n

$$\frac{1}{n} \ln L(\theta) = \frac{1}{n} \sum_{i=1}^n \ln f(y_i | \theta)$$

Karena sampel adalah IID dengan syarat keteraturan, dengan menggunakan Teorema *Khinchine*, mean sampel konvergen dalam probabilitas ke mean populasi.

$$\lim_{n \rightarrow \infty} P\left(\left|\frac{1}{n} \ln L(\hat{\theta}) - \frac{1}{n} \ln L(\theta_0)\right| < \epsilon\right) = 1 \text{ untuk } \epsilon > 0$$

Menggunakan $\theta = \hat{\theta}$, berdasarkan teorema *Likelihood Inequality*, untuk $n \rightarrow \infty$

$$\lim \text{Prob}\left\{\left[\frac{1}{n} \ln L(\hat{\theta})\right] < \left[\frac{1}{n} \ln L(\theta_0)\right]\right\} = 1 \text{ jika } \hat{\theta} \neq \theta_0$$

Dilain sisi, karena $\hat{\theta}$ adalah MLE, sehingga untuk setiap n

$$\left[\frac{1}{n} \ln L(\hat{\theta})\right] \geq \left[\frac{1}{n} \ln L(\theta_0)\right]$$

Langkah selanjutnya adalah menunjukkan bahwa

$$\frac{1}{n} \ln L(\hat{\theta}) \rightarrow \frac{1}{n} \ln L(\theta_0) \text{ maka } \hat{\theta} \rightarrow \theta_0$$

Hal ini benar untuk satu parameter dan fungsi *likelihood* adalah fungsi satu-satu. Untuk kasus umum (vektor parameter), agar dipenuhi disyaratkan

1. fungsi *likelihood* kontinu dan terdiferensial dua kali.
2. memenuhi syarat keteraturan
3. parameter teridentifikasi

Bukti sifat normal asimtotis:

Misalkan $\hat{\theta}$ adalah MLE dari θ , maka gradiennya

$$g(\hat{\theta}) = 0$$

Dengan menggunakan deret *Taylor* orde dua untuk mendekati nilai θ_0

$$g(\hat{\theta}) = g(\theta_0) + H(\bar{\theta})(\hat{\theta} - \theta_0) = 0$$

Matriks *Hessian* terevaluasi pada nilai $\bar{\theta}$ yang bernilai antara $\hat{\theta}$ dan θ_0 ,

$$\bar{\theta} = w\hat{\theta} + (1-w)\theta_0 \text{ untuk } 0 < w < 1$$

dikalikan dengan \sqrt{n} diperoleh

$$\sqrt{n}(\hat{\theta} - \theta_0) = [-H(\bar{\theta})]^{-1}[\sqrt{n}g(\theta_0)]$$

Karena $\text{plim}(\hat{\theta} - \bar{\theta}) = 0$, $\text{plim}(\hat{\theta} - \theta_0) = 0$, dan derivatif kedua adalah fungsi kontinu, maka jika limit distribusi ada maka

$$\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{d} [-H(\bar{\theta})]^{-1}[\sqrt{n}g(\theta_0)]$$

$$\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{d} \left[-\frac{1}{n}H(\theta_0)\right]^{-1}[\sqrt{n}\bar{g}(\theta_0)]$$

Selanjutnya karena $E[\bar{g}(\theta_0)] = 0$ dan $\text{Var}[\sqrt{n}\bar{g}(\theta_0)] = E\left[-\frac{1}{n}H(\theta_0)\right]$ maka menggunakan teorema limit pusat *Lindberg-Levy* diperoleh

$$\sqrt{n}\bar{g}(\theta_0) \xrightarrow{d} N\left\{0, E\left[\frac{1}{n}H(\theta_0)\right]\right\}$$

Karena $\text{plim}\left[-\frac{1}{n}H(\theta_0)\right] = -E\left[\frac{1}{n}H(\theta_0)\right]$

$$\left[\frac{1}{n}H(\theta_0)\right]^{-1}\sqrt{n}\bar{g}(\theta_0) \xrightarrow{d} N\left\{0, \left\{-E\left[\frac{1}{n}H(\theta_0)\right]\right\}^{-1}\left\{-E\left[\frac{1}{n}H(\theta_0)\right]\right\}\right\}$$

atau

$$\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{d} N\left[0, \left\{-E\left[\frac{1}{n}H(\theta_0)\right]\right\}^{-1}\right]$$

sehingga

$$\hat{\theta} \xrightarrow{a} N[\theta_0, \{I(\theta_0)\}^{-1}]$$

5.2 MLE pada distribusi multinomial

Jika Y_1, Y_2, \dots, Y_n adalah sampel *random* yang mempunyai densitas Multinomial, maka

$$f(y_i; \pi_{i1}, \dots, \pi_{ij}) = \pi_{i1}^{y_{i1}} \dots \pi_{ij}^{y_{ij}} \text{ untuk } i=1, \dots, n$$

dan densitas gabungannya adalah

$$f(y; \pi) = \prod_{i=1}^n \pi_{i1}^{y_{i1}} \dots \pi_{ij}^{y_{ij}} \text{ dan } \sum_{j=1}^J \pi_j = 1.$$

π_{ij} memuat parameter θ sehingga

$$\pi_{ij} = \pi_{ij}(\theta).$$

$y_{ij} = 1$ jika responden i memilih alternatif j ($y_i = j$) dan $y_{ij} = 0$ jika responden i memilih alternatif selain j ($y_i \neq j$). Dapat disusun fungsi *Likelihood* untuk parameter θ ,

$$L(\theta) = \prod_{i=1}^n \pi_{i1}^{y_{i1}} \dots \pi_{ij}^{y_{ij}}. \quad (5.7)$$

Fungsi *Log-likelihood*nya adalah

$$\log L(\theta) = \sum_{i=1}^n \log L_i(\theta) = \sum_{i=1}^n \log(\pi_{i1}^{y_{i1}} \dots \pi_{ij}^{y_{ij}}) = \sum_{i=1}^n \sum_{j=1}^J y_{ij} \log[\pi_{ij}] \quad (5.8)$$

MLE adalah nilai θ yang memaksimumkan fungsi $\log L(\theta)$, atau

$$\hat{\theta}_{ML} = \operatorname{argmax}_{\theta} \log L(\theta) = \operatorname{argmax}_{\theta} \sum_{i=1}^n \sum_{j=1}^J y_{ij} \log[\pi_{ij}]. \quad (5.9)$$

Derivatif pertama fungsi *log-likelihood* (5.8) adalah

$$\frac{\partial \log L(\theta)}{\partial \theta} = \sum_{i=1}^n \frac{\partial \log L_i(\theta)}{\partial \theta} = \sum_{i=1}^n \sum_{j=1}^J y_{ij} \frac{1}{\pi_{ij}} \frac{\partial \pi_{ij}}{\partial \theta}. \quad (5.10)$$

Pada titik maksimum, gradien bernilai nol sehingga

$$\sum_{i=1}^n \sum_{j=1}^J y_{ij} \frac{1}{\pi_j} \frac{\partial \pi_j}{\partial \theta} = 0. \quad (5.11)$$

Derivatif ke dua dari fungsi *Log-likelihood* (matriks Hessian) adalah

$$\frac{\partial^2 \log L(\theta)}{\partial \theta' \partial \theta} = \frac{\partial}{\partial \theta'} \sum_{i=1}^n \sum_{j=1}^J y_{ij} \frac{1}{\pi_j} \frac{\partial \pi_j}{\partial \theta} \quad (5.12)$$

$$= \sum_{i=1}^n \sum_{j=1}^J y_{ij} \left(-\frac{1}{\pi_j^2} \frac{\partial \pi_j}{\partial \theta'} \frac{\partial \pi_j}{\partial \theta} + \frac{1}{\pi_j} \frac{\partial^2 \pi_j}{\partial \theta' \partial \theta} \right)$$

Sehingga

$$H(\theta) = \sum_{i=1}^n \sum_{j=1}^J y_{ij} \left(\frac{\partial^2 \pi_j}{\partial \theta' \partial \theta} - \frac{1}{\pi_j} \frac{\partial \pi_j}{\partial \theta'} \frac{\partial \pi_j}{\partial \theta} \right) \quad (5.13)$$

Pada model MNL yang didasarkan pada Teorema 2.1,

$$\pi_k = \frac{\exp(V_k)}{\sum_{j=1}^J \exp(V_j)} = \frac{\exp(\theta' W_k)}{\sum_{j=1}^J \exp(\theta' W_j)} \quad (5.14)$$

dengan $\theta_j = (\alpha_j, \beta_j, \gamma)'$, $W_{ij} = (1, X_i, Z_{ij})'$ dan $i=1, \dots, n$. Dari model MNL tersebut, diperoleh

$$\frac{\partial \pi_k}{\partial \theta_k} = W_k \pi_k (1 - \pi_k) \quad \text{dan} \quad \frac{\partial \pi_k}{\partial \theta_l} = -\pi_k \pi_l W_l \quad \text{untuk } l \neq k.$$

$$\frac{\partial \pi_j}{\partial \theta} = \left(\frac{\partial \pi_j}{\partial \theta_1}, \dots, \frac{\partial \pi_j}{\partial \theta_J} \right).$$

Selanjutnya,

$$\begin{aligned} \frac{\partial \log L(\theta)}{\partial \theta} &= \sum_{i=1}^n \sum_{j=1}^J y_{ij} \frac{1}{\pi_j} \left(\frac{\partial \pi_j}{\partial \theta_1}, \dots, \frac{\partial \pi_j}{\partial \theta_J} \right) \\ &= \sum_{i=1}^n \left(\left(y_{i1} \frac{1}{\pi_{i1}} W_{i1} \pi_{i1} (1 - \pi_{i1}) - y_{i2} \frac{1}{\pi_{i2}} \pi_{i1} \pi_{i2} W_{i1} - \dots - y_{ij} \frac{1}{\pi_{ij}} \pi_{i1} \pi_{ij} W_{i1} \right), \dots, \right. \\ &\quad \left. \left(-y_{i1} \frac{1}{\pi_{i1}} \pi_{ij} \pi_{i1} W_{ij} - y_{i2} \frac{1}{\pi_{i2}} \pi_{ij} \pi_{i2} W_{ij} - \dots + y_{ij} \frac{1}{\pi_{ij}} W_{ij} \pi_{ij} (1 - \pi_{ij}) \right) \right) \\ &= \sum_{i=1}^n \left((y_{i1} - \pi_{i1} (y_{i1} + y_{i2} + \dots + y_{ij})) W_{i1}, \dots, (y_{ij} - \pi_{ij} (y_{i1} + y_{i2} + \dots + y_{ij})) W_{ij} \right) \\ &= \sum_{i=1}^n \left((y_{i1} - \pi_{i1}) W_{i1}, \dots, (y_{ij} - \pi_{ij}) W_{ij} \right). \end{aligned}$$



Oleh karena itu, jika $\frac{\partial \log L(\theta)}{\partial \theta} = \mathbf{0}$ maka $\sum_{i=1}^n (y_{ij} - \pi_{ij}) \mathbf{W}_j = \mathbf{0}$ untuk $j=1, \dots, J$.

Paramter π_{ij} yang memuat parameter θ merupakan parameter dari distribusi Multinomial. Beberapa model yang dapat digunakan untuk menduga nilai π_{ij} antara lain model *Logit Multinomial*, model *Nested Logit*, model *Probit*.

5.2.1 Model logit multinomial

Pada model *Logit Multinomial*, π_{ij} dirumuskan sebagai berikut

$$\pi_{ij} = P(Y_i = j | x_i) = \frac{\exp(\beta_j^t x_i)}{\sum_{k=1}^J \exp(\beta_k^t x_i)} \text{ untuk } j=1, 2, \dots, J$$

Derivatif pertama dan kedua atas fungsi *log-likelihood*nya adalah

$$\frac{\partial \log L(\beta)}{\partial \beta_j} = \sum_{i=1}^n (y_{ij} - P_j) x_i$$

$$\frac{\partial^2 \log L(\beta)}{\partial \beta_j \partial \beta_l^t} = - \sum_{i=1}^n P_j (1(j=l) - P_j) x_i x_i^t$$

Jika $\bar{x}_i = \sum_{j=1}^J \pi_j x_j$ maka

$$\frac{\partial \log L(\beta)}{\partial \beta} = \sum_{i=1}^n \sum_{j=1}^J y_{ij} (x_j - \bar{x}_i)$$

$$\frac{\partial^2 \log L(\beta)}{\partial \beta \partial \beta^t} = - \sum_{i=1}^n \sum_{j=1}^J P_j (x_j - \bar{x}_i) (x_j - \bar{x}_i)^t$$

5.2.2 Model nested logit

Pada model *Logit Multinomial*, π_{ij} dirumuskan sebagai berikut

$$\pi_{ij} = \pi_{j|l} \pi_l$$

dengan

$$\pi_{j|l} = \frac{\exp(\beta'_{j|l} x_i)}{\sum_{j=1}^{J_l} \exp(\beta'_{j|l} x_i)}; \pi_l = \frac{\exp(\gamma'_l x_i + \tau_l I_l)}{\sum_{l=1}^L \exp(\gamma'_l x_i + \tau_l I_l)}; I_l = \mathbf{h} \sum_{j=1}^{J_l} \exp(\beta'_{j|l} x_i);$$

$$\sigma_j^2 = \frac{\pi^2}{6\tau_j} \mathbf{61}$$

5.2.3 Model probit

Pada model MNP, diasumsikan bahwa vektor $\varepsilon_i' = (\varepsilon_{i1}, \dots, \varepsilon_{iJ})$ berdistribusi multivariat normal dengan mean nol dan matrik kovariansi Σ . Densitas untuk ε_i adalah

$$\phi(\varepsilon_i) = \frac{1}{(2\pi)^{J/2} |\Sigma|^{1/2}} \exp\left[-\frac{1}{2} \varepsilon_i' \Sigma^{-1} \varepsilon_i\right]$$

Probabilitas pembuat keputusan i memilih alternatif k adalah

$$\begin{aligned} p_{ik} &= P(V_{ik} + \varepsilon_{ik} > V_{ij} + \varepsilon_{ij}) && \forall j \neq k \\ &= \int I(V_{ik} + \varepsilon_{ik} > V_{ij} + \varepsilon_{ij}) \phi(\varepsilon_i) d\varepsilon_i && \forall j \neq k \end{aligned}$$

Karena hanya berbeda dalam bentuk utiliti, maka probabilitas pilihan dapat dinyatakan sebagai integral berdimensi $J-1$ atas selisih di antara *error*-nya. Misal diambil selisih terhadap alternatif k , maka dapat didefinisikan $\tilde{U}_{ijk} = U_j - U_k$, $\tilde{V}_{ijk} = V_j - V_k$ dan $\tilde{\varepsilon}_{ijk} = \varepsilon_j - \varepsilon_k$.

$$\pi_k = P(\tilde{U}_{ijk} < 0) \quad \forall j \neq k$$

$$\phi(\tilde{\varepsilon}_k) = \frac{1}{(2\pi)^{(J-1)/2} |\tilde{\Sigma}_k|^{1/2}} \exp\left[-\frac{1}{2} \tilde{\varepsilon}_k' \tilde{\Sigma}_k^{-1} \tilde{\varepsilon}_k\right]$$

Di mana $\tilde{\Sigma}_k$ adalah *covarians* dari $\tilde{\varepsilon}_k$, yang diturunkan dari Σ . Selanjutnya, probabilitas pilihan dapat disajikan dalam selisih utiliti,

$$\pi_k = \int I(\tilde{V}_{ijk} + \tilde{\varepsilon}_{ijk} < 0) \phi(\tilde{\varepsilon}_k) d\tilde{\varepsilon}_k \quad \forall j \neq k$$

yang merupakan integral berdimensi $(J-1)$ atas semua nilai yang mungkin dari selisih *error*.

5.3 Maximum simulated likelihood estimator (MSLE)

Pada model *Probit* dan *Mixed Logit*, π_{ij} merupakan integral rangkap yang tidak dapat dihitung secara analitik. Misalkan $\tilde{\pi}_{ij}$ adalah nilai π_{ij} yang dihitung menggunakan metode simulasi. Fungsi *Simulated Likelihood* diperoleh dengan cara mensubstitusi nilai π_{ij} dalam fungsi *Log-likelihood* dengan nilai simulasi $\tilde{\pi}_{ij}$. Sehingga fungsi *Simulated Log-likelihood* adalah

$$\text{simlog}L(\boldsymbol{\theta}) = \sum_{i=1}^n \text{simlog}L_i(\boldsymbol{\theta}) = \sum_{i=1}^n \sum_{j=1}^J y_j \log[\tilde{\pi}_j]. \quad (5.15)$$

$\hat{\boldsymbol{\theta}}_{MSL}$ adalah nilai $\boldsymbol{\theta}$ yang memaksimalkan $\text{simlog}L(\boldsymbol{\theta})$.

$$\hat{\boldsymbol{\theta}}_{MSL} = \arg \max_{\boldsymbol{\theta}} \text{simlog}L(\boldsymbol{\theta}) = \arg \max_{\boldsymbol{\theta}} \sum_{i=1}^n \sum_{j=1}^J y_j \log[\tilde{\pi}_j]. \quad (5.16)$$

Derivatif pertama fungsi simulated log-likelihoodnya adalah

$$\frac{\partial}{\partial \boldsymbol{\theta}} \text{simlog}L(\boldsymbol{\theta}) = \sum_{i=1}^n \sum_{j=1}^J \frac{y_j}{\tilde{\pi}_j} \frac{\partial \tilde{\pi}_j}{\partial \boldsymbol{\theta}} = \sum_{i=1}^n \sum_{j=1}^J [y_j - \tilde{\pi}_j] \frac{1}{\tilde{\pi}_j} \frac{\partial \tilde{\pi}_j}{\partial \boldsymbol{\theta}}. \quad (5.17)$$

Sifat-sifat MSLE telah dibahas oleh Gourieroux dan Monfort (1993), Lee (1992, 1995).

Teorema distribusi MSLE (Gourieroux dan Monfort, 1993)

Diketahui, sampel *random* y_1, \dots, y_n mempunyai densitas bersyarat $f(\mathbf{y}|\mathbf{x}, \boldsymbol{\theta})$ memenuhi sifat keteraturan sedemikian hingga MLE-nya adalah konsisten dan normal asimtotis. Jika densitas $f(y_i|x_i, \boldsymbol{\theta})$ diestimasi menggunakan simulator $\tilde{f}(y_i | x_i, \boldsymbol{\theta})$ yang tak bias, jika banyaknya sampel (n), banyaknya pengambilan sampel dalam simulasi (R) cukup besar ($R, n \rightarrow \infty$) dan $\frac{\sqrt{n}}{R} \rightarrow \infty$, maka MSLE mempunyai sifat asimtotis seperti MLE yaitu

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_{MSLE} - \boldsymbol{\theta}_0) \xrightarrow{d} N[0, (\mathbf{Inf}(\boldsymbol{\theta}_0))^{-1}].$$

Bukti

Misal $\check{g}_i(\boldsymbol{\theta})$ merupakan nilai simulasi dari $g_i(\boldsymbol{\theta})$. $\check{\mathbf{g}}(\boldsymbol{\theta})$ adalah mean sampel $\check{g}_i(\boldsymbol{\theta})$ sedemikian hingga $\check{\mathbf{g}}(\boldsymbol{\theta})$ adalah nilai simulasi dari $\mathbf{g}(\boldsymbol{\theta})$. Setiap individu i dilakukan pengambilan sampel sebanyak R kali. Karena dilakukan pengambilan sampel sebanyak R , maka kemungkinan menghasilkan $\check{\mathbf{g}}(\boldsymbol{\theta}_1) \neq \check{\mathbf{g}}(\boldsymbol{\theta}_2)$. Penaksir didefinisikan dengan syarat $\check{\mathbf{g}}(\hat{\boldsymbol{\theta}}) = 0$. Selanjutnya sifat-sifat estimator $\hat{\boldsymbol{\theta}}$ dapat diturunkan melalui dua tahap seperti dalam penaksir tradisional.

Tahap 1: Menentukan distribusi $\check{g}(\theta^*)$

θ^* merupakan nilai parameter sesungguhnya. Untuk mengidentifikasi komponen-komponen $\check{g}(\theta^*)$ dapat disusun persamaan

$$\begin{aligned}\check{g}(\theta^*) &= \check{g}(\theta^*) + g(\theta^*) - g(\theta^*) + E_r \check{g}(\theta^*) - E_r \check{g}(\theta^*) \\ &= \bar{g}(\theta^*) + [E_r \check{g}(\theta^*) - \bar{g}(\theta^*)] + [\check{g}(\theta^*) - E_r \check{g}(\theta^*)]\end{aligned}$$

$\bar{g}(\theta^*)$ adalah mean sampel yang sama dengan penaksir tradisional, yaitu

$$\sqrt{n}\bar{g}(\theta_0) \xrightarrow{d} N\{0, E_0[\frac{1}{n} H(\theta_0)]\}$$

Suku ke dua, $[E_r \check{g}(\theta^*) - \bar{g}(\theta^*)]$ merupakan besarnya bias antara nilai simulasi dengan nilai sesungguhnya. Jika tidak ada perbedaan antara $\bar{g}(\theta^*)$ dengan nilai harapan simulasi maka

$$E_r \check{g}(\theta^*) = \bar{g}(\theta^*)$$

Suku ketiga, $[\check{g}(\theta^*) - E_r \check{g}(\theta^*)]$ merupakan simulasi *noise*.

Selanjutnya dapat disusun

$$\check{g}(\theta) = A + B + C$$

dan perlu menguji sifat bias simulasi B dan simulasi *noise* C.

$$\begin{aligned}C &= [\check{g}(\theta^*) - E_r \check{g}(\theta^*)] \\ &= \frac{1}{n} \sum_{i=1}^n (\check{g}_i(\theta^*) - E_r \check{g}_i(\theta^*)) = \frac{1}{n} \sum_{i=1}^n d_i\end{aligned}$$

nilai d_i menyatakan deviasi nilai simulasi terhadap nilai harapannya. Mean C adalah null dengan variansi untuk setiap i adalah

$$\text{Var}(d_i) = S_i/R$$

S_i adalah variansi untuk i (dari R pengambilan sampel simulasi). Dengan teorema Limit Pusat diperoleh

$$\sqrt{n}C \xrightarrow{d} N(0, S/R) \implies C \xrightarrow{a} N(0, \frac{S}{nR})$$

dengan S merupakan mean populasi untuk S_i . Untuk n cukup besar walaupun R tertentu maka simulasi *noise* menjadi tidak ada. Hal ini mengindikasikan bahwa untuk menghilangkan simulasi *noise* dapat dilakukan dengan mengambil sampel observasi n cukup besar.

Selanjutnya untuk suku bias B . Jika simulasi $\check{g}(\theta)$ tak bias untuk $g(\theta)$ maka suku bias B adalah nul. Bagaimanapun juga jika terdapat bias, maka efek bias pada distribusi $\check{g}_i(\theta)$ harus dipertimbangkan. Biasanya suku $g_i(\theta)$ didefinisikan sebagai fungsi dari statistik $l_i(\theta)$ di mana $E[\check{l}_i(\theta)] = l_i(\theta)$, dan $\check{l}_i(\theta)$ adalah nilai simulasi dari $l_i(\theta)$.

Dengan deret *Taylor* dapat digunakan untuk mendekati nilai $\check{g}_i(\theta)$ menggunakan $g_i(\theta)$ dapat dinyatakan sebagai

$$\check{g}_i(\theta) = g_i(\theta) + \frac{\partial g(l_i(\theta))}{\partial l_i} [\check{l}_i(\theta) - l_i(\theta)] + \frac{1}{2} \frac{\partial^2 g(l_i(\theta))}{\partial l_i^2} [\check{l}_i(\theta) - l_i(\theta)]^2$$

$$\check{g}_i(\theta) - g_i(\theta) = \frac{\partial g(l_i(\theta))}{\partial l_i} [\check{l}_i(\theta) - l_i(\theta)] + \frac{1}{2} \frac{\partial^2 g(l_i(\theta))}{\partial l_i^2} [\check{l}_i(\theta) - l_i(\theta)]^2$$

$$\check{g}_i(\theta) - g_i(\theta) = g_i'[\check{l}_i(\theta) - l_i(\theta)] + \frac{1}{2} g_i''[\check{l}_i(\theta) - l_i(\theta)]^2$$

Karena $\check{l}_i(\theta)$ unbiased untuk $l_i(\theta)$, maka

$$E_r g_i'[\check{l}_i(\theta) - l_i(\theta)] = g_i' [E_r(\check{l}_i(\theta) - l_i(\theta))] = 0$$

sehingga

$$E_r \check{g}_i(\theta) - g_i(\theta) = \frac{1}{2} g_i'' E_r [\check{l}_i(\theta) - l_i(\theta)]^2$$

$$E_r \check{g}_i(\theta) - g_i(\theta) = \frac{1}{2} g_i'' \text{Var}_r[\check{l}_i(\theta)] = \frac{1}{2} g_i'' \frac{Q}{R}$$

oleh karena itu

$$E_r \check{g}(\theta) - g(\theta) = \frac{1}{n} \sum_i (E_r \check{g}_i(\theta) - g_i(\theta)) = \frac{1}{n} \sum_i g_i'' \frac{Q}{2R} = \frac{Z}{R}$$

Karena $B = Z/R$ maka $\sqrt{n}B = \frac{\sqrt{n}}{R} Z$. Jika R fixed maka $B \neq 0$.

Jika $\frac{R}{\sqrt{n}} \rightarrow \infty$ maka $B = \frac{Z}{R} \xrightarrow{p} 0$

kesimpulan:

$$\sqrt{n}\tilde{g}(\theta^*) = \sqrt{n}(A + B + C)$$

$$\sqrt{n}A \xrightarrow{d} N(0, W) \quad : \text{ sama dengan estimator tradisional}$$

$$\sqrt{n}B = \frac{\sqrt{n}}{R}Z \quad : \text{ bias simulasi}$$

$$\sqrt{n}C \xrightarrow{d} N(0, \frac{S}{R}) \quad : \text{ simulation noise}$$

dengan $A = \bar{g}(\theta^*)$; $B = [E_r \tilde{g}(\theta^*) - \bar{g}(\theta^*)]$; $C = [\tilde{g}(\theta^*) - E_r \tilde{g}(\theta^*)]$;

$$W = -E_0 \left[\frac{1}{n} H(\theta_0) \right]^{-1};$$

S merupakan mean populasi untuk S_r ;

Tahap 2: Menurunkan distribusi $\hat{\theta}$ dari distribusi $\tilde{g}(\theta^*)$

Sebagaimana penaksir dengan MLE, distribusi $\hat{\theta}$ dapat diturunkan dari distribusi $g(\theta^*)$ menggunakan deret *Taylor* diperoleh

$$\sqrt{n}(\hat{\theta} - \theta^*) = -\tilde{D}^{-1} \sqrt{n}\tilde{g}(\theta^*) = -\tilde{D}^{-1} \sqrt{n}(A + B + C)$$

dengan \tilde{D} merupakan derivatif $g(\theta^*)$ terhadap parameternya. Sehingga

$$\hat{\theta} = \theta^* - \tilde{D}^{-1}(A + B + C)$$

5.4 Prosedur uji MLE

Prosedur uji yang biasa digunakan dalam MLE adalah Uji rasio *likelihood*, Uji Wald.

5.4.1 Uji rasio *likelihood*

Variabel *random* $\mathbf{Y} = (Y_1, \dots, Y_n)$ yang mempunyai fungsi densitas gabungan $f(\mathbf{y}; \theta)$ dan dipunyai hipotesis $H_0: \theta \in \Omega_0$ versus $H_1: \theta \in \Omega - \Omega_0$. Rasio *Likelihood* didefinisikan sebagai

$$\lambda(y) = \frac{\max_{\theta \in \Omega_0} f(y, \theta)}{\max_{\theta \in \Omega} f(y, \theta)} = \frac{f(y, \hat{\theta}_0)}{f(y, \hat{\theta}_{MLE})}$$

dengan $\hat{\theta}_0$ adalah nilai parameter θ di bawah H_0 benar.

Teorema Uji Rasio Likelihood (Greene, 2005)

Dengan asumsi keteraturan (*regularity*) dan di bawah H_0 benar, untuk $n \rightarrow \infty$

$$-2 \log(\lambda) \sim \chi^2$$

dengan derajat bebas sama dengan banyaknya parameter yang diuji dalam H_0 .

Bukti

Untuk penyederhanaan, misal diuji hipotesis

$$H_0: \theta = \theta_0.$$

Menggunakan deret Taylor orde dua mendekati nilai $\ln L(\theta_0)$ disekitar $\ln(\hat{\theta}_{MLE})$ diperoleh

$$\ln L(\theta_0) = \ln L(\hat{\theta}_{MLE}) + \frac{\partial \ln L}{\partial \theta'} \Big|_{\hat{\theta}_{MLE}} (\theta_0 - \hat{\theta}_{MLE}) + \frac{1}{2} (\theta_0 - \hat{\theta}_{MLE})' \frac{\partial^2 \ln L}{\partial \theta \theta' \partial \theta'} \Big|_{\hat{\theta}_{MLE}} (\theta_0 - \hat{\theta}_{MLE}) + R$$

dengan R adalah suku sisa. Dari sifat MLE diperoleh

$$\frac{\partial \ln L}{\partial \theta'} \Big|_{\hat{\theta}_{MLE}} = 0$$

sehingga dapat disusun

$$-2(\ln L(\hat{\theta}_{MLE}) - \ln L(\theta_0)) = -(\theta_0 - \hat{\theta}_{MLE})' \frac{\partial^2 \ln L}{\partial \theta \theta' \partial \theta'} \Big|_{\hat{\theta}_{MLE}} (\theta_0 - \hat{\theta}_{MLE}) + R$$

Dari sifat MLE, untuk $n \rightarrow \infty$ di bawah H_0 benar

$$\sqrt{n}(\theta_0 - \hat{\theta}_{MLE}) \xrightarrow{d} N\left(0, -\left(p \lim \frac{1}{n} \frac{\partial^2 \ln L}{\partial \theta \theta' \partial \theta'}\right)^{-1}\right)$$

$$- [I(\hat{\theta}_{MLE})]^{-1} = - \left(p \lim \frac{1}{n} \frac{\partial^2 \mathbf{h} L}{\partial \theta \partial \theta'} \right)^{-1}$$

sehingga

$$- 2 \left(\mathbf{h} L(\hat{\theta}_{MLE}) - \mathbf{h} L(\theta_0) \right) \xrightarrow{d} \chi^2$$

5.4.2 Uji wald

Jika H_0 benar yaitu $c(\theta) = \mathbf{q}$, maka $c(\hat{\theta}_{MLE}) \cong \mathbf{q}$ sebab MLE bersifat konsisten. H_0 akan ditolak jika $c(\hat{\theta}_{MLE}) \neq \mathbf{q}$.

Teorema Limit distribusi untuk statistik uji Wald (Greene, 2005)

Diketahui statistik Wald

$$W = [c(\hat{\theta}) - \mathbf{q}] (asy.Var[c(\hat{\theta}) - \mathbf{q}])^{-1} [c(\hat{\theta}) - \mathbf{q}].$$

Jika H_0 benar dan untuk $n \rightarrow \infty$ maka W mempunyai distribusi *chi-square* dengan derajat bebas sama dengan parameter yang diuji dalam H_0 (banyaknya persamaan $[c(\hat{\theta}) - \mathbf{q}] = \mathbf{0}$).

Bukti:

Karena $\hat{\theta}$ merupakan MLE dari θ , maka

$$[c(\hat{\theta}) - \mathbf{q}] \xrightarrow{d} N(0, asy.Var[c(\hat{\theta}) - \mathbf{q}]) \quad \text{untuk } n \rightarrow \infty$$

dan

$$\left(asy.Var[c(\hat{\theta}) - \mathbf{q}] \right)^{\frac{1}{2}} [c(\hat{\theta}) - \mathbf{q}] \xrightarrow{d} N(0, I_p)$$

I_p matriks identitas berukuran p . p adalah banyaknya persamaan $[c(\hat{\theta}) - \mathbf{q}] = \mathbf{0}$.

Selanjutnya

$$\left(\left(asy.Var[c(\hat{\theta}) - \mathbf{q}] \right)^{\frac{1}{2}} [c(\hat{\theta}) - \mathbf{q}] \right)^t \left(asy.Var[c(\hat{\theta}) - \mathbf{q}] \right)^{-\frac{1}{2}} [c(\hat{\theta}) - \mathbf{q}] \xrightarrow{d} \chi_p^2$$

$$\left([c(\hat{\theta}) - \mathbf{q}] \right) \left(asy.Var[c(\hat{\theta}) - \mathbf{q}] \right)^{-1} [c(\hat{\theta}) - \mathbf{q}] \xrightarrow{d} \chi_p^2$$



5.4.3 Lagrange multiplier test

Jika H_0 benar, $c(\theta) = 0$, maka $\hat{\theta}_R$ mendekati titik maksimum dari **log-likelihood**. Statistik uji didasarkan pada *slope log-likelihood*.

Teorema Limit distribusi untuk statistik Lagrange Multiplier

Didefinisikan statistik *Lagrange Multiplier*

$$LM = \left(\frac{\partial \ln L(\hat{\theta}_R)}{\partial \hat{\theta}_R} \right)' [I(\hat{\theta}_R)]^{-1} \left(\frac{\partial \ln L(\hat{\theta}_R)}{\partial \hat{\theta}_R} \right)$$

dengan

$$[I(\hat{\theta}_R)]^{-1} = \left(-E_0 \left[\frac{\partial^2 \ln L(\hat{\theta}_R)}{\partial \hat{\theta}_R \partial \hat{\theta}_R'} \right] \right)^{-1} = \left(E_0 \left[\left(\frac{\partial \ln L(\hat{\theta}_R)}{\partial \hat{\theta}_R} \right) \left(\frac{\partial \ln L(\hat{\theta}_R)}{\partial \hat{\theta}_R} \right)' \right] \right)^{-1}$$

Di bawah H_0 benar, LM mempunyai limit distribusi *chi-square* dengan derajat bebas sama dengan banyaknya parameter yang diuji.

Bukti

Karena $E \left(\frac{\partial \ln L(\hat{\theta}_R)}{\partial \hat{\theta}_R} \right) = 0$ dan $Var \left(\frac{\partial \ln L(\hat{\theta}_R)}{\partial \hat{\theta}_R} \right) = [I(\hat{\theta}_R)]$ maka menggunakan teorema Limit Pusat *Lindberg-Levy* diperoleh

$$[I(\hat{\theta}_R)]^{-\frac{1}{2}} \left(\frac{\partial \ln L(\hat{\theta}_R)}{\partial \hat{\theta}_R} \right) \xrightarrow{d} N(0, I_p)$$

sehingga

$$LM = \left(\frac{\partial \ln L(\hat{\theta}_R)}{\partial \hat{\theta}_R} \right)' [I(\hat{\theta}_R)]^{-1} \left(\frac{\partial \ln L(\hat{\theta}_R)}{\partial \hat{\theta}_R} \right) \xrightarrow{d} \chi_p^2$$

5.5 Simulasi monte carlo

Dalam tahun 1981, Lerman dan Manski menyampaikan metode simulasi *Monte Carlo* untuk menghitung integral dimensi tinggi. Mereka mengusulkan

$$E[h(U)] = \int h(u) \cdot f(u) \cdot du$$



dengan $U \sim f(u)$ dan $h(u)$ adalah sebuah fungsi tertentu, yang secara asimtotis ekuivalen dengan menghitung mean sampel $h(u)$. Oleh karena itu, adalah memungkinkan mengambil variabel *random* sebanyak R kali dari distribusi $f(u)$, misalkan u^r dengan $r=1,2,\dots,R$ dan

$$E_R[h(U)] = \frac{1}{R} \sum_{r=1}^R h(u^r)$$

Jika $E(h(U)) < \infty$ dan pengambilan bersifat iid dari distribusi U ,

$u^r \sim f(u)$ dengan $r=1,2,\dots,R$

sehingga dengan *strong law of large number*

$$\tilde{h}_R = \frac{1}{R} \sum_{r=1}^R h(u^r) \xrightarrow{a.s.} E[h(u)] = \int h(u) f(u) du$$

$$P(\lim_{R \rightarrow \infty} \tilde{h}_R = E[h(u)]) = 1 \text{ dan } E[\tilde{h}_R] = E[h(u)]$$

Jika $\sigma_h^2 \equiv \text{Var}[h(u)] < \infty$, maka *Lindberg-Levy central limit theorem* menghasilkan

$$\text{Var}[\tilde{h}_R] = \frac{\sigma_h^2}{R} \rightarrow 0 \text{ untuk } R \rightarrow \infty$$

Hasil terakhir ini dapat digunakan untuk mengukur ketepatan simulasi, menggantikan σ_h^2 dengan kovariansi sampel

$$\tilde{\sigma}_h^2 = \frac{1}{R} \sum_{r=1}^R [h(u^r) - \tilde{h}_R] \cdot [h(u^r) - \tilde{h}_R]^t$$

5.6 Metode optimisasi

Perhitungan untuk MLE atau persamaan penaksir yang lain seringkali memerlukan prosedur iterasi. Terdapat beberapa metode iterasi, diantaranya adalah metode *Newton-Raphson*, metode *Fisher Scoring*, metode BHHH, metode BFGS, metode DFP.

5.6.1 Metode Newton Raphson (NR) dan Fisher Scoring

Optimisasi fungsi log-likelihood $LL(\theta)$ dapat menggunakan deret Taylor's orde ke dua yaitu mengabaikan nilai pada suku orde tiga atau lebih. Nilai pendekatan terbaik dari $\theta^{(k+1)}$ (pada iterasi ke $k+1$) dengan deret Taylor's orde ke dua, yaitu mendekati $LL(\theta^{(k+1)})$ di sekitar $LL(\theta^{(k)})$,

$$LL(\theta^{(k+1)}) = LL(\theta^{(k)}) + (\theta^{(k+1)} - \theta^{(k)})' g^{(k)} + \frac{1}{2} (\theta^{(k+1)} - \theta^{(k)})' H^{(k)} (\theta^{(k+1)} - \theta^{(k)}) + o(\|\theta^{(k+1)} - \theta^{(k)}\|^2)$$

di mana H adalah matriks *Hessian* yaitu matriks yang mempunyai elemen

$$\frac{\partial^2 LL(\theta)}{\partial \theta_h \partial \theta_j} \text{ dan } g \text{ adalah vektor yang mempunyai elemen } \frac{\partial LL(\theta)}{\partial \theta_j}.$$

Oleh karena itu,

$$g = \frac{\partial LL(\theta)}{\partial \theta} \text{ dan } H = \frac{\partial^2 LL(\theta)}{\partial \theta' \partial \theta} \quad (5.18)$$

$H^{(k)}$ dan $g^{(k)}$ adalah H dan g yang terevaluasi pada $\theta = \theta^{(k)}$.

Selanjutnya mencari nilai $\theta^{(k+1)}$ yang memaksimumkan persamaan $LL(\theta^{(k+1)})$,

$$\frac{\partial LL(\theta^{(k+1)})}{\partial \theta^{(k+1)}} = g^{(k)} + H^{(k)} (\theta^{(k+1)} - \theta^{(k)}) = 0 \quad (5.19)$$

Sehingga persamaan Newton-Raphson-nya adalah

$$\theta^{(k+1)} = \theta^{(k)} - (H^{(k)})^{-1} g^{(k)} \quad (5.20)$$

Persamaan iterasi untuk *Fisher Scoring* adalah

$$\theta^{(k+1)} = \theta^{(k)} + (I^{(k)})^{-1} g^{(k)} \quad (5.21)$$

atau

$$I^{(k)} \theta^{(k+1)} = I^{(k)} \theta^{(k)} + g^{(k)} \quad (5.22)$$

di mana $I^{(k)}$ adalah pendekatan ke- k untuk matriks informasi Fisher, $I^{(k)}$ mempunyai elemen $-E \left(\frac{\partial^2 LL(\theta)}{\partial \theta_h \partial \theta_j} \right)$ yang terevaluasi pada $\theta^{(k)}$.

Untuk mengurangi banyaknya iterasi, dapat dilakukan pembobot skalar λ pada setiap langkah iterasi. Sehingga persamaan (5.21) menjadi

$$\theta^{(k+1)} = \theta^{(k)} + \lambda \cdot (I^{(k)})^{-1} g^{(k)} \quad (5.23)$$

Pendekatan dengan deret *Taylor's orde pertama* untuk $LL(\theta^{(k+1)})$ adalah

$$LL(\theta^{(k+1)}) = LL(\theta^{(k)}) + \lambda(I^{(k)})^{-1}g^{(k)}$$

Langkah-langkah pengambilan nilai λ pada iterasi ke-k adalah sebagai berikut:

1. Diawali dengan pengambilan nilai $\lambda=(1/2)^t$ dengan $t=0$ dan menghitung $\theta^{(k+1)}$
2. Membandingkan nilai $LL(\theta^{(k+1)})$ terhadap $LL(\theta^{(k)})$. Jika $LL(\theta^{(k+1)}) > LL(\theta^{(k)})$ maka ke langkah $\theta^{(k+1)}$ (persamaan 2.59) dan mulai iterasi baru. Jika $LL(\theta^{(k+1)}) < LL(\theta^{(k)})$ maka ke langkah ke 3.
3. Mengambil $\lambda=(1/2)^t$ dengan $t=(t+1)$, dan kembali ke langkah 2. langkah ke 3 ini dilakukan sampai diperoleh nilai λ sangat kecil (mendekati nol)

5.6.2 Metode BHHH

Prosedur iterasi yang lain adalah metode BHHH (Berndt, Hall, Hall, Hausman). Iterasi pada prosedur ini adalah

$$\theta^{(k+1)} = \theta^{(k)} + \lambda.(B_k)^{-1}g^{(k)} \tag{5.24}$$

di mana $B^{(k)}$ merupakan pendekatan dari matriks *Hessian* yang dirumuskan dengan

$$B^{(k)} = \frac{1}{n} \sum_i s_i(\theta^k)[s_i(\theta^k)]'$$

dan

$$s_i(\theta^k) = \left. \frac{\partial \log \pi_i(\theta)}{\partial \theta} \right|_{\theta=\theta^k} = \sum_{j=1}^J \frac{\partial \log \pi_{ij}(\theta)}{\partial \theta} \Big|_{\theta=\theta^k} = \sum_{j=1}^J \frac{1}{\pi_{ij}(\theta)} \frac{\partial \pi_{ij}(\theta)}{\partial \theta} \Big|_{\theta=\theta^k}$$

Jika θ adalah vektor parameter berdimensi $(K \times 1)$ maka

$$s_i(\theta^k)[s_i(\theta^k)]' = \begin{pmatrix} s_i^1 s_i^1 & s_i^1 s_i^2 & \dots & s_i^1 s_i^K \\ s_i^1 s_i^2 & s_i^2 s_i^2 & \dots & s_i^2 s_i^K \\ \dots & \dots & \dots & \dots \\ s_i^1 s_i^K & s_i^2 s_i^K & \dots & s_i^K s_i^K \end{pmatrix} \tag{5.25}$$

dengan s_i^j adalah elemen ke j dari $s_i(\theta^k)$.

Terdapat dua keunggulan prosedur BHHH dibanding prosedur NR (Train, 2003):

1. Menghitung B_k adalah lebih mudah dan lebih cepat dihitung dibanding dengan H_k . B_k hanya membutuhkan derivatif pertama dari fungsi *likelihood*, sedangkan

H_k membutuhkan derivatif kedua dari fungsi *likelihood*.

2. Dalam BHHH dijamin bahwa setiap langkah iterasi akan menghasilkan nilai *log-likelihood* yang selalu lebih besar dibanding tahap sebelumnya.

5.6.3 Metode BFGS dan DFP

Metode BFGS ini disampaikan oleh Broyden, Fletcher, Goldfarb dan Shanno pada tahun 1970. Persamaan iterasi BFGS adalah (Chong dan Zak, 1996)

$$B_{k+1} = B_k + \frac{\Delta \mathbf{g}^{(k)} (\Delta \mathbf{g}^{(k)})'}{(\Delta \mathbf{g}^{(k)})' \Delta \boldsymbol{\theta}^{(k)}} - \frac{[B_k \Delta \boldsymbol{\theta}^{(k)}][B_k \Delta \boldsymbol{\theta}^{(k)}]'}{[\Delta \boldsymbol{\theta}^{(k)}]' [B_k \Delta \boldsymbol{\theta}^{(k)}]} \quad (5.26)$$

dengan

$$\Delta \boldsymbol{\theta}^{(k)} = \boldsymbol{\theta}^{(k+1)} - \boldsymbol{\theta}^{(k)}$$

$$\Delta \mathbf{g}^{(k)} = \mathbf{g}^{(k+1)} - \mathbf{g}^{(k)}$$

Sehingga BFGS tidak memerlukan derivatif ke dua dari fungsi *log-likelihood*.

Sedangkan pada metode iterasi DFP (Davidson-Fletcher-Powell), persamaan iterasinya adalah

$$H_{k+1}^{DFP} = H_k + \frac{\Delta \boldsymbol{\theta}^{(k)} (\Delta \boldsymbol{\theta}^{(k)})'}{(\Delta \boldsymbol{\theta}^{(k)})' \Delta \mathbf{g}^{(k)}} - \frac{[H_k \Delta \mathbf{g}^{(k)}][H_k \Delta \mathbf{g}^{(k)}]'}{[\Delta \mathbf{g}^{(k)}]' [H_k \Delta \mathbf{g}^{(k)}]} \quad (5.27)$$

Algoritma BFGS dan DFP adalah

1. Pada $k=0$, menentukan nilai penduga awal untuk $\boldsymbol{\theta}^{(0)}$ dan H_0 (matriks simetri yang definit positif).
2. Menghitung

$$\mathbf{g}^{(k)} = \left. \frac{\partial LL(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^{(k)}}$$

Jika $\mathbf{g}^{(k)} = \mathbf{0}$ maka berhenti yang berarti bahwa $\hat{\boldsymbol{\theta}} = \boldsymbol{\theta}^{(k)}$. Jika $\mathbf{g}^{(k)} \neq \mathbf{0}$ maka

$$\mathbf{d}^{(k)} = -H_k \mathbf{g}^{(k)}$$

3. Menghitung

$$\alpha_k = \arg \min LL(\boldsymbol{\theta}^{(k)} + \alpha \mathbf{d}^{(k)})$$



$$\boldsymbol{\theta}^{(k+1)} = \boldsymbol{\theta}^{(k)} + \alpha_k \cdot \mathbf{d}^{(k)}$$

4. Menghitung

$$\Delta \boldsymbol{\theta}^{(k)} = \alpha_k \cdot \mathbf{d}^{(k)} ; \Delta g^{(k)} = g^{(k+1)} - g^{(k)}$$

untuk BFGS: \mathbf{H}_{k+1}^{BFGS} pada persamaan (5.28)

untuk DFP: \mathbf{H}_{k+1}^{DFP} pada persamaan (5.29)

$k=k+1$, mengulangi langkah 2

Bab 6

Menghitung Nilai Probabilitas pada Distribusi Normal Multivariat

6.1 Pendahuluan

Menghitung nilai probabilitas variabel *random* yang mempunyai distribusi multivariat normal, merupakan salah satu kendala dalam model *Probit*. Persamaan probabilitas dalam model *Probit* tidak berbentuk persamaan tertutup dan harus diselesaikan secara numerik. Beberapa prosedur non simulasi telah banyak digunakan seperti metode kuadrat untuk mendekati hitungan integral menggunakan fungsi pembobot. Metode ini hanya efektif untuk hitungan integral dimensi rendah. Metode ini dapat digunakan untuk integral dengan dimensi tidak lebih dari empat atau lima (Geweke, 1996). Metode non simulasi yang lain adalah *Clark Algoritma*, yang telah disampaikan oleh Daganzo, dkk. (1977). Metode ini sangat tidak akurat sebab akurasi dipengaruhi oleh modelnya (Horowitz, dkk., 1982).

Untuk menghitung probabilitas dalam model *Probit* dapat juga digunakan metode simulasi. Banyak cara simulasi yang dapat digunakan dalam model *Probit*, secara ringkas dijelaskan oleh Hajivassiliou, dkk. (1996) dan menyimpulkan bahwa *Geweke-Hajivassiliou-Keane* (GHK) adalah metode yang paling baik paling baik. GHK simulator menghasilkan nilai antara nol s/d satu, *unbiased* dan bersifat konsisten (Paul C., dkk., 2001)

Permasalahan yang sering muncul dalam statistika adalah menghitung probabilitas normal multivariat. Dalam bab ini ini akan dibahas metode menghitung nilai probabilitas variabel *random* yang mempunyai distribusi multivariat normal menggunakan program R. Pembahasan dimulai dari distribusi normal univariat, distribusi normal bivariat, distribusi normal trivariat dan distribusi normal multivariat. Terakhir adalah pembahasan mengenai penggunaan Metode GHK dalam model *Multinomial Probit*.

6.2 Distribusi normal univariat

Distribusi Normal (*Gaussian*) yang dinotasikan dengan $N(\mu, \sigma^2)$ mempunyai mean μ dan variansi σ^2 mempunyai densitas

$$f(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right] \text{ dengan } \mu, \sigma \in \mathbb{R} \quad (6.1)$$

Selanjutnya, jika variabel *random* Z berdistribusi normal standar (mean $\mu = 0$ dan variansi $\sigma^2 = 1$). Fungsi distribusi $F(z_0)$ yang dinotasikan dengan $\Phi(z_0)$ adalah

$$\Phi(z_0) = P(Z \leq z_0) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z_0} \exp[-t^2/2] dt \quad (6.2)$$

Untuk mendapatkan nilai $F(z_0)$ untuk suatu nilai z_0 , memerlukan transformasi deret Taylor

$$\exp(-z^2/2) = \sum_{i=1}^{\infty} (-1)^i \frac{z^{2i}}{2^i i!}$$

sehingga

$$\begin{aligned} \Phi(z) &= F_z(z) = F_z(0) + \frac{1}{\sqrt{2\pi}} \int_0^z \exp[-t^2/2] dt \\ &= \frac{1}{2} + \frac{1}{\sqrt{2\pi}} \sum_{i=0}^{\infty} (-1)^i \frac{z^{2i+1}}{(2i+i)2^i i!} \end{aligned} \quad (6.3)$$

(Paolella, 2006). Invers fungsi $F(z_0)$, yaitu $F^{-1}(p)$, $0 < p < 1$, juga dapat dihitung. Perhitungan yang sederhana dan mempunyai akurasi tinggi telah dilakukan oleh Moro (1995).

6.3 Distribusi normal bivariat

Variabel *random* (X, Y) yang berdistribusi normal bivariat, dengan

$$\text{Var}(X) = \text{Var}(Y) = 1 \text{ dan } \text{Korr}(X, Y) = r,$$

fungsi distribusinya dapat dituliskan dalam bentuk. Distribusi normal bivariat dapat dituliskan dalam bentuk

$$\Phi(b, \rho) = \frac{1}{2\pi\sqrt{1-\rho^2}} \int_{-\infty}^{b_1} \int_{-\infty}^{b_2} \exp\left[-\frac{(x^2 - 2\rho xy + y^2)}{2(1-\rho^2)}\right] dy dx \quad (6.4)$$

Tersa dan wellan, (1990) menghitung probabilitas normal bivariat dengan mendefinisikan probabilitas dalam bentuk

$$L(h, k, \rho) = \frac{1}{2\pi\sqrt{1-\rho^2}} \int_h^\infty \int_k^\infty \exp\left[-\frac{(x^2 - 2\rho xy + y^2)}{2(1-\rho^2)}\right] dy dx \quad (6.5)$$

dan $\Phi(\mathbf{b}, \rho) = L(-b_1, -b_2, \rho)$.

Bukti:

$$\begin{aligned} L(h, k, \rho) &= \frac{1}{2\pi\sqrt{1-\rho^2}} \int_h^\infty \int_k^\infty \exp\left[-\frac{(y - \rho x)^2 + (x^2 - \rho^2 x^2)}{2(1-\rho^2)}\right] dy dx \\ &= \frac{1}{2\pi\sqrt{1-\rho^2}} \int_h^\infty \int_k^\infty \exp\left[-\frac{(y - \rho x)^2}{2(1-\rho^2)}\right] \cdot \exp\left[-\frac{(1-\rho^2)x^2}{2(1-\rho^2)}\right] dy dx \\ &= \int_h^\infty \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{x^2}{2}\right] \left[\int_k^\infty \frac{1}{\sqrt{2\pi}\sqrt{1-\rho^2}} \exp\left[-\frac{(y - \rho x)^2}{2(1-\rho^2)}\right] dy \right] dx \\ &= \int_h^\infty \phi(x) \left[\int_k^\infty \frac{1}{\sqrt{2\pi}\sqrt{1-\rho^2}} \exp\left[-\frac{1}{2}\left(\frac{y - \rho x}{\sqrt{1-\rho^2}}\right)^2\right] \right] \sqrt{1-\rho^2} \cdot d\left[\frac{y - \rho x}{\sqrt{1-\rho^2}}\right] dx \\ &= \int_h^\infty \phi(x) \left[\int_{\frac{k - \rho x}{\sqrt{1-\rho^2}}}^\infty \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{w^2}{2}\right] dw \right] dx \\ &= \int_h^\infty \phi(x) \left[1 - \Phi\left(\frac{k - \rho x}{\sqrt{1-\rho^2}}\right) \right] dx \\ L(h, k, \rho) &= \int_h^\infty \phi(x) \Phi\left(\frac{\rho x - k}{\sqrt{1-\rho^2}}\right) dx = \Phi(-h, -k, \rho) \end{aligned}$$

sebab $\Phi(\mathbf{b}, \rho) = L(-b_1, -b_2, \rho)$.

Oleh karena itu,

$$\Phi(h, k, \rho) = \int_{-\infty}^h \phi(x) \Phi\left(\frac{k - \rho x}{\sqrt{1 - \rho^2}}\right) dx \quad (\text{terbukti}).$$

Drezner dan Wesolowsky (1990) mengembangkan rumus:

$$L(h, k, r) = \Phi(-h)\Phi(-k) + \frac{1}{2\pi} \int_0^r \frac{1}{\sqrt{(1-r^2)}} \exp\left[-\frac{(h^2 - 2rhk + k^2)}{2(1-r^2)}\right] dr \quad (6.6)$$

untuk menghitung probabilitas normal bivariate. Derivatif parsial terhadap koefisien korelasi dapat ditulis menjadi,

$$\frac{\partial L(h, k, r)}{\partial r} = \frac{1}{2\pi\sqrt{(1-r^2)}} \exp\left[-\frac{(h^2 - 2rhk + k^2)}{2(1-r^2)}\right]$$

Jika mensubstitusi $r = \sin \theta$ maka,

$$L(h, k, r) = \Phi(-h)\Phi(-k) + \frac{s}{2\pi} \int_0^{\sin^{-1}(\rho)} \exp\left[-\frac{(h^2 - 2hk \sin(\theta) + k^2)}{2[\cos(\theta)]^2}\right] d\theta \quad (6.7)$$

rumus tersebut akan singular (tidak terdefinisi) jika $r = 1$ dan akan kehilangan akurasi untuk $|\rho| \neq 1$. Drezner dan Wesolowsky mengintegrasikan antara r dan s bukannya dari 0 dan r , sehingga

$$L(h, k, \rho) = \Phi(-h)\Phi(-k) + \frac{s}{2\pi} \int_\rho^s \frac{1}{\sqrt{(1-r^2)}} \exp\left[-\frac{(h^2 - 2rhk + k^2)}{2(1-r^2)}\right] dr \quad (6.8)$$

$$\text{di mana } L(h, k, s) = \begin{cases} \Phi(-\max(h, k)) & \text{jika } s = 1 \\ \max(0, \Phi(-h) - \Phi(k)) & \text{jika } s = -1 \end{cases}$$

Jika substitusi $x = \sqrt{1 - r^2}$ digunakan maka

$$L(h, k, \rho) = L(h, k, s) - \frac{s}{2\pi} \int_0^{\sqrt{1-\rho^2}} \frac{1}{\sqrt{1-x^2}} \exp\left[-\frac{(h^2 - 2shk\sqrt{1-x^2} + k^2)}{2x^2}\right] dx \quad (6.10)$$

Integrasi numerik persamaan ini sangat akurat ketika $|r| \approx 1$ kecuali ketika h adalah cukup dekat (tetapi tidak sama) dengan sk . Untuk menghindari masalah ketika $h \approx sk$, Drezner dan Wesolowsky menuliskan dalam bentuk

$$L(h, k, \rho) = L(h, k, s) - \frac{s}{2\pi} \int_0^{\sqrt{1-\rho^2}} \frac{1}{\sqrt{1-x^2}} \exp\left[-\frac{(h-sk)}{2x^2}\right] \cdot \exp\left[-\frac{shk}{1+\sqrt{1-x^2}}\right] dx \quad (6.11)$$

Integral dalam penyajian ini adalah *smooth* kecuali untuk suku $\exp\left(-\frac{(h-sk)^2}{2x^2}\right)$ yang menghasilkan error cukup besar ketika h cukup dekat dengan sk .

Menggunakan deret *Taylor*:

$$\frac{1}{\sqrt{1-x^2}} \exp\left[-\frac{shk}{1+\sqrt{1-x^2}}\right] = \exp\left[-\frac{shk}{2}\right] \left(1 + \frac{(4-shk)x^2}{8}\right) + O(x^4)$$

$$\begin{aligned} L(h, k, \rho) &= L(h, k, s) - \frac{s}{2\pi} \int_0^{\sqrt{1-\rho^2}} \frac{1}{\sqrt{1-x^2}} \exp\left[-\frac{(h-sk)}{2x^2}\right] \cdot \exp\left[-\frac{shk}{2}\right] \left(1 + \frac{(4-shk)x^2}{8}\right) dx \\ &= \frac{s}{2\pi} \int_0^{\sqrt{1-\rho^2}} \exp\left[-\frac{(h-sk)}{2x^2}\right] \cdot \left(\frac{\exp\left[-\frac{shk}{1+\sqrt{1-x^2}}\right]}{\sqrt{1-x^2}} - \exp\left(-\frac{shk}{2}\right) \left(1 + \frac{(4-shk)x^2}{8}\right) \right) dx \quad (6.12) \end{aligned}$$

Algoritma Drezner dan Wesolowsky menghitung kedua integral secara numerik dan integral suku

$$\exp\left[-\frac{(h-sk)}{2x^2}\right] - \exp\left(-\frac{shk}{2}\right) \left(1 + \frac{(4-shk)x^2}{8}\right)$$

secara analitik dengan menggunakan rumus:

$$\int_0^a \exp\left(-\frac{b^2}{2x^2}(1+cx^2)\right) dx = a\left(1 - \frac{c(b^2-a^2)}{3}\right) \exp\left(-\frac{b^2}{2a^2}\right) - b\left(1 - \frac{cb^2}{3}\right) \sqrt{2\pi} \Phi\left(-\frac{b}{a}\right) \quad (6.13)$$

dengan $a = \sqrt{1-\rho^2}$, $b=|h-sk|$ dan $c = (4-shk)/8$

6.4 Probabilitas normal trivariat

Distribusi normal trivariat dapat dituliskan sebagai berikut:

$$\Phi(b, R) = \frac{1}{(2\pi)^{3/2} |R|^{1/2}} \int_{-\infty}^{b_1} \int_{-\infty}^{b_2} \int_{-\infty}^{b_3} \exp\left[-\frac{x^T R^{-1} x}{2}\right] dx_1 dx_2 dx_3 \quad (6.14)$$

di mana $b=(b_1, b_2, b_3)$ dan $R=(\rho_{ij})$ adalah matrik korelasi. Penyajian distribusi normal trivariat berdasarkan distribusi normal bivariat dalam bentuk (Genz,1992)

$$\Phi(b, R) = \frac{1}{(2\pi)^{1/2}} \int_{-\infty}^{b_1} \exp\left[-\frac{x^2}{2}\right] F(x) dx \quad (6.15)$$

di mana

$$F(x) = \Phi\left(\left(\frac{b_2 - \rho_{21}x}{(1 - \rho_{21}^2)^{1/2}}, \frac{b_3 - \rho_{31}x}{(1 - \rho_{31}^2)^{1/2}}\right), \frac{\rho_{32} - \rho_{31}\rho_{21}}{(1 - \rho_{21}^2)^{1/2}(1 - \rho_{31}^2)^{1/2}}\right)$$

Terdapat beberapa metode yang dapat digunakan untuk menghitung $F(\cdot)$ pada persamaan (6.15), yaitu

1. Menggunakan transformasi $x=\Phi^{-1}(t)$, sehingga

$$\Phi(b, R) = \int_0^{\Phi(b_1)} F(\Phi^{-1}(t)) dt \quad (6.16)$$

dan selanjutnya itegrasi numerik menggunakan aturan *Gauss-Legendre* pada interval $[0, \Phi(b_1)]$.

2. Menggunakan metode *Drezner* (1992), yang menggunakan modifikasi aturan *Gauss-Hermite*. Aturan tersebut sangat cocok untuk integral dalam bentuk

$$\frac{1}{\sqrt{2\pi}} \int_0^{\infty} \exp\left(-\frac{x^2}{2}\right) f(x) dx$$

sehingga persamaan 15 ditranformasi ke dalam interval $[0, \infty]$. Misal $y = x - b_1$ dan selanjutnya

$$\Phi(b, R) = \frac{1}{(2\pi)^{1/2}} \int_{-\infty}^0 \exp\left[-\frac{(y + b_1)^2}{2}\right] F(y + b_1) dy$$

$$= \frac{\exp(-b_1^2/2)}{(2\pi)^{1/2}} \int_{-\infty}^0 \exp\left[-\frac{y^2/2 - yb_1}{2}\right] F(y + b_1) dy \quad (6.17)$$

jika $z=-y$ maka

$$\Phi(b, R) = \frac{\exp(-b_1^2/2)}{(2\pi)^{1/2}} \int_0^{\infty} \exp\left[-z^2/2 + zb_1\right] F(b_1 - z) dz \quad (6.18)$$

Kedua metode sangat akurat untuk menghitung integral pada daerah terbatas (Genz,1993).

Rumus untuk menghitung nilai probabilitas normal trivariat yang lain didasarkan pada rumus *Plackett*, yang didasarkan atas derivatif parsial.

$$\frac{\partial \Phi(b, R)}{\partial \rho_{21}} = \frac{\exp(-f_3(\rho_{21})/2)}{2\pi\sqrt{1-\rho_{21}^2}} \Phi(u_3(\rho_{21}))$$

di mana

$$f_3(r) = \frac{b_1^2 + b_2^2 - 2b_1b_2}{(1-r^2)} \text{ dan}$$

$$u_3(r) = \frac{b_3(1-r^2) - b_1(\rho_{31} - r\rho_{32}) - b_2(\rho_{32} - r\rho_{31})}{((1-r^2)(1-r^2 - \rho_{31}^2 - \rho_{32}^2 + 2r\rho_{31}\rho_{32}))^{1/2}}$$

terdapat dua metode *Plackett*, yaitu yang pertama

$$\rho_{21}^* = \rho_{31}\rho_{32} \pm \sqrt{(1-\rho_{31}^2)(1-\rho_{32}^2)} \text{ dan } R^* = \begin{bmatrix} 1 & \rho_{21}^* & \rho_{31} \\ \rho_{21}^* & 1 & \rho_{32} \\ \rho_{31} & \rho_{32} & 1 \end{bmatrix}$$

metode *Plackett* yang ke dua adalah

$$\Phi(b, R) = \Phi(b, R') + \frac{1}{2\pi} \int_0^1 \left(\rho_{21} \frac{\exp(-f_3(\rho_{21}t)/2)}{\sqrt{1-\rho_{21}^2t^2}} \Phi(\hat{u}_3(t)) + \rho_{31} \frac{\exp(-f_2(\rho_{31}t)/2)}{\sqrt{1-\rho_{31}^2t^2}} \Phi(\hat{u}_2(t)) \right) dt \quad (6.19)$$

di mana $f_2(r) = \frac{b_1^2 + b_3^2 - 2rb_1b_3}{(1-r^2)}$ dan $R' = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & \rho_{32} \\ 0 & \rho_{32} & 1 \end{bmatrix}$

$$\hat{u}_2(t) = \frac{b_3(1 - \rho_{31}^2 t^2) - b_1 t(\rho_{21} - \rho_{31} \rho_{32}) - b_3(\rho_{32} - t^2 \rho_{21} \rho_{31})}{((1 - \rho_{31}^2 t^2)(1 - \rho_{31}^2 t^2 - \rho_{21}^2 t^2 - \rho_{32}^2 + 2t^2 \rho_{31} \rho_{21} \rho_{32}))^{1/2}}$$

$$\hat{u}_3(t) = \frac{b_3(1 - \rho_{21}^2 t^2) - b_1 t(\rho_{31} - \rho_{21} \rho_{32}) - b_2(\rho_{32} - t^2 \rho_{31} \rho_{21})}{((1 - \rho_{21}^2 t^2)(1 - \rho_{21}^2 t^2 - \rho_{31}^2 t^2 - \rho_{32}^2 + 2t^2 \rho_{31} \rho_{21} \rho_{32}))^{1/2}}$$

Dalam kasus ini, $\Phi(b, R^1)$ dihitung menggunakan $\Phi(b_1)\Phi((b_2, b_3), \rho_{32})$. Metode ke dua ini telah diimplementasikan oleh Drezner (1994).

6.5 Distribusi multivariat normal

Fungsi densitas multivariat normal dapat disajikan dalam bentuk

$$F(a, b) = \frac{1}{\sqrt{|\Sigma|} (2\pi)^m} \int_{a_1}^{b_1} \int_{a_2}^{b_2} \dots \int_{a_m}^{b_m} \exp\left(-\frac{1}{2} \theta' \Sigma^{-1} \theta\right) d\theta \quad (6.20)$$

di mana $\theta = (\theta_1, \theta_2, \dots, \theta_m)^t$ dan Σ adalah matrik kovarian mxm yang simetrik positif definit.

Pada korelasi konstan, $\sigma_{ij} = \rho$, untuk $i \neq j$ dan $\sigma_{ii} = 1$, nilai $F(\cdot)$ dapat dihitung dengan akurat menggunakan normal independen (Tong, 90).

$$F(b) = \frac{1}{(2\pi)^{1/2}} \int_{-\infty}^{\infty} \exp\left[-\frac{t^2}{2}\right] \prod_{i=1}^m \Phi\left(\frac{b_i + \sqrt{\rho}t}{\sqrt{1-\rho}}\right) dt \quad (6.21)$$

Secara umum, untuk mengitung integral rangkap, fungsi densitasnya ditranformasi menggunakan faktor *Cholesky*, yaitu $\theta = Cy$ di mana $CC^t = \Sigma$.

Sehingga $\theta' \Sigma^{-1} \theta = y^t C^t C^{-1} C^{-1} C y = y^t y$ dan $d\theta = |C| dy = |\Sigma|^{1/2} dy$. Karena $a \leq \theta = Cy \leq b$ maka

$$\frac{(a_i - \sum_{j=1}^{i-1} c_j y_j)}{c_i} \leq y_i \leq \frac{(b_i - \sum_{j=1}^{i-1} c_j y_j)}{c_i} \quad \text{untuk } i=1,2,\dots,m$$

Oleh karena itu,

$$F(a, b) = \frac{1}{\sqrt{|\Sigma|} (2\pi)^m} \int_{a_1}^{b_1} \exp\left(-\frac{y_1^2}{2}\right) \int_{a_2'(y_1)}^{b_2'(y_1)} \exp\left(-\frac{y_2^2}{2}\right) \dots \int_{a_m'(y_1, \dots, y_{m-1})}^{b_m'(y_1, \dots, y_{m-1})} \exp\left(-\frac{y_m^2}{2}\right) dy \quad (6.22)$$

dengan

$$a_i'(y_1, \dots, y_{i-1}) = \left(a_i - \sum_{j=1}^{i-1} c_{ij} y_j\right) / c_{ii} \text{ dan } b_i'(y_1, \dots, y_{i-1}) = \left(b_i - \sum_{j=1}^{i-1} c_{ij} y_j\right) / c_{ii}$$

Untuk masing-masing y_i dapat ditransformasi menggunakan $y_i = \Phi^{-1}(z_i)$, dengan

$$\Phi(y) = \frac{1}{(2\pi)^{1/2}} \int_{-\infty}^y \exp\left[-\frac{\theta^2}{2}\right] d\theta$$

yang merupakan fungsi distribusi normal standar univariat. Setelah transformasi itu, $F(a, b)$ menjadi

$$F(a, b) = \int_{d_1}^{e_1} \int_{d_2(z_1)}^{e_2'(z_1)} \dots \int_{d_m(z_1, \dots, z_{m-1})}^{e_m(z_1, \dots, z_{m-1})} dz \quad (6.23)$$

dengan

$$d_i(z_1, \dots, z_{i-1}) = \Phi\left(\left[a_i - \sum_{j=1}^{i-1} c_{ij} \Phi^{-1}(z_j)\right] / c_{ii}\right) \text{ dan}$$

$$e_i(z_1, \dots, z_{i-1}) = \Phi\left(\left[b_i - \sum_{j=1}^{i-1} c_{ij} \Phi^{-1}(z_j)\right] / c_{ii}\right)$$



Bentuk integral ini lebih sederhana dibanding integral aslinya. Integrasi wilayah (*region*) adalah lebih kompleks, sehingga tidak bisa ditangani secara langsung menggunakan algoritma numerik standar untuk integral rangkap. Penyelesaian masalah ini menggunakan tranformasi $z_i = d_i + w_i (e_i - d_i)$. Setelah tranformasi akhir ini

$$F(a, b) = (e_1 - d_1) \int_0^1 (e_2 - d_2) \dots \int_0^1 (e_m - d_m) \int_0^1 dw \quad (6.24)$$

dengan $d_i = \Phi \left([a_i - \sum_{j=1}^{i-1} c_{ij} \Phi^{-1}(d_j + w_j(e_j - d_j))] / c_{ii} \right)$ dan

$$e_i = \Phi \left((b_i - \sum_{j=1}^{i-1} c_{ij} \Phi^{-1}(d_j + w_j(e_j - d_j))) / c_{ii} \right)$$

Integrasi pertama terhadap variabel $w_1, w_2 \dots$ dst.

$$F(a, b) = (e_1 - d_1) \int_0^1 (e_m - d_m) \dots \int_0^1 (e_2 - d_2) \int_0^1 dw_1 \dots dw_m \quad (6.25)$$

Integral ini dapat dinyatakan dalam algoritma sebagai berikut.

1. Input $\Sigma, \alpha, \varepsilon, a, b$ dan N_{\max}
2. Menghitung faktor *Cholesky* bagian bawah, C untuk Σ .
3. Nilai awal $Itsum=0, N=0, Varsum=0, d_1 = \Phi(a_1/c_{1,1}), e_1 = \Phi(b_1/c_{1,1})$ dan $f_1 = e_1 - d_1$.
4. *Repeat*.
 - a. Membangkitkan variabel *random uniform* $w_1, w_2, \dots, w_{m-1} \in [0, 1]$
 - b. untuk $i=2, 3, \dots, m$
 - c. $y_{i-1} = \Phi^{-1}(d_{i-1} + w_{i-1}(e_{i-1} - d_{i-1}))$; $d_i = \Phi \left((a_i - \sum_{j=1}^{i-1} c_{ij} y_j) / c_{ii} \right)$
 $e_i = \Phi \left((b_i - \sum_{j=1}^{i-1} c_{ij} y_j) / c_{ii} \right)$ dan $f_i = (e_i - d_i) f_{i-1}$
 - d. $N=N+1$; $\delta = (f_m - Itsum) / N$; $Itsum = Itsum + \delta$; $Varsum = (N-2)Varsum / N + \delta^2$
 dan $Error = \alpha \sqrt{Varsum}$
 Until $Error < \varepsilon$ atau $N = N_{\max}$
 - e. *Output*: $Itsum, Error, dan N$

Parameter input α adalah faktor confidensi montecarlo untuk eror standar. Misalnya untuk $\alpha = 2,5$, kita mengharapkan eror aktual dalah F menjadi lebih kecil dari $Error$

99%. N_{\max} adalah parameter input yang merupakan total jumlah perhitungan.

Untuk $a_i = -\infty$ atau $b_i = \infty$ maka $d_i = 0$ atau $e_i = 1$.

Dalam beberapa aplikasi, kadang kita menghitung integral

$$G(g) = \frac{1}{\sqrt{|\Sigma|} (2\pi)^m} \int_{a_1}^{b_1} \int_{a_2}^{b_2} \dots \int_{a_m}^{b_m} \exp\left(-\frac{1}{2} \theta^t \Sigma^{-1} \theta\right) g(\theta) d\theta \quad (6.26)$$

Untuk menghitung integral ini memerlukan modifikasi dari algoritma di atas. Karena $g(\theta)$ yang tergantung pada θ_m dan w_m . Sehingga perlu menghitung $\theta = Cy$ dan menghitung $f_m(g(\theta))$ dalam langkah 4c.

6.6 Menghitung nilai probabilitas dalam program R

Terdapat beberapa *software* yang cukup reliabel dan efisien untuk kasus $m=1$ dan $m=2$ (Drezner, dkk., 1990) dan (Cox, dkk., 1991). Algoritma *Schervish* (Schervish, 1984) telah diimplementasikan untuk m kurang dari 8. Algoritma ini menggunakan integrasi secara *locally adaptive numerical* yang didasarkan pada *Simpson's rule*. Algoritma ini membutuhkan waktu yang lama untuk m lebih dari 4 (Genz, 1993). Pendekatan lain adalah dengan metode *Monte Carlo* atau metode *Subregion Adaptive*. Metode ini cukup reliabel dan efisien untuk menghitung $F(\cdot)$. (Genz, A., 1992). Genz (1993) telah melakukan perbandingan beberapa metode untuk menghitung nilai probabilitas normal multivariat. Metode yang dibandingkan adalah *acceptance-rejection sampling*, metode *Deák*, metode *Genz* dan metode *Schervish*. Disimpulkan bahwa metode *Genz* adalah yang paling efisien.

Pada distribusi normal, untuk menghitung fungsi distribusi maupun fungsi densitas dalam program R, dapat digunakan perintah

`>pnorm(x, μ , σ):` untuk menghitung nilai fungsi densitas normal

`>dnorm(x, μ , σ):` untuk menghitung nilai fungsi distribusi normal

Dalam program R juga terdapat beberapa paket yang dapat digunakan untuk menghitung nilai probabilitas multivariat. Perhitungan probabilitas yang didasarkan pada algoritma Genz (1992, 1993) dapat diakses dalam "*fmultivar Package*," "*mnormt Package*," "*mvtnorm Package*," "*adapt package*." Paket lain yang didasarkan pada metode *Schervish's* dan metode "*Joe*" dapat diperoleh pada "*mProbit Package*" (R Core



Team, 2007).

Seiring dengan kemajuan teknologi komputasi, saat ini memungkinkan menghitung nilai probabilitas dari variabel *random* yang berdistribusi normal multivariat. Metode yang banyak diaplikasikan dalam program R adalah metode *Genz* yang disusun berdasarkan simulasi *Monte Carlo*. Aplikasi metode ini dalam bidang ekonometrika, adalah untuk simulasi nilai probabilitas dikenal dengan nama metode GHK. Dengan adanya simulasi probabilitas (GHK) ini, maka dengan program R adalah sangat memungkinkan untuk menyusun model *multinomial Probit*.

Bab 7

Uji *Goodness of Fit* dalam DCM

7.1 Pendahuluan

Dalam statistika parametrik, distribusi atas variabel *random* menjadi dasar untuk melakukan uji statistik. Pada beberapa pengujian, penyimpangan asumsi ini akan beresiko pada keabsahan uji statistiknya. Asumsi yang paling sering digunakan adalah distribusi normal. Untuk melakukan uji kenormalan data dapat digunakan beberapa metode-metode berikut yaitu seperti *Anderson-Darling test*, *Kolmogorov-Smirnov test*, *Pearson Chi-square test*, dan *Cramer-von Mises test*. Metode lain yang dapat digunakan adalah metode *grafis*, *Shapiro-Wilk test*, dan *Fisher's cumulatif test*.

Dalam regresi linear, kita dapat melakukan pengujian asumsi normalitas

$$Y_i \sim N(\beta X_i, \sigma^2)$$

Pengujian dilakukan berdasarkan nilai residual

$$e_i = (Y_i - \hat{Y}) \sim N(0, \sigma^2) \text{ dengan } \hat{Y}_i = \hat{\beta} X_i$$

menggunakan plot distribusi normal. Metode lainnya adalah *Q-Q plot*, *Shapiro-Wilk's test (Wilk's test)*, *Kolmogorov-Smirnov test (K-S test)*. Demikian juga dalam model pemilihan diskrit, penaksiran parameter maupun uji statistik didasarkan atas distribusi komponen variabel randomnya. Dalam makalah ini dibahas uji *goodness of fit* dalam model DCM.

Utilitas yang digunakan oleh responden untuk membuat keputusan dinyatakan sebagai:

$$U_{ij} = V_{ij} + \varepsilon_{ij} \quad i=1, \dots, n \text{ dan } j=1, \dots, J.$$

$$V_{ij} = f(Z_{ij}, X_i) = \alpha_j + \beta_j X_i + \gamma Z_{ij}$$

Responden memilih pilihan yang mempunyai utilitas terbesar, sehingga memilih pilihan k jika dan hanya jika $U_{ik} > U_{ij} \forall j \neq k$ (Train, 2003). Nilai utilitas U_{ij} tidak

terobservasi, yang diketahui (terobservasi) adalah variabel (Y_i, X_i, Z_i) . Y_i berdistribusi



multinomial. Model DCM merupakan probabilitas responden i memilih pilihan k yang dapat dinyatakan sebagai:

$$\begin{aligned}\pi_{ik} &= P(U_{ik} > U_{ij}) \quad \forall j \neq k \\ &= \int_{\varepsilon} I(\varepsilon_j - \varepsilon_k < V_k - V_j) f(\varepsilon_i) d\varepsilon_{i1} \dots d\varepsilon_{ij} \quad \forall j \neq k \\ &= \int_{\varepsilon} I(\varepsilon_{j,k} < V_{k,j}) f(\varepsilon_i) d\varepsilon_{i1} \dots d\varepsilon_{ij} \quad \forall j \neq k\end{aligned}$$

Peneliti tidak mempunyai informasi mengenai variabel ε_{ij} . Asumsi yang berbeda mengenai distribusi ε_{ij} akan menghasilkan persamaan π_{ik} yang berbeda. Pemilihan model lebih mempertimbangkan sifat-sifat pilihan.

1. Jika diasumsikan ε_{ij} berdistribusi *extreme value* tipe I dan independen untuk semua i dan j maka akan diperoleh model Logit yang mempunyai persamaan

$$\pi_{ik} = \frac{\exp(V_{ik})}{\sum_{j=1}^J \exp(V_{ij})} \quad \text{untuk } k=1,2,\dots,J$$

Model *Logit* akan cocok jika alternatifnya memenuhi sifat *Independence from Irrelevant Alternatives* (IIA). Oleh karena itu, model *Logit* tidak sesuai untuk kasus di mana antar alternatif/pilihan saling berkorelasi. Demikian juga pada data panel, pengukuran berulang pada subjek yang sama mengakibatkan responnya menjadi saling berkorelasi.

2. Jika diasumsikan $\varepsilon_{i,k}$ berdistribusi normal dengan mean nol maka akan diperoleh model Probit yang mempunyai persamaan

$$\pi_{ik} = \int_{-\infty}^{-V_{i1,k}} \dots \int_{-\infty}^{-V_{ij,k}} \phi(\varepsilon_{i,k}) d\varepsilon_{i1,k} \dots d\varepsilon_{ij,k}$$

Model Probit tidak mensyaratkan sifat IIA. Model Probit dapat mengakomodasi adanya korelasi antar alternatif, sehingga dapat diaplikasikan pada data panel. Kesulitan dalam model Probit adalah persamaan probabilitasnya memuat integral, sehingga dalam estimasi parameternya memerlukan proses simulasi (Mc Fadden dan Train, 2000).

Persamaan probabilitas (π_{ik}) memuat parameter q yang dapat diestimasi menggunakan metode MLE dengan berdasarkan data observasi (Y_i, X_i, Z_i). Fungsi *likelihood* untuk parameter q ,

$$L(\theta) = \prod_{i=1}^n \pi_{i1}^{y_{i1}} \dots \pi_{ij}^{y_{ij}} .$$

MLE adalah nilai θ yang memaksimumkan fungsi $\log L(\theta)$, atau

$$\hat{\theta}_M = \arg \max_{\theta} \log L(\theta) = \arg \max_{\theta} \sum_{i=1}^n \sum_{j=1}^J y_{ij} \log[\pi_{ij}] .$$

Dari nilai $\hat{\theta}$, digunakan untuk menghitung persamaan probabilitas $\hat{\pi}_j$. Sifat asimtotis penaksir MLE dapat digunakan untuk menguji masing-masing penaksir. Diharapkan, jika $y_i = k$ maka $\hat{\pi}_k \rightarrow 1$ dan $\hat{\pi}_j \rightarrow 0 \quad \forall j \neq k$.

7.2 Uji goodness of fit

Dalam DCM, nilai utilitas U_{ij} tidak terobservasi, sehingga uji *goodness of fit* didasarkan pada observasi Y_i dan nilai $\hat{\pi}_i = (\hat{\pi}_{i1}, \dots, \hat{\pi}_{ij})$. Dari observasi Y_i dapat diperoleh nilai $p_i = (p_{i1}, \dots, p_{ij})$. Misalkan dari n responden dapat dikelompokkan menjadi n' yang mempunyai kesamaan sifat (nilai variabel X_i dan Z_{ij}) dan n_i adalah banyaknya responden dalam kelompok i , maka

$$p_{ij} = n_{ij}/n_i.$$

Jika n_i cukup besar, maka uji asumsi distribusi (p_i mempunyai distribusi sama dengan distribusi dari $\hat{\pi}_i$) dapat dilakukan menggunakan *Kolmogorov-Smirnov test*. Dari persamaan probabilitas, $\hat{\pi}_k$ merupakan distribusi kumulatif. Pada model Probit, $\hat{\pi}_i$ berdistribusi normal. Pada model Logit, $\hat{\pi}_i$ berdistribusi *extreme value* tipe I. Hipotesisnya adalah

$$H_0 : F_X(x) = F_Y(x) \text{ vs } H_0 : F_X(x) \neq F_Y(x)$$

$$H_0 \text{ ditolak jika } D = \max\{D_+, D_-\} > d_{m,n,\alpha}$$

$$D_+ = \max_x (F_X(x) - F_Y(x))$$

$$D_- = \max_x (F_Y(x) - F_X(x))$$

dengan

$$F_X(x_i) = p_{ij} \text{ dan } F_Y(x_i) = \hat{\pi}_j$$



Pada $\alpha = 0.05$, $d_{m,n,\alpha} = 1.36 \sqrt{\frac{m+n}{mn}}$. m dan n masing-masing merupakan banyaknya sampel dalam sampel *random* X dan Y . Jika $m=n$ maka $d_{m,n,\alpha} = 1.36 \sqrt{\frac{2}{n}}$.

Untuk mendapatkan n_i cukup besar hanya mungkin dilakukan pada observasi mengenai pendapat responden. Untuk mengukur kecocokan model, statistik yang dapat menggunakan:

1. *Pearson Chi-Square Test*

$$\chi^2 = \sum_{i=1}^{n'} \sum_{j=1}^J \frac{(p_{ij} - \hat{\pi}_{ij})^2}{(1 - \hat{\pi}_{ij}) \hat{\pi}_{ij} / n_i} \sim \chi_{(n'-r)}^2, r \text{ adalah banyaknya parameter.}$$

2. *Devians*

$$D = 2 \sum_{i=1}^{n'} \sum_{j=1}^J \left\{ n_i p_{ij} \log \left(\frac{p_{ij}}{\hat{\pi}_{ij}} \right) + n_i (1 - p_{ij}) \log \left(\frac{1 - p_{ij}}{1 - \hat{\pi}_{ij}} \right) \right\}$$

D berdistribusi $\chi_{(n'-r)}^2$, r adalah banyaknya parameter.

3. *Hosmer-Lemeshow Test*

Dari n responden dikelompokkan menjadi 10 kelompok berdasarkan nilai $\hat{\pi}_{ij}$ pada masing-masing alternatif. Kelompok I berisi responden dengan $\hat{\pi}_{ij} \leq 0.1$ dan Kelompok ke 10 berisi responden dengan $\hat{\pi}_{ij} > 0.9$. Statistik ujinya adalah

$$\chi^2 = \sum_{k=1}^{10} \frac{(o_{kj} - n_k \hat{\pi}_{kj})^2}{n_k (1 - \hat{\pi}_{kj}) \hat{\pi}_{kj}}, \hat{\pi}_{kj} = \sum \frac{\hat{\pi}_{kj}}{n_k}$$

o_{kj} adalah banyaknya responden yang memilih alternatif j pada kelompok k .

$\hat{\pi}_{kj}$ adalah rata-rata probabilitas responden memilih alternatif j pada kelompok k .

Uji kecocokan model yang lain adalah statistik *Pseudo R²* yang identik dengan nilai R^2 (koefisien deterministik) (Mc Fadden, 1974)

$$\text{pseudo } R^2 = 1 - \frac{D_1}{D_0}$$

D_1 merupakan nilai devian model yang diuji, sedangkan D_0 adalah nilai devians untuk semua parameter sama dengan nol. Jika model secara sempurna memprediksi nilai Y_{ij} maka $\log L(\hat{\theta}) = 0$ (atau nilai deviansnya nol). Sehingga nilai maksimum dari pseudo R^2 adalah satu. Pemasalahan dalam penggunaan pseudo R^2 ini adalah tidak adanya kaidah untuk menyatakan pada nilai berapa sedemikian hingga model dikatakan baik. Permasalahan kedua adalah peningkatan nilai pseudo R^2 pada penambahan variabel independen tidak dapat menjelaskan seberapa penting variabel tersebut (Koppelman dan Chandra, 2006). Oleh karena itu, untuk menjelaskan seberapa penting suatu variabel lebih baik menggunakan nilai signifikan masing-masing variabel independennya (Maddala, 1983).

7.3 Studi kasus

Data mengenai pengaruh lama pendampingan dalam bulan (X) terhadap kesuksesan mengurangi kenakalan remaja. Diperoleh data pada tabel 1 sebagai berikut

Tabel 7.1. Data hasil pendampingan.

| X_i | Jumlah anak (n_i) | Jumlah Sukses (y_i) | p_i |
|-------|-----------------------|-------------------------|-------|
| 6 | 38 | 18 | 0.47 |
| 7 | 42 | 23 | 0.55 |
| 8 | 36 | 22 | 0.61 |
| 9 | 30 | 19 | 0.63 |
| 10 | 48 | 32 | 0.67 |
| 11 | 45 | 32 | 0.71 |
| 12 | 40 | 30 | 0.75 |

Dapat disusun persamaan *representative utility*

$$V_i = \beta_0 + \beta_1 X_i$$

Berdasarkan model Probit diperoleh

$$\hat{\beta}_0 = -0.70946 \text{ dan } \hat{\beta}_1 = 0.11582$$

Hasil estimasi probabilitas sukses dari model Probit disajikan pada tabel 7.2.

Tabel 7.2. Nilai probabilitas model *probit*

| X_i | $\hat{\pi}_i$ | $n_i \hat{\pi}_i$ | $ p_i - \hat{\pi}_i $ | $(y_i - n_i \hat{\pi}_i)$ |
|-------|---------------|-------------------|-----------------------|---------------------------|
| 6 | 0.494 | 18.780 | 0.0242 | -0.780 |

| X_i | $\hat{\pi}_i$ | $n_i \hat{\pi}_i$ | $ p_i - \hat{\pi}_i $ | $(y_i - n_i \hat{\pi}_i)$ |
|-------|---------------|-------------------|-----------------------|---------------------------|
| 7 | 0.540 | 22.694 | 0.0097 | 0.306 |
| 8 | 0.586 | 21.094 | 0.0241 | 0.906 |
| 9 | 0.630 | 18.912 | 0.0004 | 0.088 |
| 10 | 0.673 | 32.313 | 0.0132 | -0.313 |
| 11 | 0.714 | 32.122 | 0.0039 | -0.122 |
| 12 | 0.752 | 30.075 | 0.0019 | -0.075 |

Selanjutnya, dengan uji *Kolmogorov-Smirnov*

$$D = 0.02407 \text{ dan } d_{(7,7,0.05)} = 1.36 \sqrt{2/7} = 0.726951$$

Sehingga dapat disimpulkan H_0 tidak ditolak, yang berarti bahwa asumsi distribusi normal dapat diterima. Jika didasarkan pada statistik *Pearson Chi Square* diperoleh $c^2 = 0.18$, maka dapat disimpulkan bahwa model Probit sesuai dengan data.

Dari data pada tabel 1, jika digunakan model Logit maka diperoleh

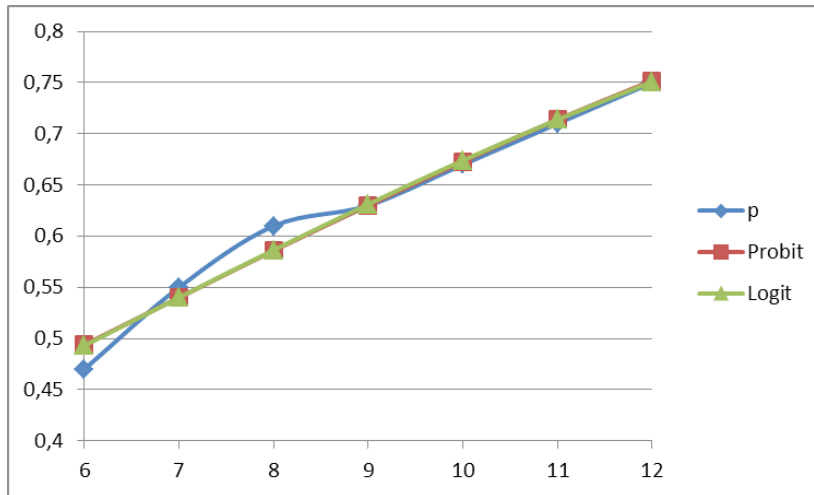
$$\hat{\beta}_0 = -1.15774 \text{ dan } \hat{\beta}_1 = 0.18841$$

Hasil estimasi probabilitas sukses dari model Logit disajikan pada tabel 3.

Tabel 7.3. Nilai Probabilitas Model Logit

| X_i | $\hat{\pi}_i$ | $n_i \hat{\pi}_i$ | $ p_i - \hat{\pi}_i $ | $(y_i - n_i \hat{\pi}_i)$ |
|-------|---------------|-------------------|-----------------------|---------------------------|
| 6 | 0.494 | 18.741 | 0.0232 | -0.741 |
| 7 | 0.540 | 22.688 | 0.0098 | 0.312 |
| 8 | 0.586 | 21.114 | 0.0235 | 0.886 |
| 9 | 0.630 | 18.940 | 0.0013 | 0.060 |
| 10 | 0.673 | 32.353 | 0.0140 | -0.353 |
| 11 | 0.714 | 32.129 | 0.0039 | -0.129 |
| 12 | 0.752 | 30.034 | 0.0009 | -0.034 |

Hasil prediksi menggunakan model *Probit* maupun *Logit* dapat disajikan dalam Gambar 7.1



Gambar 7.1. Prediksi menggunakan model *probit* maupun *logit*.

Berdasarkan uji *Kolmogorov-Smirnov*, diperoleh

$$D = 0.02349 \text{ dan } d_{(7,7,0.05)} = 1.36 \sqrt{2/7} = 0.726951$$

Sehingga dapat disimpulkan H_0 tidak ditolak, yang berarti bahwa asumsi distribusi *extreme value* tipe I dapat diterima. Jika didasarkan pada statistik Pearson Chi Square diperoleh $\chi^2 = 0.171$, maka dapat disimpulkan bahwa model Logit sesuai dengan data.

Koefisien regresi memberikan tanda efek parsial untuk masing-masing x_j pada respon probabilitas dan signifikansi x_i dihitung pada tingkat signifikansi yang ditentukan. Estimasi koefisien, standar error dan nilai fungsi *log-likelihood* dari model Logit dan Probit dapat dilihat dalam *output* beberapa *software* statistik seperti Minitab, SPSS dan Splus. Ukuran kecocokan model biasanya dilaporkan dengan istilah *percent correctly predicted*, yang dihitung sebagai berikut. Untuk setiap i , kita menghitung estimasi probabilitas pada y_i yang mempunyai nilai satu. Jika nilai $p(x)$ lebih dari 0.5, maka prediksi y_i adalah satu dan jika nilai $p(x)$ kurang dari 0.5, prediksi y_i adalah null. Persentase dihitung dari banyaknya prediksi y_i yang sesuai dengan nilai aktual y_i . Ukuran ini sedikit bermanfaat, tetapi mungkin akan mempunyai nilai persentase yang besar walaupun tanpa memperhatikan modelnya. Misalnya dari sampel berukuran 200, 180 observasi mempunyai nilai $y_i=0$ dan dari aturan prediksi diperoleh 150 nilai prediksi y adalah null. Walaupun tidak satupun prediksi y pada nilai satu yang tepat, kita masih memprediksikan 70% prediksi adalah benar. Oleh karena itu, lebih baik menghitung presentasi pada nilai $y=0$ dan $y=1$.

Pada model Probit, $p(0) \approx 0,4$ dan untuk model Logit, $p(0) = 0,25$. Sehingga estimator koefien dari model Logit sama dengan 1,6 (atau $0,4/0,25$) kali koefisien model Probit, atau koefisien model Probit sama dengan 0,625 kali koefisien model Logit. Pada model linear, $g(0) = 1$, sehingga

koefisien model *Linear* = 0,25 kali koefisien model *Logit* dan

koefisien model *Probit* = 2,5 kali koefisien model *Linear*

Perbandingan yang lebih akurat adalah dengan mengalikan koefisien Probit dengan $\phi(\hat{\beta}_0 + \bar{x}\hat{\beta})$ dan mengalikan koefisien Logit dengan

$$\frac{\exp(\hat{\beta}_0 + \bar{x}\hat{\beta})}{[1 + \exp(\hat{\beta}_0 + \bar{x}\hat{\beta})]^2}.$$

Untuk menguji kesesuaian pemilihan distribusi yang digunakan dalam model DCM dapat dilakukan menggunakan statistik *Kolmogorov-Smirnov*, sedangkan untuk menguji kecocokan model dengan data dapat dilakukan menggunakan statistik *Pearson Chi Square*.

Referensi

- Agresti, A. (1990). *Categorical Data Analysis*. John Wiley and Son
- Bhat, C. R. (2001). Quasi-random maximum simulated likelihood estimation of the mixed multinomial logit model. *Transportation Research*, 35B(7), 677-695.
- Ben-Akiva, M., & M. Bierlaire. (1999). 'Discrete choice methods and their applications in short term travel decisions', in R. Hall, ed., *The Handbook of Transportation Science*, Kluwer, Dordrecht, The Netherlands, pp. 5-33.
- Bhat, C. R. (2003). Econometric choice formulation alternative model structures, estimation techniques, and emerging direction. *International Conference on Travel Behaviour Research Lucerna*.
- Bierlaire, M. (1998). Discrete choice models, in M. Labbe, G. Laporte, K. Tanczos, and P. Toint, eds., *Operations Research and Decision Aid Methodologies in Traffic and Transportation Management*, Springer-Verlag, Heidelberg, Germany, pp. 203-227.
- Bolduc, D. (1992). 'Generalized autoregressive errors: The multinomial Probit model. *Transportation Research B* 26. 155-170
- Bolduc, D. (1999). A practical technique to estimate multinomial Probit models in transportation. *Transportation Research, B* 33, 63-79.
- Bolduc, D., B. Fortin, & M. Fournier. (1996). The impact of incentive policies on the practice location of doctors: A multinomial Probit analysis. *Journal of Labor Economics*, 14, 703-732.
- Bunch, D., & R. Kitamura. (1989). Multinomial Probit estimation revisited: Testing new algorithms and evaluation of alternative model specification of household car ownership', Transportation Research Group Report UCD-TRG-RR-4, University of California, Davis.
- Chu, C. (1989). A paired combinational Logit model for travel demand analysis. *Proceedings of Fifth World Conference on Transportation Research*, 4, 295-309.
- Contoyannis, P. (2001). *An Introduction to Simulation-Based Estimation, Working Paper Department of Economics and Related Studies*, University of York.
- Cox, D.R., & Wermuth, N. (1991). A simple approximation for bivariate and normal integrals. *International Statistics Review* 59, 263-269
- Daganzo, C., F. Bouthelier, & Y. Sheffi. (1977). Multinomial probit and qualitative choice: A computationally efficient algorithm. *Transportation Science*, 11, 338-358.

- Davidson, & Russel. (1999). *Econometric Theory and Method*.
- Drezner, Z. (1992). *Computation of Multivariate Normal Integral, ACM Transactions on Mathematics Software, 18*. 450-460
- Drezner, Z. (1994). *Computation of the Trivariate Normal Integral, Mathematics of Computation, 62*, 289-294
- Drezner, Z., & Wesolowsky, G.O. (1989). On the computation of bivariate normal integral. *Journal of Statist. Comput. Simul, 35*. 101-107
- Forinash, C., & F. Koppelman. (1993). Application and interpretation of nested logit models of intercity mode choice. *Transportation Research Record, 1413*, 98-106.
- Genz, A. (1992). Numerical computation of multivariate normal probabilities. *J. Computational and Graphical Statist., 1*, 141-149.
- Genz, A. (1993). Comparison of methods for the computation of multivariate normal probabilities. *Computing Science and Statistics, 25*, 400-405.
- Geweke, J. (1996). *Monte Carlo Simulation and Numerical Integration*, in D. Kendrick and J. Rust, eds., *Handbook of Computational Economics*, (hal. 731-800). Amsterdam: Elsevier Science, 731-800.
- Greene W. (2005). *Econometrics Analysis*. 5 Editions. Prentice Hall
- Haaijer, M., M. Wedel, M. Vriens, & T. Wansbeek. (1998). Utility covariances and context effects in conjoint MNP models. *Marketing Science, 17*, 236-252.
- Hajivassiliou, V. A., & P. A. Ruud. (1994). *Classical Estimation Methods for LDV Models Using Simulation*, Ch. 40 in *Handbook of Econometrics*, Vol. 4, ed. R. F. Engle and D. L. McFadden, Amsterdam, Elsevier, 2383-441.
- Hajivassiliou, V., D. McFadden, & P. Ruud. (1996). Simulation of multivariate normal rectangle probabilities and their derivatives: Theoretical and computational results. *Journal of Econometrics, 72*, 85-134.
- Harris, M.N dan Macquarie L.R, & Siouclis AJ. (2000). Comparison of alternative estimators for binary panel probit models, *Melbourne Institute Working Paper* no 3/00
- Henningsen, & Arne. (2007). The Micecon Package, <http://www.r-project.org/>. diakses 12 Januari 2015
- Horowitz J.L. & Savin N.E. (2001). Binary response models: Logits, probits and semiparametrics. *Journal of Economic Perspectives, 15*, 43-56
- Horowitz, J., J. Sparmann, & C. Daganzo. (1982). An investigation of the accuracy of the Clark approximation for the multinomial Probit model. *Transportation Science 16*, 382-401



- Hosmer W, & L. Stanley. (1989). *Applied Logistic Regression*, John Wiley & Sons
- Joe, H. (1995). Approximations to multivariate normal rectangle probabilities based on conditional expectations. *J. Amer. Statist. Assoc.*, 90, 957-964.
- Karlstrom, A. (2001). Developing generalized extreme value models using the Piekands representation theorem', Working Paper, Infrastructure and Planning, Royal Institute of Technology, Stockholm, Sweden.
- Koppelman F.S., & Bhat Chandra. (2006). A Self Instructing Course in Mode Choice Modeling: Multinomial and Nested Logit Models, U.S. Department of Transportation Federal Transit Administration.
- Koppelman, F., & C. Wen. (2000). The paired combination Logit model: Properties, estimation and application. *Transportation Research*, 34, 75-89.
- Kosuke Imai, & David A. van Dyk. (2005). A Bayesian analysis of the multinomial Probit model using marginal data augmentation. *Journal of Econometrics*, 124, 311 - 334
- Maddala, G. S. (1983). *Limited Dependent and Qualitative Variables in Econometrics*. Cambridge: Cambridge University Press.
- McFadden, D. (1974). *Conditional Logit Analysis of Qualitative Choice Behavior*, dalam *Zarembka, Frontiers in Econometrics* (hal. 105-142). New York: Academic Press
- McFadden, D., & Train K. (2000). Mixed MNL models for discrete response. *Journal of Applied Econometrics*, 15(5), 447-470.
- Moro, B. (1995). The full Monte. *Risk* 8(2), 53-57.
- Munizaga, Heydecker, & Ortuzar. (1997). On error structure of discrete choice models. *Traffic Engineering and Control*, 38 (11), 593-597.
- Nugraha J, Haryatmi S., & Guritno S. (2006). Model discrete choice dan regresi logistik. *Prosiding Seminar Nasional MIPA, UNY*.
- Nugraha J., Haryatmi S., & Guritno S. (2007). Bias maximum likelihood estimator (MLE) dalam model multinomial logit pada respons saling berkorelasi. *Prosiding Seminar Nasional MIPA, UNY*
- Lee, B. (1999). Calling patterns and usage of residential toll service under self-selecting tariffs. *Journal of Regulatory Economics*, 16, 45-82.
- Paolella, & Marc S. (2006). *Fundamental Probability (A Computational Approach)*. John Wiley & Sons Ltd,
- Prentice. (1988). Correlated binary regression with covariates specific to each binary observation. *Biometrics*, 44, 1043-1048.

- Revelt, D. & Train K. (1998). Mixed logit with repeated choices: Households' choice of appliance efficiency level. *Review of Economics and Statistics* 80(4), 647-657
- Robert E. McCulloch, Nicholas G. Polson, & Peter E. Rossi. (2000). A bayesian analysis of the multinomial probit model with fully identified parameters. *Journal of Econometrics* 99, 173-193
- Rodriguez G. (2001). *Generalized Linear Models*. Princeton University
- Ruud P.A. (1996). *Approximation and Simulation of Multinomial Probit Model: An analysis of Covariance Matrix Estimation*. Berkeley: Working Paper University of California
- Schervish, M. J. (1984). Multivariate normal probabilities with error bound. *Appl. Statist.*, 33, 81-94.
- Tong, Y.L. (1990). *The Multivariate Normal Distribution*. New York: Springer-Verlag
- Train, K. (1986). *Qualitative Choice Analysis*. Cambridge: MIT Press
- Train, K. (2000). *Halton Sequences for Mixed Logit, Working Paper No. E00-278*. Berkeley: Department of Economics, University of California
- Train, Kenneth. (2003). *Discrete Choice Methods with Simulation*. Cambridge: UK Press
- Vovsha, P. (1997). *The cross-nested Logit model: Application to mode choice in the Tel Aviv metropolitan area'*, *Conference Presentation, 76th*. Washington: Transportation Research Board Meetings
- Walker J., Ben-Akiva M., Bolduc D. (2004). Identification of the Logit Kernel (or Mixed Logit) Model, Working Paper Massachusetts Institute of Technology.
- Wen, C.-H., & F. Koppelman. (2001). The generalized nested Logit model. *Transportation Research*, 35, 627-641.
- Yai, T., S. Iwakura, & S. Morichi. (1997). Multinomial Probit with structured covariance for route choice behavior. *Transportation Research*, 31, 195-207.

Glosarium

| | | |
|------------------------------|---|--|
| Choice set | : | himpunan semua pilihan/alternatif yang dapat dipilih oleh responden |
| Data | : | "keterangan" yang berhasil dicatat atau direkam mengenai suatu hal |
| Fungsi likelihood | : | merupakan probabilitas yang didefinisikan sebagai fungsi dari parameter |
| Maximum Likelihood Estimator | : | metode untuk mencari penaksir parameter yang dengan memaksimumkan fungsi likelihood berdasarkan data sampel |
| Logit | : | transformasi logaritma natural dari nilai odds. |
| matrik Hessian | : | matrik definit negative yang elemen-elemennya merupakan derivative kedua fungsi log likelihood |
| Model pemilihan diskrit | : | model yang menggambarkan pembuat keputusan memilih diantara alternatif yang tersedia |
| Parameter | : | sebarang nilai yang menjelaskan ciri populasi |
| Populasi | : | keseluruhan objek pengamatan yang menjadi perhatian yang jumlahnya bisa tak terhingga maupun terhingga |
| P-value | : | nilai probabilitas kesalahan tipe I terkecil sedemikian hingga dapat menolak hipotesis nol |
| representative utility | : | fungsi linear atas variabel-variabel dari karakteristik responden dan pilihan |
| Sampel | : | suatu himpunan bagian dari populasi |
| Statistik | : | sebarang nilai yang menjelaskan ciri sampel |
| Variabel | : | suatu karakteristik yang nilainya antar objek pengamatan atau antar waktu pada objek yang sama bisa berbeda beda |
| Variabel diskrit | : | variabel yang mempunyai kemungkinan nilai terbatas (anggota domainnya terbatas) |

- Variabel kontinu : variabel yang kemungkinan nilainya tak terhitung (bagian dari bilangan real)
- Variabel dependen : variabel yang nilainya dipengaruhi oleh variabel yang lain. yaitu variabel independen
- Variabel independen : variabel yang nilainya berpengaruh terhadap nilai variabel yang lain.
- Variabel random : Variabel yang nilai berupa bilangan real dan mengacu pada fungsi probabilitas yang merupakan observasi dari suatu eksperimen
- Ukuran asosiasi : Untuk mengetahui derajat keeratan hubungan antara dua faktor

Indeks Subjek

B

Bayesian 2, 147, 149

D

Densitas xii, 4, 5, 6, 7, 51, 54, 66, 73, 78, 79, 80, 84, 93, 97, 98, 103, 108, 112, 121, 129, 133

Distribusi normal ix, 121, 122

E

Estimasi vii, xi, 2, 5, 10, 19, 20, 23, 24, 27, 37, 51, 59, 77, 85, 89, 90, 137, 140, 141, 142

G

Gev 2, 37, 51

Ghk viii, 2, 66, 67, 69, 71, 121, 133

Goodness of fit 135, 137

I

lia 3, 9, 10, 20, 78, 136

K

Korelasi viii, xi, 3, 20, 24, 26, 27, 28, 30, 31, 37, 60, 61, 62, 78, 80, 81, 82, 83, 85, 86, 90, 124, 126, 129, 137

Kovariansi 21, 22, 23, 24, 26, 31, 51, 53, 59, 60, 61, 62, 63, 69, 97, 107, 116

L

Logit 152

Logit viii, ix, xi, 2, 5, 9, 10, 13, 17, 20, 37, 40, 66, 77, 78, 79, 82, 83, 84, 85, 86, 89, 90, 106, 107, 141, 142, 144, 147, 148, 149, 150, 152

M

Maksimum likelihood 2, 20

Matrik hessian 152

Maximum 2

Maximum likelihood estimator 152

Mixed logit 2, 77, 78, 79, 150

N

Nested 37, 105, 146, 149, 150

Newton-raphson 12, 71, 116, 117

P

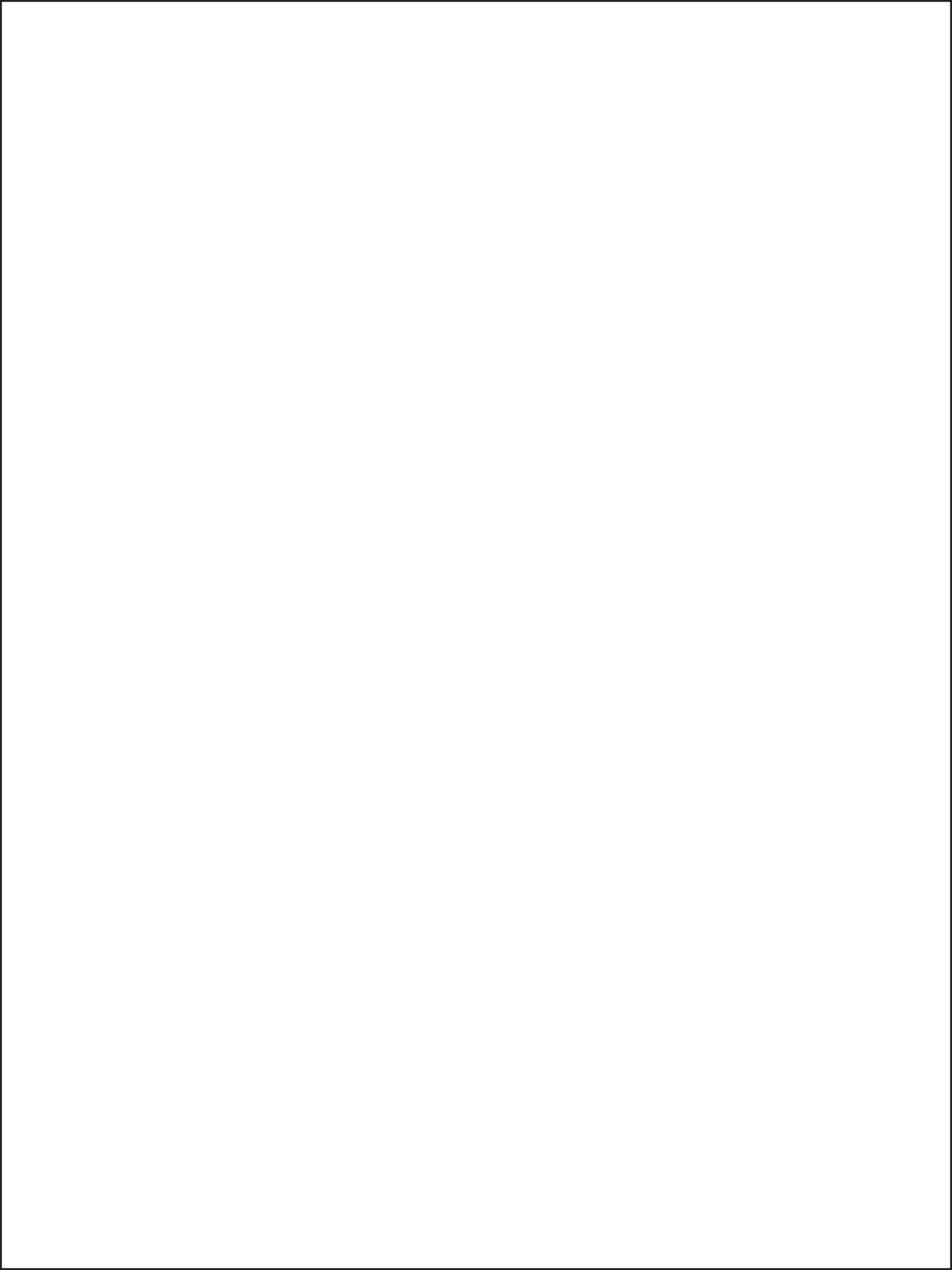
Parameter vii, xi, xii, 2, 5, 10, 11, 12, 13, 14, 17, 19, 20, 21, 23, 24, 26, 27, 28, 31, 32, 33, 34, 35, 36, 37, 51, 56, 57, 59, 60, 61, 62, 64, 65, 71, 74, 79, 80, 81, 83, 85, 86, 87, 88, 89, 90, 93, 94, 95, 98, 101, 102, 103, 104, 105, 109, 112, 113, 114, 118, 132, 135, 137, 138, 139, 152
Probabilitas ix, xi, 5, 7, 9, 31, 51, 52, 53, 65, 78, 79, 107, 126, 132, 140, 141
Probit viii, xi, 2, 3, 37, 51, 62, 65, 66, 71, 107, 121, 140, 141, 146, 149
P-value 152

R

Regresi xi, 1, 4, 13, 14, 15, 16, 17, 18, 85, 89, 135, 142

S

Simulasi vii, viii, 2, 20, 51, 65, 67, 68, 77, 79, 82, 90, 107, 108, 109, 110, 111, 115, 116, 121, 133, 137
Statistik wald 12, 113





UNIVERSITAS ISLAM INDONESIA

978-602-45001-9-1



9 786024 500191

Biografi :



Dr. Jaka Nugraha merupakan dosen tetap pada Program Studi Statistika Universitas Islam Indonesia (UII) sejak tahun 1995. Menyelesaikan S1 (1994), S2 (1999) dan S3 (2010) pada Program Studi Statistika Universitas Gadjah Mada. Disamping aktif mengajar di beberapa program studi di UII, juga aktif dalam kegiatan ilmiah di tingkat nasional maupun internasional. Sejumlah karya ilmiah telah dipublikasikan dalam beberapa jurnal nasional dan internasional. Saat ini menduduki jabatan sebagai Kepala Badan Pengembangan Akademik (BPA) di UII. Pengalaman mengajar dan penelitian di bidang analisis data kategorik memunculkan ide penulisan buku ini.

Alamat kantor : Program Studi Statistika, FMIPA-UII

Kampus Terpadu UII, jln Kaliurang KM 14,5

email

: jnugraha@uii.ac.id atau jk.nugraha@gmail.com