PROSIDING

onferensi Nasional Matematika XIII

24 - 27 Juli 2006

Editor: S. B. Waluya, dkk

MATIKA INDONESIA 1976

The Indonesian Mathematical Society (IndoMS)

Jurusan Matematika
FMIPA
Universitas Negeri Semarang

BADAN PENERBIT

ISBN: 979.704.457.2

ISBN : 979704457-2

PROSIDING

Konferensi Nasional Matematika XIII

Semarang, 24 27 Juli 2006

The Indonesian
Mathematical Society
(IndoMS)

Bekerjasama dengan

Jurusan Matematika
Fakultas Matematika dan Ilmu Pengetahuan Alam
Universitas Negeri Semarang

EDITOR:

S. B. Waluya
Zaenal Abidin
Muhammad Kharis
Riza Arifudin
David Mubarok
Alamsyah
Ary Woro Kurniasih
Putriaji Hendikawati

PENATA LETAK:

Zaenal Abidin

DESAIN COVER:

Zaenal Abidin

TEBAL BUKU:

883 + xvi

PENERBIT:

Jurusan Matematika
Fakultas Matematika dan Ilmu Pengetahuan Alam
Universitas Negeri Semarang
Kampus Sekaran, Gunungpati, Semarang

BEKERJASAMA DENGAN

Badan Penerbit Universitas Diponegoro

Cetakan pertama, Desember 2006
ISBN No. 979-704-457-2

Tim Penilai Makalah (*Reviewer*):

1. **Sri Wahyuni, Prof. Dr.**
   Jurusan Matematika, FMIPA – Universitas Gadjah Mada
2. **Christiana Rini Indrati, Dr., M.Si.**
   Jurusan Matematika, FMIPA – Universitas Gadjah Mada
3. **Widodo, Dr.**
   Jurusan Matematika, FMIPA – Universitas Gadjah Mada
4. **Lina Aryati, Dr.**
   Jurusan Matematika, FMIPA – Universitas Gadjah Mada
5. **Retantyo Wardoyo, Dr.**
   Jurusan Matematika, FMIPA – Universitas Gadjah Mada
6. **Pudji Astuti, Dr.**
   FMIPA Institut Teknologi Bandung
   Kelompok Keilmuan/Keahlian Aljabar
7. **A. Muchlis, Dr.**
   Jurusan Matematika, FMIPA – Institut Teknologi Bandung
8. **Irawati, Dr.**
   Jurusan Matematika, FMIPA – Institut Teknologi Bandung
9. **Hendra Gunawan, Prof., Ph.D.**
   Jurusan Matematika, FMIPA – Institut Teknologi Bandung
10. **Edy Tri Baskoro, Prof. Dr.**
    Jurusan Matematika, FMIPA – Institut Teknologi Bandung
11. **Robert Saragih, Dr.**
    Jurusan Matematika, FMIPA – Institut Teknologi Bandung
12. **Sutanawir Darwis, Dr.**
    Jurusan Matematika, FMIPA – Institut Teknologi Bandung
13. **Kiki Ariyanti Sugeng, Dr.**
    Jurusan Matematika, FMIPA – Universitas Indonesia
14. **Yusuf Fuad, Dr.**
    Jurusan Matematika, FMIPA – Universitas Negeri Surabaya
15. **Siti Khabibah, Dr.**
    Jurusan Matematika, FMIPA – Universitas Negeri Surabaya
16. **Mashadi, Dr.**
    Jurusan Matematika, FMIPA – Universitas Riau
17. **Edi Cahyono, Dr.**
    Jurusan Matematika, FMIPA – Universitas Haluoleo
18. **I Nyoman Budiantara, Prof.**
    Jurusan Matematika, FMIPA – Institut Teknologi Surabaya
19. **Sutarto Hadi, Dr.**
    Jurusan Matematika, FMIPA – Universitas Lambung Mangkurat
20. **Zulkardi, Prof. Dr.**
    Jurusan Matematika, FMIPA – Universitas Sriwijaya
21. **Yansen Marpaung, Prof. Dr.**
    Jurusan Pendidikan MIPA, FKIP – Universitas Sanata Dharma
22. **Amin Suyitno, Drs., M.Pd.**
    Jurusan Matematika, FMIPA – Universitas Negeri Semarang
23. **Stevanus Budi Waluya, Dr.**
    Jurusan Matematika, FMIPA – Universitas Negeri Semarang
24. **Sukestiyarno, Prof. Dr.**
    Jurusan Matematika, FMIPA – Universitas Negeri Semarang
25. **Sukirman, Drs., M.Pd.**
    Jurusan Matematika, FMIPA – Universitas Negeri Yogyakarta
26. **Budi Nurani, Dr.**
    Jurusan Matematika, FMIPA – Universitas Padjadjaran

27. **Toto Nusantoro, Dr.**
Jurusan Matematika, FMIPA – Universitas Negeri Malang
28. **Purwanto, Dr.**
Jurusan Matematika, FMIPA – Universitas Negeri Malang
29. **Swasono Raharjo, Dr.**
Jurusan Matematika, FMIPA – Universitas Negeri Malang
30. **Abdur Rahman As'ari, Drs., M.Pd., M.A.**
Jurusan Matematika, FMIPA – Universitas Negeri Malang
31. **Ipung Yuwono, Dr.**
Jurusan Matematika, FMIPA – Universitas Negeri Malang
32. **Akbar Sutawijaya, Prof. Dr.**
Jurusan Matematika, FMIPA – Universitas Negeri Malang
33. **Cholis Sa'dijah, Dr., M.Pd., M.A.**
Jurusan Matematika, FMIPA – Universitas Negeri Malang

SIFAT-SIFAT PENDUGA KUADRAT TERKECIL DAN PENDUGA KEMUNGKINAN MAKSIMUM PADA MODEL REGRESI LINEAR DENGAN BERBAGAI PROPORSI NILAI AMATAN NOL

(PROPERTIES OF LEAST SQUARE AND MAXIMUM LIKELIHOOD ESTIMATORS OF LINEAR REGRESSION MODEL WITH VARIOUS PROPORTION OF ZERO OBSERVATIONS)

Fitria Virgantari[1] dan Siswadi[2]

[1]Program Studi Matematika, FMIPA, Universitas Pakuan Bogor
E-mail: fitriav@yahoo.com
[2]Departemen Matematika, FMIPA, Institut Pertanian Bogor
Email: siswadi@inrr.org

Abstract. This article aims to study properties of ordinary least square (OLS) and maximum likelihood (ML) estimators of linear regression model which containing zero observations on dependent variable. This data structure is often met at biometrics or econometrics study (see [4]). The regression model to study is linear regression model with two dependent variables that is $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$, where ε drawn from standard normal distribution, and $\beta_0=0$, $\beta_1=1$ and $\beta_2=1$. Proportion of zero observations studied is 5%, 10%, 25%, 50% and 75%; each repeated 30 times from 100 random sample. The results show that coefficient of ML and OLS estimator is similiar at 5% until 10% of zero observations. The difference value of both estimators is greater after 10% proportion of zero. ML estimator is unbiased at 40% of zero observations, while OLS estimator is never unbiased with downward bias. In general can be concluded that bias of ML is smaller than OLS. Mean Square Error (MSE) of ML is also much smaller than OLS and even close to zero, so that can be said that ML estimator fulfill criterion of unbiased and consistent estimator. However, variance of OLS estimator is smaller than ML estimator after 25% proportion of zero. It indicates that OLS has higher of precision level but has lower accuracy level than ML estimator at model studied.

Keywords: OLS estimator, ML estimator, zero observations, bias, variance of estimator, MSE

Abstrak. Tulisan ini bertujuan untuk mengkaji sifat penduga suatu model regresi linear yang banyak mengandung nilai amatan nol (*zero observations)* pada peubah tak bebasnya. Struktur data ini sering dijumpai pada data biometrika atau ekonometrika (lihat [4]). Penduga yang dikaji adalah penduga kuadrat terkecil biasa (OLS/*Ordinary Least Square*) dan penduga kemungkinan maksimum (ML/*maximum likelihood*) yang memisahkan antara nilai amatan nol dan tidak. Model regresi yang akan dikaji adalah model regresi linear berganda dengan dua peubah bebas yaitu $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$ di mana ε dibangkitkan dari distribusi normal baku, serta nilai $\beta_0=0$, $\beta_1=1$ dan $\beta_2=1$. Proporsi nilai amatan nol yang dicobakan adalah 5%, 10%, 25%, 50% dan 75%; masing-masing diulang sebanyak 30 kali dari sampel acak berukuran 100. Hasil analisis menunjukkan bahwa nilai dugaan koefisien penduga ML dan OLS relatif sama pada proporsi nilai amatan nol 5% sampai 10%. Dugaan ML bersifat tak bias pada proposi nilai amatan nol 40%, sedangkan dugaan OLS tidak pernah tak bias pada semua proporsi nilai amatan nol dengan bias ke bawah. Perbedaan nilai kedua penduga semakin besar dengan semakin banyaknya proporsi nilai amatan nol. Namun secara umum dapat disimpulkan bahwa bias nilai dugaan ML jauh lebih kecil daripada OLS. Rata-rata kuadrat tengah galat (MSE) penduga ML juga jauh lebih kecil daripada OLS dan mendekati nol, sehingga dapat dikatakan bahwa penduga ML memenuhi kriteria sifat asimtotis, yaitu asimtotis tak bias dan konsisten.

Akan tetapi, keragaman penduga OLS lebih kecil dari penduga ML setelah proporsi nilai amatan nol 10%. Jadi secara umum dapat dikatakan bahwa penduga ML adalah penduga yang mempunyai tingkat ketepatan lebih tinggi daripada OLS, namun tingkat ketelitiannya lebih rendah dibandingkan OLS pada model yang dikaji.

Kata kunci: penduga OLS, penduga ML, nilai amatan nol, bias, ragam penduga, MSE

1. Pendahuluan

Analisis regresi merupakan analisis yang digunakan untuk menjelaskan pola hubungan antara peubah tak bebas dengan satu atau lebih peubah bebas dari segugus pengamatan. Analisis regresi telah menjadi salah satu hal yang pokok dalam menyelesaikan berbagai permasalahan, sehingga banyak tulisan dan penelitian mengenai regresi ini dikembangkan.

Analisis regresi pertama kali diperkenalkan oleh Galton (1822-1911), seorang ahli genetika. Galton membandingkan tinggi badan anak laki-laki dengan tinggi badan ayahnya dan menunjukkan bahwa tinggi badan anak laki-laki dari ayah yang tinggi setelah beberapa generasi cenderung mundur *(regressed)* mendekati nilai tengah populasi (Walpole [8]). Saat ini analisis regresi digunakan pada semua jenis peramalan, tidak harus berimplikasi regresi mendekati nilai tengah populasi. Selain untuk peramalan, analisis regresi juga digunakan untuk tujuan pemeriksaan peubah, spesifikasi model, serta pendugaan parameter.

Hubungan antara peubah bebas dan tak bebas dalam suatu model regresi bersifat stokastik, dalam arti bahwa hubungan tersebut tidak selalu menduga nilai yang sebenarnya dari peubah yang dijelaskan; jadi selalu ada faktor acak yang tidak terjelaskan dalam model hubungan tersebut, dan biasa dilambangkan dengan ε (sisaan/galat) (Kennedy [3]). Dalam model regresi diasumsikan bahwa ε merupakan suatu peubah acak yang berdistribusi normal dengan nilai tengah nol dan ragam σ^2, serta tidak berkorelasi (bebas satu sama lain). Apabila model yang dipostulatkan benar, maka faktor ε tadi akan menunjukkan kecenderungan yang mendukung asumsi yang berlaku.

Pendugaan parameter dalam persamaan regresi biasa dilakukan melalui metode kuadrat terkecil dengan asumsi kenormalan, kebebasan dan kehomogenan ragam. Namun, fenomena yang terjadi kadang menghasilkan respons yang berstruktur kontinu dengan kisaran yang mungkin sangat besar, sehingga menyebabkan timbulnya masalah heteroskedastisitas (ketidakhomogenan ragam). Hal ini akan berimplikasi pada metode pendugaan parameter dari model yang digunakan.

Pendugaan dengan metode kuadrat terkecil (*Ordinary Least Square/OLS*) akan cukup baik didekati apabila asumsi kenormalan, kebebasan dan kehomogenan ragam dipenuhi. Namun, semakin banyak nilai amatan nol pada data yang diperoleh akan menyebabkan timbulnya masalah heteroskedastisitas. Penggunaan metode OLS akan menghasilkan penduga yang berbias dan tidak konsisten karena asumsi yang mendasari tidak dipenuhi. Sedangkan penghilangan nilai amatan nol *(zero observations)* tersebut akan mengurangi ukuran sampel dan tidak mencerminkan keadaan yang sebenarnya. Metode pendugaan alternatif yang dapat dilakukan adalah metode kemungkinan maksimum (*Maximum Likelihood/ML*). Penduga dari metode ini didapatkan dengan memecah data dalam dua bagian (yang bernilai nol dan bukan) sehingga fungsi kepekatannya merupakan fungsi kepekatan bersama.

Tulisan ini bertujuan untuk mengkaji sifat penduga OLS dan ML pada model regresi dengan struktur data yang mengandung banyak nilai amatan nol pada peubah tak bebasnya. Proporsi nilai amatan nol yang dicobakan adalah 5%, 10%, 25%, 50% dan 75%. Sedangkan sifat penduga yang dikaji adalah sifat ketakbiasan, keragaman penduga, dan kuadrat tengah galat penduga, berdasarkan data bangkitan dari komputer.

2. Landasan Teori

2.1. Penduga Kuadrat Terkecil Biasa (*Ordinary Least Square/OLS*)

Dalam persamaan regresi linear, hubungan antara peubah tak bebas (Y) dengan peubah bebas ($X_1, X_2, \ldots, X_p$) dapat dirumuskan dalam bentuk persamaan :

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \ldots + \beta_p X_{pi} + \varepsilon_i \tag{1}$$

atau dalam bentuk matriks:

$$Y = \mathrm{X}\beta + \varepsilon \tag{2}$$

di mana X adalah matriks peubah bebas berukuran nxk, Y adalah vektor peubah tak bebas berukuran nx1, β adalah vektor parameter berukuran kx1, ε adalah vektor galat (sisaan) berukuran nx1; n adalah banyaknya pengamatan, dan k=p+1 adalah banyaknya parameter.

Penduga kuadrat terkecil merupakan penduga yang didapatkan dengan meminimumkan jumlah kuadrat sisaan atau :

$$\varepsilon'\varepsilon = Y'Y - 2\beta' X'Y + \beta' X'X\beta \tag{3}$$

Sebagai nilai dugaan, maka akan dipilih $\hat{\beta}$ sedemikian rupa sehingga nilai $\varepsilon'\varepsilon$ akan minimum. Caranya adalah dengan mendiferensialkan persamaan (3) terhadap β dan kemudian disamakan dengan nol, yaitu:

$$\frac{\partial(\varepsilon'\varepsilon)}{\partial\beta} = -2X'Y + 2X'X\beta = 0 \tag{4}$$

sehingga akan didapatkan:

$$\hat{\beta} = (X'X)^{-1}(X'Y) \tag{5}$$

dengan $X'X$ adalah matriks nonsingular (berpangkat penuh). Apabila matriks $X'X$ tidak berpangkat penuh, maka penduga β dicari dengan matriks kebalikan umum. Penduga tersebut bersifat tidak unik, dan solusi umumnya (Kshirsagar [5]) adalah:

$$\ddot{\beta} = \hat{\beta} + (I - H)z \tag{6}$$

di mana $H = S^{-}S$ adalah matriks idempoten berukuran pxp yang mempunyai sifat $H^2 = H$, $SH = S$, pangkat H = pangkat S = pangkat X = tr H; dan z adalah vektor sembarang; sedangkan

$$\hat{\beta} = S^{-}X'Y \tag{7}$$

di mana S^{-} adalah kebalikan umum dari $S = X'X$.

Penduga OLS merupakan penduga yang sering dipakai pada analisis regresi klasik. Hal ini berdasarkan pada kenyataan bahwa:

1. Penduga OLS relatif mudah dan selalu tersedia pada *software-software* statistika, bahkan pada kalkulator.
2. Penduga OLS dirancang dengan meminimumkan jumlah kuadrat sisaan, sehingga memenuhi kriteria kuadrat terkecil.
3. Penduga OLS merupakan penduga tak bias.
4. Penduga OLS merupakan penduga tak bias linear terbaik.
5. Penduga OLS memenuhi kriteria asimtotik.

2.2. Penduga Kemungkinan Maksimum (*Maximum Likelihood*/ML)

Dalam statistika klasik dianggap bahwa populasi adalah tunggal dan bisa membangkitkan banyak sampel acak; dalam hal ini parameter populasi tetap *(fixed)*. Sedangkan dalam metode ML dianggap bahwa sampel *fixed*, tetapi bisa dibangkitkan dari berbagai populasi induk, masing-masing dengan parameter tertentu. Jadi dalam hal ini parameter populasi adalah variabel; dan dari semua parameter tersebut yang dipilih adalah yang memberikan peluang maksimum bahwa populasinya membangkitkan sampel yang teramati.

Penduga ML merupakan penduga dengan sifat teoritis yang lebih kuat daripada penduga OLS; namun penduga ML tersebut mempunyai sifat-sifat asimtotik yang diperlukan, yaitu asimtotik tak bias, asimtotik efisien dan asimtotik berdistribusi normal. Dalam perhitungannya, penduga ini harus didasari dengan asumsi distribusi tertentu, misalnya distribusi normal. Apabila asumsi-asumsi dipenuhi, penduga β dari metode OLS dan ML akan menghasilkan nilai yang sama. Sedangkan penduga σ^2 yang dihasilkan metode ML bersifat bias, namun akan sama dengan penduga OLS dengan semakin besarnya ukuran sampel.

Metode ML yang digunakan untuk data dengan banyak pengamatan bernilai nol pertama kali diperkenalkan oleh Tobin [6]. Tobin menghubungkan studinya berdasarkan analisis probit, sehingga modelnya kemudian dikenal sebagai model Tobit. Prinsip dari metode ini adalah memecah data ke dalam dua bagian (yang bernilai nol dan bukan) sehingga fungsi kepekatannya merupakan fungsi kepekatan bersama.

Misalkan Y^* adalah peubah yang berdistribusi normal dengan ragam σ^2. Anggaplah $(y_1^*, y_2^*, \ldots, y_n^*)$ adalah sampel berukuran n dan data tercatat hanya pada nilai-nilai Y^* yang lebih besar dari nol. Untuk nilai-nilai $Y^* \leq 0$, dimasukkan nilai 0, atau:

$$Y = \begin{cases} Y^*, untuk\ Y^* \rangle 0 \\ 0, selainnya \end{cases} \tag{8}$$

di mana $Y_i^* = X_i\beta + \varepsilon_i$, β adalah vektor parameter berukuran kx1, X_i adalah vektor peubah bebas berukuran kx1 dan ε adalah sisaan yang merupakan peubah acak bebas dan berdistribusi normal dengan nilai tengah nol dan ragam σ^2; n adalah banyaknya pengamatan, dan k=p+1 adalah banyaknya parameter.

Sampel $y_1, y_2, \ldots, y_n$ disebut dengan sampel tersensor, sedangkan Y disebut dengan peubah tersensor (lihat [7]). Pada pengamatan $Y = 0$ yang berarti $Y^* \leq 0$, maka;

$$P(Y = 0) = P(Y^* \leq 0) \tag{9}$$

Secara teoritis, pendugaan parameter pada model regresi tersensor menurut Maddala (1983) dilakukan dengan memisahkan peng-amatan y_i yang sama dengan nol dan y_i yang lebih besar dari nol. Misalkan N_0 adalah banyaknya pengamatan di mana $y_i=0$ dan N_1 adalah banyaknya pengamatan di mana $y_i>0$, dan didefinisikan:

$$F_i = F(x_i'\beta, \sigma^2) = \int_{-\infty}^{x_i'\beta} \frac{1}{\sigma\sqrt{2\pi}} e^{-t^2/2\sigma^2} dt$$

$$f_i = f(x_i'\beta, \sigma^2 = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x_i'\beta}{\sigma}\right)^2} \tag{11}$$

$$\Phi_i = \int_{-\infty}^{x_i'\beta/\sigma} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt$$

$$\phi_i = \sigma f_i = \frac{1}{\sqrt{\;}} e^{-\frac{1}{2}\left(\frac{x_i' \beta}{\sigma}\right)^2}$$

Φ_i dan ϕ_i masing-masing adalah fungsi distribusi dan fungsi kepekatan normal baku yang dievaluasi pada $x_i' \beta / \sigma$. Misalkan pula $\gamma_i = \frac{\phi_i}{1-\Phi_i}$ dan

$Y_1' = (y_1, y_2, \ldots, y_{N_1})$ adalah vektor berukuran $1xN_1$ pada pengamatan y_i yang lebih besar dari nol

$X_1' = (x_1, x_2, \ldots, x_{N_1})$ adalah matriks berukuran kxN_1 bagi nilai x_i untuk y_i yang lebih besar dari nol

$X_0' = (x_{N_1+1}, \ldots, x_N)$ adalah matriks berukuran kxN_0 bagi nilai x_i untuk y_i sama dengan nol

$\gamma_0' = (\gamma_{N_1+1}, \ldots, \gamma_N)$ adalah vektor $1xN_0$ bagi nilai γ_i untuk nilai y_i sama dengan nol,

maka untuk pengamatan $Y=0$:

$$P(Y=0) = P(Y^* \le 0) = P(\varepsilon \le -x_i' \beta)$$

$$\int_{-\infty}^{-x_i'\beta} f(u)\,du = \int_{x_i'\beta}^{\infty} f(u)\,du$$

$$= 1 - F(x_i' \beta, \sigma^2) = (1-F_i)$$

Berdasarkan (9), maka $P(Y=0) + P(Y>0) = 1$, sehingga fungsi kepekatan peluang dari y_i adalah:

$$f(y_i) = (1-F_i) I_{(0)}(y_i) + \frac{1}{\sqrt{2\pi\sigma^2}} exp\left\{-\frac{1}{2\sigma^2}\left(y_i - x_i'\beta\right)^2\right\} I_{(0,\infty)}(y_i) \tag{15}$$

dengan I adalah fungsi indikator yang didefinisikan $I_A(y_i) = \begin{cases} 1, y_i \in A \\ 0, y_i \notin A \end{cases}$

Dalam model regresi linear, fungsi *likelihood* merupakan fungsi kepekatan peluang bersama dari peubah tak bebas Y_i. Karena faktor sisaan merupakan peubah acak yang saling bebas, maka Y_i juga merupakan peubah acak bebas, sehingga fungsi kemungkinan bagi persamaan (15) adalah:

$$L = \prod_{i=1}^{n} f(y_i) = \prod_0 (1-F_i) \prod_1 \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{y_i - x_i'\beta}{\sigma}\right)^2} \tag{17}$$

di mana suku pertama meliputi N_0 pengamatan untuk $y_i=0$ dan suku kedua untuk N_1 pengamatan pada $y_i>0$, dan log fungsi kemungkinannya adalah:

$$\log L = \sum_0 \log(1-F_i) + \sum_1 \log\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right) - \sum_1 \frac{1}{2\sigma^2}\left(y_i - x_i'\beta\right)^2 \tag{18}$$

di mana $\sum_0$ meliputi N_0 pengamatan untuk $y_i=0$ dan $\sum_1$ untuk N_1 pengamatan pada $y_i>0$.

Penduga parameter didapatkan dengan menggunakan turunan pertama dari *log L* terhadap β dan σ^2 yaitu:

$$\frac{\partial \log L}{\partial \beta} = -\sum_{0} \frac{1}{(1-F_i)} \frac{\partial F_i}{\partial \beta} - \frac{1}{\sigma^2} \sum_{1} (y_i - x_i'\beta)(-x_i) = 0$$

Yang akan menhghasilkan:

$$\hat{\beta} = (X_1'X_1)^{-1} X_1'Y - \sigma (X_1'X_1)^{-1} X_0'\gamma_0$$

$$= \hat{\beta}_{LS} - \sigma (X_1'X_1)^{-1} X_0'\gamma_0$$

di mana $\hat{\beta}_{ls}$ adalah penduga OLS dari β yang didapatkan dari N_1 pengamatan bukan nol.

2.4. Sifat-sifat Penduga

Penduga *(estimator)* adalah fungsi dari contoh acak sedangkan dugaan *(estimate)* adalah nilai dari penduga (Casella and Berger [1]). Ada banyak cara yang dapat digunakan untuk mencari penduga dari segugus data sampel, namun hanya sedikit yang merupakan penduga yang baik, dan dari penduga-penduga yang baik tersebut, hanya ada satu yang merupakan penduga terbaik. Ada beberapa kriteria yang dapat dipakai untuk mendapatkan penduga yang baik tersebut, di antaranya adalah sifat tak bias, ragam minimum, minimum MSE, konsisten serta sifat asimtotis.

Tak bias

Bias suatu penduga didefinisikan sebagai selisih antara nilai harapan dengan nilai parameter yang sebenarnya, atau:

$$Bias = E\left(\hat{\beta}\right) - \beta$$

Jadi $\hat{\beta}$ dikatakan sebagai penduga tak bias dari β jika nilai harapan dari $\hat{\beta}$ sama dengan nilai β atau $E\left(\hat{\beta}\right) = \beta$; sebaliknya jika $E\left(\hat{\beta}\right) \neq \beta$ maka dikatakan penduga tersebut berbias. Suatu penduga dikatakan tak bias jika nilai biasnya sama dengan nol. Hal ini berarti bahwa penduga tak bias akan konvergen pada nilai parameter dengan semakin meningkatnya ukuran sampel.

Ragam minimum

Penduga $\hat{\beta}$ dikatakan sebagai penduga terbaik apabila ragam dari $\hat{\beta}$ mempunyai nilai paling kecil (minimum) dari penduga-penduga tak bias lainnya, atau:

$$E\left[\hat{\beta} - E\left(\hat{\beta}\right)\right]^2 \langle E\left[\bar{\beta} - E\left(\bar{\beta}\right)\right]^2 \text{ atau}$$

$$Var\left(\hat{\beta}\right) \langle Var\left(\bar{\beta}\right)$$

di mana $\bar{\beta}$ adalah penduga lain dari parameter β

Suatu penduga $\hat{\beta}$ dikatakan bersifat efisien apabila mempunyai sifat tak bias dan ragam minimum dari penduga-penduga tak bias lainnya, atau mempunyai kedua sifat yang telah disebutkan di atas.

Minimum MSE

Kriteria ini juga merupakan kombinasi dari sifat ketakbiasan dan ragam minimum. Suatu penduga mempunyai minimum MSE jika nilai harapan dari kuadrat selisih penduga di sekitar nilai parameter populasinya mempunyai nilai paling kecil; atau dirumuskan sebagai:

$$MSE = E\left(\hat{\beta} - \beta\right)^2 = Var\left(\hat{\beta}\right) + \left(Bias\left(\hat{\beta}\right)\right)^2$$

Sifat asimtotis

Sifat ini dimaksudkan untuk melihat perilaku dari distribusi sampling penduga dalam ukuran sampel yang cukup besar. Distribusi sampling penduga pada umumnya berubah dengan berubahnya ukuran sampel. Sifat-sifat asimtotis yang diinginkan dari suatu penduga adalah asimtotis tak bias, konsisten dan asimtotis efisien.

a. $\hat{\beta}$ adalah penduga tak bias asimtotis dari β jika

$$\lim_{n\to\infty} E\left(\hat{\beta}\right) = \beta$$

b. Jika $\lim_{n\to\infty} MSE\left(\hat{\beta}\right) = 0$

maka $\hat{\beta}$ adalah penduga yang konsisten dari β

c. $\hat{\beta}$ merupakan penduga yang asimtotis efisien dari β jika memenuhi ketiga syarat berikut:

- $\hat{\beta}$ mempunyai distribusi asimtotis dengan nilai tengah dan ragam tertentu
- $\hat{\beta}$ konsisten
- Tidak ada penduga konsisten lain dari β yang mempunyai ragam asimtotis yang lebih kecil dari $\hat{\beta}$

3. Contoh Kasus dan Pembahasan

Untuk mengkaji perbedaam antara model regresi biasa dengan model regresi tersensor, berikut ini akan disajikan hasil analisis terhadap data bangkitan sebanyak 100 pengamatan dengan proporsi nilai amatan nol 5%, 10%, 25%, 50% dan 75%.

Data yang dibangkitkan meliputi peubah X_1 dan X_2 serta ε yang masing-masing bebas dan dibangkitkan dari distribusi normal baku (dengan nilai tengah nol dan ragam 1). Sedangkan Y dibangkitkan sesuai dengan model:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon \qquad (27)$$

dengan nilai $\beta_0 = 0$, $\beta_1 = 1$ dan $\beta_2 = 1$. Data yang dibangkitkan berukuran 100 dan diulang sebanyak 30 kali. Ke-30 nilai dugaan tersebut dipandang sebagai distribusi sampling dari $\hat{\beta}$ dan selanjutnya digunakan untuk mencari bias, ragam dan MSE dari nilai dugaan yang diperoleh dengan cara berikut.

Nilai bias dugaan

Nilai harapan dari distribusi sampling $\hat{\beta}$ dihitung berdasarkan nilai rata-rata dari seratus nilai penduga yang didapatkan dari hasil simulasi, yaitu:

$$E\left(\hat{\beta}\right)=\bar{\hat{\beta}}=\frac{\sum_{i=1}^{N}\hat{\beta}_i}{N}$$

Sedangkan nilai bias dari $\hat{\beta}$ dihitung dengan mengurangkan nilai β yang sebenarnya dengan nilai harapannya:

$$\hat{Bias}=\hat{\beta}-\beta$$

Ragam dugaan

Ragam dari distribusi sampling $\hat{\beta}$ diduga dengan menggunakan rumus klasik untuk menduga ragam yaitu:

$$Var\left(\hat{\beta}\right)=\frac{\sum_{i=1}^{N}\left(\hat{\beta}-\bar{\hat{\beta}}\right)^2}{\quad}$$

Kuadrat tengah galat (Mean Square Error/MSE)

MSE dari $\hat{\beta}$ diduga berdasarkan rata-rata dari kuadrat selisih antara penduga dengan nilai sebenarnya yaitu:

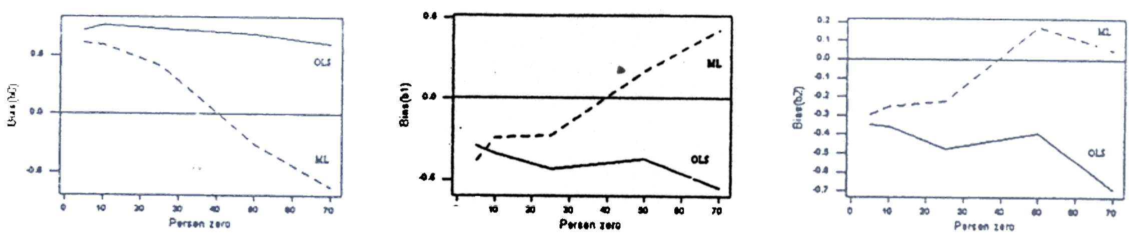
$$\hat{MSE}=\frac{\sum_{i=1}^{N}\left(\hat{\beta}-\beta\right)^2}{N}$$

Pembangkitan (simulasi data), pendugaan dan analisis dilakukan dengan bantuan *software* Eviews 4 dan Minitab 14. Hasil pendugaannya tercantum pada Tabel 1 berikut ini.

Tabel Nilai rata-rata dugaan parameter model regresi

% zero	Dugaan OLS			Dugaan ML		
	β_0	β_1	β_2	β_0	β_1	β_2
5	0.7251	0.7116	0.6482	0.6140	0.7919	0.7057
10	0.7710	0.6613	0.6418	0.6034	0.7609	0.7481
25	0.7448	0.5637	0.5229	0.4221	0.7726	0.7759
50	0.6938	0.6300	0.6077	-0.2562	1.1732	1.1750
75	0.6043	0.4521	0.3107	-0.6262	1.4313	1.0505

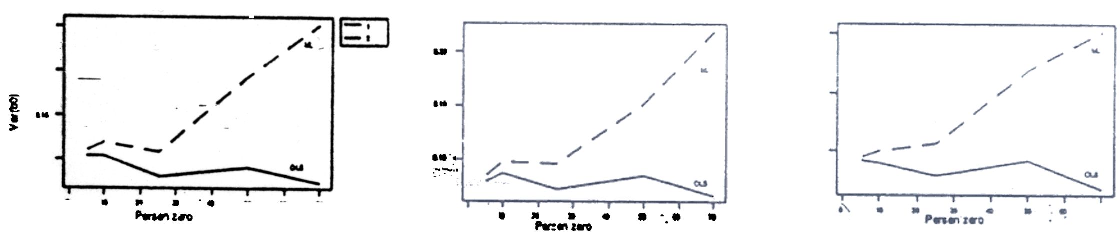
Berdasarkan tabel tersebut terlihat bahwa secara umum, nilai koefisien dugaan OLS dan ML relatif sama pada proporsi nilai amatan nol 5% sampai 10%. Perbedaan nilai dugaan keduanya semakin besar dengan semakin banyaknya proporsi nilai amatan nol. Hal ini dapat dilihat pula pada Gambar 1. Berdasarkan gambar tersebut terlihat pula bahwa bias koefisien β_0 dugaan OLS lebih besar daripada dugaan ML (Gambar 1a), sedangkan koefisien β_1 dan β_2 dugaan ML lebih besar daripada dugaan OLS. Pada Gambar 1a, 1b dan 1c terlihat bahwa bias dugaan koefisien β_0, β_1 dan β_2 metode ML sama dengan nol (tak bias) pada proporsi nilai amatan nol 40%, sedangkan dugaan OLS selalu berbias pada semua proporsi nilai amatan nol. Nilai dugaan ML cenderung berbias ke atas setelah proporsi nilai amatan nol 40%. Nilai bias tiap koefisien tersebut dapat dilihat pada Lampiran 1.



a. Bias (β_0)　　b. Bias (β_1)　　c. Bias (β_2)

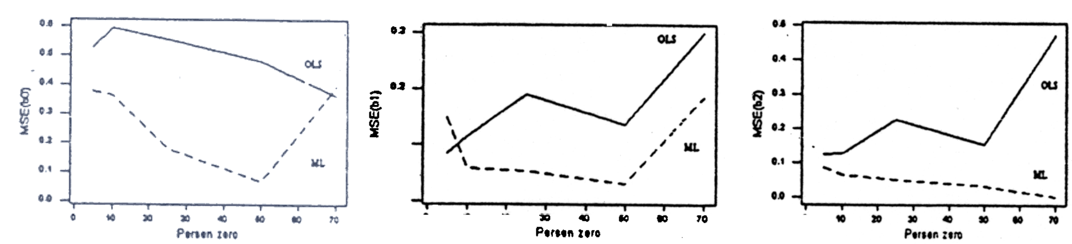
Gambar 1. Rata-rata bias dugaan $\beta_0, \beta_1, \beta_2$ metode OLS dan ML pada berbagai proporsi nilai amatan nol

Nilai ragam dugaan dan MSE dari metode OLS dan ML dapat dilihat pada Lampiran 2 dan 3; sedangkan gambaran grafisnya dapat dilihat pada Gambar 2 dan 3 di bawah ini.



a. Var(β_0)　　b. Var(β_1)　　c. Var(β_2)

Gambar 2. Ragam penduga $\beta_0, \beta_1, \beta_2$ metode OLS dan ML pada berbagai proporsi nilai amatan nol



a. MSE (β_0)　　b. MSE (β_1)　　c. MSE (β_2)

Gambar 3. MSE dugaan $\beta_0, \beta_1, \beta_2$ metode OLS dan ML pada berbagai proporsi nilai amatan nol

Berdasarkan Gambar 2 (a), (b) dan (c) terlihat bahwa pada proporsi nilai amatan nol sampai dengan sekitar 10%, ragam dugaan β_0, β_1, dan β_2 metode ML dan OLS bisa dikatakan sama. Namun dengan semakin banyaknya proporsi nilai amatan nol, ragam dugaan ML semakin besar dan lebih besar daripada dugaan OLS. Hal ini menunjukkan bahwa penduga OLS lebih efisien daripada penduga ML atau dapat dikatakan juga bahwa tingkat ketelitian penduga OLS lebih baik daripada penduga ML.

Sedangkan berdasarka Gambar 3 (a), (b), dan (c) terlihat bahwa nilai MSE dugaan ML jauh lebih kecil daripada OLS bahkan mendekati nol pada β_2. Hal ini menunjukkan bahwa penduga ML merupakan penduga yang lebih konsisten daripada OLS.

4. Simpulan

Berdasarkan hasil analisis yang telah dilakukan maka dapat disimpulkan bahwa sampai dengan 10% proporsi nilai amatan nol, nilai dugaan OLS dan ML relatif sama. Nilai dugaan ML tak bias pada proporsi nilai amatan nol 40%, dan sesudah itu penduga ML cenderung berbias ke atas; namun secara umum dapat dikatakan bahwa penduga ML jauh lebih tepat daripada penduga OLS, dengan kata lain lebih bersifat takbias. Kuadrat tengah galat yang lebih kecil dan lebih mendekati nol menunjukkan bahwa penduga ML lebih konsisten secara asimtotis. Sedangkan penduga OLS mempunyai ragam penduga yang jauh lebih kecil daripada penduga ML, yang menunjukkan bahwa penduga OLS mempunyai tingkat ketelitian yang lebih tinggi daripada penduga ML.

Daftar Pustaka

[1] Casella, G. and R. L. Berger. 1990. *Statistical inference.* California: Wadsworth Inc.

[2] Heien, D. and C. R. Wessells. 1990. Demand system estimation with microdata: A censored regression approach. *Journal of Business & Economic Statistics* Vol. 8 (3), 365-371.

[3] Kennedy, P. *A Guide to econometrics.* 1996. Cambridge: The MIT Press.

[4] Maddala, G. S. 1983. *Limited dependent and qualitative variables in econometrics.* New York, University Press.

[5] Kshirsagar, A. M. 1983. *A course in linear models.* New York : Marcel Dekker, Inc.

[6] Tobin, J. 1958. Estimation of relationships for limited dependent variables. *Econometrica* Vol. 26, 24-36.

[7] Virgantari, F. dan Siswadi. 2005. Model regresi tersensor kiri pada nilai amatan nol. *Prosiding Seminar Nasional Matematika 30 Juli 2005 (dalam proses penerbitan).* Universitas Indonesia. Jakarta.

[8] Walpole, R. E. 1982. *Pengantar Statistika.* Jakarta : Gramedia Pustaka Utama.

A. Lampiran 1. Nilai rata-rata bias dugaan OLS dan ML pada berbagai proporsi zero

% zero	Dugaan OLS			Dugaan ML		
	Bias(β_0)	Bias(β_1)	Bias(β_2)	Bias(β_0)	Bias(β_1)	Bias(β_2)
5	0.7251	-0.2884	-0.3518	0.6140	-0.3860	-0.2943
10	0.7710	-0.3387	-0.3582	0.6034	-0.2391	-0.2519
25	0.7448	-0.4363	-0.4771	0.4221	-0.2274	-0.2241
50	0.6938	-0.3700	-0.3923	-0.2562	0.1732	0.1750
75	0.6043	-0.5479	-0.6893	-0.6262	0.4313	0.0505

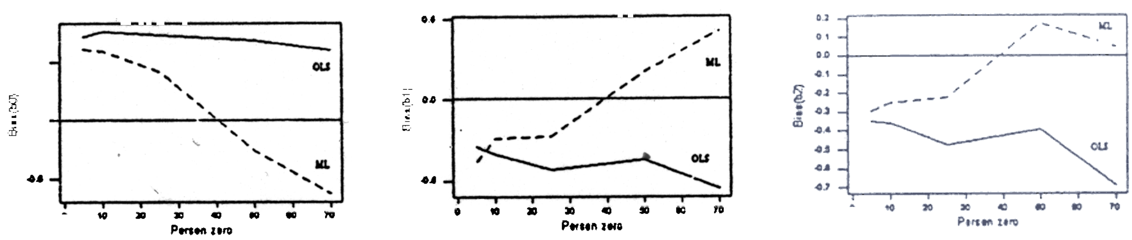
B. Lampiran 2. Nilai ragam dugaan OLS dan ML pada berbagai proporsi zero

% zero	Dugaan OLS			Dugaan ML		
	Var(β_0)	Var(β_1)	Var(β_2)	Var(β_0)	Var(β_1)	Var(β_2)
5	0.107	0.079	0.093	0.113	0.085	0.096
10	0.107	0.087	0.091	0.121	0.097	0.101
25	0.082	0.072	0.081	0.110	0.096	0.107
50	0.093	0.084	0.093	0.194	0.152	0.170
75	0.077	0.067	0.070	0.252	0.220	0.202

C. Lampiran 3. Nilai MSE dugaan OLS dan ML pada berbagai proporsi zero

% zero	Dugaan OLS			Dugaan ML		
	MSE(β_0)	MSE(β_1)	MSE(β_2)	MSE(β_0)	MSE(β_1)	MSE(β_2)
5	0.5258	0.0832	0.1238	0.3770	0.1490	0.0866
10	0.5944	0.1147	0.1283	0.3641	0.0572	0.0635
25	0.5547	0.1904	0.2276	0.1782	0.0517	0.0502
50	0.4814	0.1369	0.1539	0.0656	0.0300	0.0306
75	0.3652	0.3002	0.4751	0.3921	0.1860	0.0026

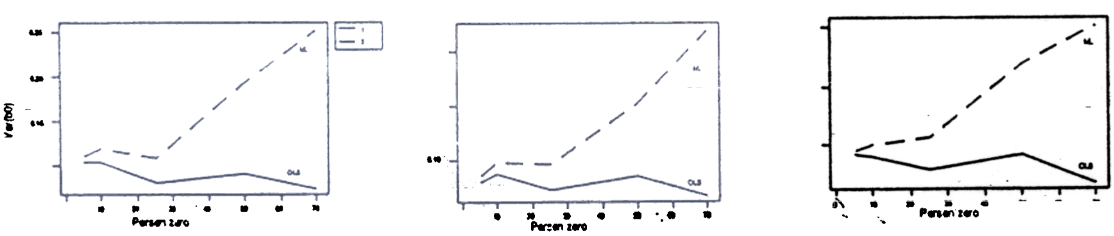
pada Gambar 1. Berdasarkan gambar tersebut terlihat pula bahwa bias koefisien β_0 dugaan OLS lebih besar daripada dugaan ML (Gambar 1a), sedangkan koefisien β_1 dan β_2 dugaan ML lebih besar daripada dugaan OLS. Pada Gambar 1a, 1b dan 1c terlihat bahwa bias dugaan koefisien β_0, β_1 dan β_2 metode ML sama dengan nol (tak bias) pada proporsi nilai amatan nol 40%, sedangkan dugaan OLS selalu berbias pada semua proporsi nilai amatan nol. Nilai dugaan ML cenderung berbias ke atas setelah proporsi nilai amatan nol 40%. Nilai bias tiap koefisien tersebut dapat dilihat pada Lampiran 1.



a. Bias (β_0) b. Bias (β_1) c. Bias (β_2)

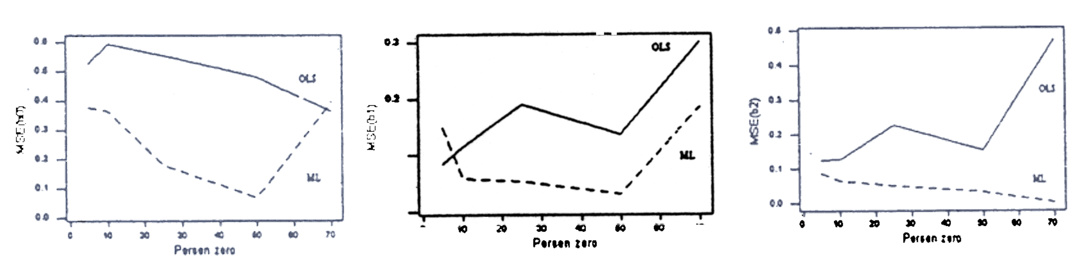
Gambar Rata-rata bias dugaan $\beta_0, \beta_1, \beta_2$ metode OLS dan ML pada berbagai proporsi nilai amatan nol

Nilai ragam dugaan dan MSE dari metode OLS dan ML dapat dilihat pada Lampiran 2 dan 3; sedangkan gambaran grafisnya dapat dilihat pada Gambar 2 dan 3 di bawah ini.



a. Var(β_0) b. Var(β_1) c. Var(β_2)

Gambar 2. Ragam penduga $\beta_0, \beta_1, \beta_2$ metode OLS dan ML pada berbagai proporsi nilai amatan nol



a. MSE (β_0) b. MSE (β_1) c. MSE (β_2)

Gambar 3. MSE dugaan $\beta_0, \beta_1, \beta_2$ metode OLS dan ML pada berbagai proporsi nilai amatan nol

Berdasarkan Gambar 2 (a), (b) dan (c) terlihat bahwa pada proporsi nilai amatan nol sampai dengan sekitar 10%, ragam dugaan β_0, β_1, dan β_2 metode ML dan OLS bisa dikatakan sama. Namun dengan semakin banyaknya proporsi nilai amatan nol, ragam dugaan ML semakin besar dan lebih besar daripada dugaan OLS. Hal ini menunjukkan bahwa penduga OLS lebih efisien daripada penduga ML atau dapat dikatakan juga bahwa tingkat ketelitian penduga OLS lebih baik daripada penduga ML.

Sedangkan berdasarka Gambar 3 (a), (b), dan (c) terlihat bahwa nilai MSE dugaan ML jauh lebih kecil daripada OLS bahkan mendekati nol pada β_2. Hal ini menunjukkan bahwa penduga ML merupakan penduga yang lebih konsisten daripada OLS.

4. Simpulan

Berdasarkan hasil analisis yang telah dilakukan maka dapat disimpulkan bahwa sampai dengan 10% proporsi nilai amatan nol, nilai dugaan OLS dan ML relatif sama. Nilai dugaan ML tak bias pada proporsi nilai amatan nol 40%, dan sesudah itu penduga ML cenderung berbias ke atas; namun secara umum dapat dikatakan bahwa penduga ML jauh lebih tepat daripada penduga OLS, dengan kata lain lebih bersifat takbias. Kuadrat tengah galat yang lebih kecil dan lebih mendekati nol menunjukkan bahwa penduga ML lebih konsisten secara asimtotis. Sedangkan penduga OLS mempunyai ragam penduga yang jauh lebih kecil daripada penduga ML, yang menunjukkan bahwa penduga OLS mempunyai tingkat ketelitian yang lebih tinggi daripada penduga ML.

Daftar Pustaka

[1] Casella, G. and R. L. Berger. 1990. *Statistical inference.* California: Wadsworth Inc.

[2] Heien, D. and C. R. Wessells. 1990. Demand system estimation with microdata: A censored regression approach. *Journal of Business & Economic Statistics* Vol. 8 (3), 365-371.

[3] Kennedy, P. *A Guide to econometrics.* 1996. Cambridge: The MIT Press.

[4] Maddala, G. S. 1983. *Limited dependent and qualitative variables in econometrics.* New York, University Press.

[5] Kshirsagar, A. M. 1983. *A course in linear models.* New York : Marcel Dekker, Inc.

[6] Tobin, J. 1958. Estimation of relationships for limited dependent variables. *Econometrica* Vol. 26, 24-36.

[7] Virgantari, F. dan Siswadi. 2005. Model regresi tersensor kiri pada nilai amatan nol. *Prosiding Seminar Nasional Matematika 30 Juli 2005 (dalam proses penerbitan).* Universitas Indonesia. Jakarta.

[8] Walpole, R. E. 1982. *Pengantar Statistika.* Jakarta : Gramedia Pustaka Utama.

A. Lampiran 1. Nilai rata-rata bias dugaan OLS dan ML pada berbagai proporsi zero

	Dugaan OLS			Dugaan ML		
% zero	Bias(β_0)	Bias(β_1)	Bias(β_2)	Bias(β_0)	Bias(β_1)	Bias(β_2)
5	0.7251	-0.2884	-0.3518	0.6140	-0.3860	-0.2943
10	0.7710	-0.3387	-0.3582	0.6034	-0.2391	-0.2519
25	0.7448	-0.4363	-0.4771	0.4221	-0.2274	-0.2241
50	0.6938	-0.3700	-0.3923	-0.2562	0.1732	0.1750
75	0.6043	-0.5479	-0.6893	-0.6262	0.4313	0.0505