BAB II

MODEL AUTOKORELASI

Dalam statistika khususnya yang berhubungan dengan pendugaan parameter model regresi linier dikenal metode kuadrat terkecil biasa (Ordinary Least Square Estimators) sebagai pemerkira linear terbaik tak bias (BLUE : Best Linear Unbiased Estimator). Salah satu asumsi untuk menggunakan menggunakan metode ini adalah tidak ada korelasi diantara kesalahan pengganggu (ε).

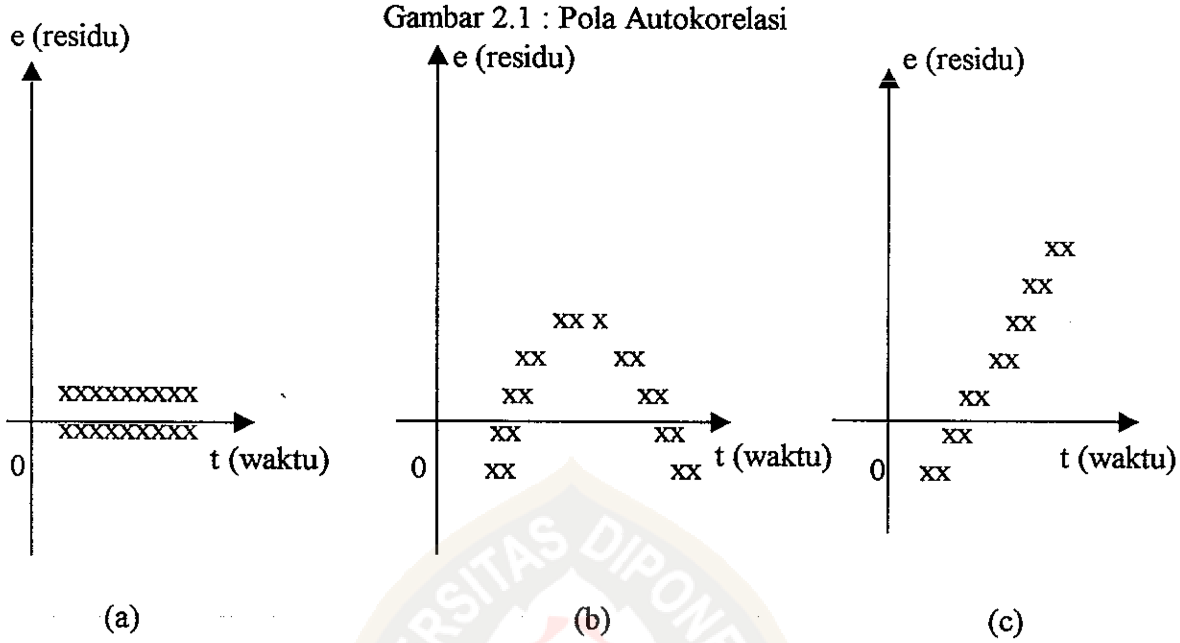
$$E(\varepsilon_i \ \varepsilon_j) = 0, \ i \neq j$$

Jika $E(\varepsilon_i . \varepsilon_j) = 0, \ i \neq j$, maka model tersebut bisa disebut sebagai "Model Autokorelasi". Pemaksaan pemakaian model regresi dengan mengabaikan adanya korelasi diantara kesalahan pengganggu (residu) padahal berkorelasi akan berakibat penaksiran yang kurang sesuai dari data yang bersangkutan. Sehingga harus segera diatasi.

Autokorelasi merupakan korelasi antara anggota seri observasi yang disusun menurut urutan waktu (seperti data time series) atau menurut urutan tempat (ruang seperti data cross-section). Biasanya jenis data ini dijumpai di bidang ekonomi, kedokteran pertanian dan di bidang lain yang melibatkan adanya perlakuan (treatments).

Korelasi diantara residu ditunjukkan dengan adanya pola diantara residu bila diplot dalam grafik, beberapa pola diantaranya yaitu :

Gambar 2.1 : Pola Autokorelasi



Dari grafik hubungan waktu (t) dengan residu (e) di atas nampak bahwa gambar 2.1a tidak berpola sistematis,sehingga tidak berkorelasi, sedang dari gambar 2.1b dan 2.1c menunjukkan adanya korelasi serial diantara residu, karena gambar 2.1b mempunyai trend kuadratik dan gambar 2.1c mempunyai trend linear menaik.

2.1. Pengaruh Bias Spesifikasi Model terhadap Model Regresi

Terdapat banyak sebab terjadinya model autokorelasi. Salah satu diantaranya yaitu tidak dimasukkannya variabel penting dalam model regresi. Dalam analisa data seringkali terjadi bahwa seorang peneliti memulai dengan model regresi tertentu yang dapat diterima walau pun belum sempurna betul. Misalnya setelah dilakukan analisa residu, ternyata terdapat hubungan sistematis diantara komponen residu. Berarti residu tersebut merupakan fungsi dari suatu variabel tertentu. Kondisi model seperti ini dikatakan dalam model terjadi bias spesifikasi.

Untuk mengatasinya kita harus mengusahakan suatu model regresi baru, sehingga didapat model regresi dengan residu tidak berkorelasi, $E(\varepsilon_I, \varepsilon_j) = 0$ sesuai yang diinginkan.

Contoh :

Diberikan model regresi hubungan antara biaya marginal (Y) dengan output (X), sebagai berikut :

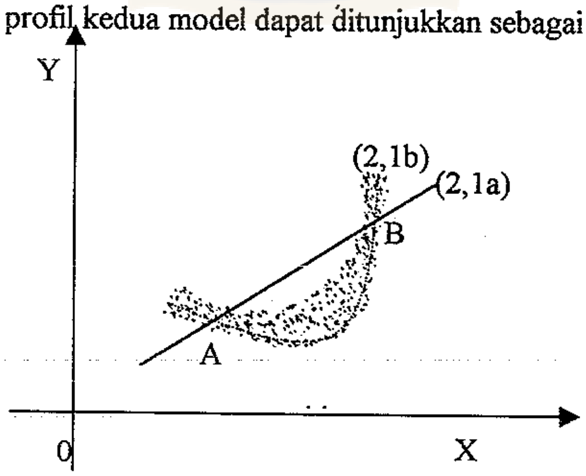
$$Y = \beta_0 + \beta_1 X + U_1 \qquad (2.1a)$$

Tetapi setelah dilakukan analisa, ternyata residu (U_1) saling berkorelasi dan menunjukkan tren kuadrat sehingga model sebenarnya adalah :

$$Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \varepsilon \qquad (2.1b)$$

Dengan ε berdistribusi normal dengan rata-rata (mean) 0 (nol) dan variansi σ^2, dinyatakan dengan $\varepsilon \sim N(0, \sigma^2)$.

Dalam gambar, profil kedua model dapat ditunjukkan sebagai berikut :



Nampak bahwa data biaya marginal (Y) antara titik A dengan B jika digunakan model (2.1a) akan "under estimate" dari biaya sebenarnya, kurva yang seharusnya tidak linier dan karena dilinierkan maka akan timbul residu $U_I = \beta_2 x_i^2 + \varepsilon$ yang saling berkorelasi.

Selanjutnya kasus ini akan dikembangkan untuk membangkitkan model regresi dari data longitudinal dalam bab III.

Berikut ini akan ditunjukkan hubungan antara residu berkorelasi terhadap pemerkira OLS. Sebelumnya kita ambil model regresi linier :

$$Y_t = \beta_0 + \beta_1 X_t + \varepsilon_t \qquad (2.2)$$

t menunjukkan observasi pada waktu t.

Karena ε_t tidak diketahui nilai sebenarnya, untuk langkah permulaan dalam mendapatkan ε_t kita dapat mengassumsikan bahwa :

$$\varepsilon_t = \rho\varepsilon_{t-1} + U_t \; (-1 \leq \rho \leq +1) \qquad (2.3)$$

dengan ρ adalah koefisien otokovarian (ρ = rho) dan U_t adalah kesalahan pengganggu. Sehingga jika assumsi OLS dipenuhi, maka akan berlaku.

$$E(U_t) = 0$$

$$Var(U_t) = \sigma^2$$

$$Kov(U_t, U_{t+s}) = 0, \; s \neq 0 \qquad (2.4)$$

Model (2.3) disebut autoregressif order pertama, karena model tersebut menunjukkan regresi antar ε_t dengan dirinya sendiri dengan beda kala satu (ε_{t-1}).

Dari definisi korelasi antara dua variabel, nilai ρ dapat dinyatakan dengan :

$$\rho = \frac{E\{\varepsilon_t - E(\varepsilon_t)\}\{\varepsilon_{t-1} - E(\varepsilon_{t-1})\}}{\sqrt{Var(\varepsilon_t)}\sqrt{Var(\varepsilon_{t-1})}}$$

$$\rho = \frac{E(\varepsilon_t \varepsilon_{t-1})}{Var(\varepsilon_{t-1})}, \text{ dengan } E(\varepsilon_t) = 0,\ Var(\varepsilon_t) = Var(\varepsilon_{t-1}) \qquad (2.5)$$

Dari (2.5) dapat dikatakan pula bahwa ρ merupakan koefisien arah (slope) dan nampak pula bahwa gerakan ε_t terdiri dari dua bagian, yaitu bagian pertama oleh gerakan sistematis ε_{t-1} dan gerakan acak U_t.

Dengan koefisien autokorelasi order pertama dapat ditunjukkan bahwa :

$$Var(b^*) = \frac{\sigma^2}{\sum_{i=1}^{n} x_t^2}\left(1 + 2\rho \frac{\sum_{t=1}^{n-1} x_t x_{t+1}}{\sum_{t=1}^{n} x_t^2} + 2\rho^2 \frac{\sum_{t=1}^{n-1} x_t x_{t+2}}{\sum_{t=1}^{n} x_t^2} + \ldots\ldots \right.$$

$$\left. + 2\rho^{n-1} \frac{\sum_{t=1}^{n-1} x_1 x_n}{\sum_{t=1}^{n} x_t^2}\right) \qquad (2.6)$$

Adanya korelasi diantara residu menyebabkan variansi dari (2.6) bukanlah variansi yang minimum, sehingga diperlukan cara untuk membuat variansi tersebut minimum. Ketidakminimuman variansi parameter dari model autokorelasi dapat diperlihatkan sebagai berikut :

Dari model : $y_t = \beta_0 + \beta_1 x_t + \varepsilon_t$ dan dengan metode OLS didapat :

$$\hat{\beta}_1 = \sum c_t y_t = \sum c_t(\beta_0 + \beta_1 x_t + \varepsilon_t)$$

$$= \sum c_t \beta_0 + \beta_1 \sum c_t x_t + \sum c_t \varepsilon_t \text{ ....(1)}$$

dengan $c_t = \dfrac{(X_t - \overline{X})}{\sum(X_t - \overline{X})^2}$

Karena $\sum c_t = 0$, $\sum c_t x_t = 1$, maka :

$$\hat{\beta}_1 = b = \beta_1 + \sum c_t \varepsilon_t \rightarrow \hat{\beta}_1 - \beta_1 = \sum c_t \varepsilon_t \text{ ....(2)}$$

dari (1) di dapat $\mathrm{Var}(\hat{\beta}_1) = E(\hat{\beta}_1 - \beta_1)^2 = E(\sum c_t \varepsilon_t)^2$

$$\mathrm{Var}(\hat{\beta}_1) = E(\sum c_t^2 \varepsilon_t^2 + 2\sum\sum c_t c_s \varepsilon_t \varepsilon_s)$$

$$\mathrm{Var}(\hat{\beta}_1) = \sum c_t^2 E(\varepsilon_t^2) + 2\sum\sum c_t c_s E(\varepsilon_t \varepsilon_s),\ t \neq s$$

$$\mathrm{Var}(\hat{\beta}_1) = \sum c_t^2 E(\varepsilon_t^2) + 2[\sum c_t c_{t+1} E(\varepsilon_t \varepsilon_{t+1}) + \text{ .....}$$

$$c_1 c_n E(c_1 \varepsilon_n) \text{ ....(3)}$$

dan dari (2.5) didapat hubungan $E(\varepsilon_t \varepsilon_{t+1}) = \rho \mathrm{Var}\ \varepsilon_t$

dengan cara analog akan didapatkan :

- $E(\varepsilon_t \varepsilon_{t+2}) = \rho^2 \mathrm{Var}\ \varepsilon_t$

$\vdots \qquad \ddots \qquad \vdots$

- $E(\varepsilon_1 \varepsilon_n) = \rho^{n-1} \mathrm{Var} \varepsilon_t$ ....(4)

Jadi dengan substitusi (4) ke (2) didapatkan :

$$\text{Var}(\hat{\beta}_1) = \Sigma c_t^2 \sigma^2 + 2\,\Sigma\,[c_t c_{t+1} \rho \sigma^2 + c_t c_{t+2} \rho \sigma^2 + c_1 c_n \rho^{n-1} \sigma^2]$$

$$\text{Var}(\hat{\beta}_1) = \Sigma c_t^2 \sigma^2 + 2\sigma^2\,\Sigma\,[c_t c_{t+1} \rho + c_t c_{t+2} \rho^2 + \ldots + c_1 c_n \rho^{n-1}]$$

$$\text{Var}(\hat{\beta}_1) = \frac{\sigma^2}{\Sigma x_t^2}\,[1 + 2\rho \sum_{t=1}^{n-1} x_t x_{t+1} + 2\rho^2 \sum_{t=1}^{n-2} x_t x_{t+2} + 2\rho^{n-1} x_1 x_n]$$

$$\text{Var}(b^*) = \frac{\sigma^2}{\Sigma x_t^2}\,[1 + 2\rho \sum_{t=1}^{n-1} x_t x_{t+1} + 2\rho^2 \sum_{t=1}^{n-2} x_t x_{t+2} + 2\rho^{n-1} x_1 x_n]$$

Persamaan (2.6) merupakan varian pemerkira OLS yang biasa dengan memperhatikan adanya korelasi serial order-pertama.

Dari (2.6) dan Var (b) = $\sigma^2\ \Sigma x_t^2$ jelas bahwa Var (b) < Var (b^*)

Jelas bahwa Var (b^*) tidak minimum lagi.

Demikian pula dengan Var (b) = Var (b_0) suatu variansi dari estimator β_1 dengan metode OLS didapat :

$$b_0 = \frac{\Sigma(x_t - \rho x_{t-1})(y_t - \rho y_{t-1})}{\Sigma(x_t - \rho x_{t-1})^2} + C \qquad (2.7a)$$

$$\text{Var}(b_0) = \frac{\sigma^2}{\Sigma(x_t - \rho x_{t-1})^2} + D \qquad (2.7b)$$

Dengan C, D = faktor koreksi

Dari (2.6), (2.7a) dan (2.7b) dengan nilai $\rho = 0$, maka rumus bo = b dan Var (b^*) = Var (bo) = Var (b).

Kalau kita gunakan Var (b^*), kita akan menaikan ketelitian pemerkira b, sebab standar errornya "underestimate".

2.2. Model regresi Berganda (Multiple)

Model regresi yang melibatkan lebih dari satu variabel regressor (variabel bebas) disebut model regresi multiple.

Misalkan X_1, X_2 adalah variabel regressor dan Y adalah variabel response, maka :

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon \tag{2.8}$$

Disebut model regresi linier multiple dengan dua variabel regressor dengan parameter tak diketahui (unknown parameter) β_0 , β_1 dan β_2. Parameter β_0 disebut intersep.

Jika jangkauan data diambil $X_1 = X_2 = 0$, maka β_0 disebut pula sebagai rata-rata (mean) Y ketika $X_1 = X_2 = 0$. Parameter β_1 mengindikasikan perubahan rata-rata pada response (Y) perunit perubahan dalam X_1 ketika X_2 adalah konstan. Demikian pula β_2 menunjukkan perubahan rata-rata dalam X_2 ketika X_1 adalah konstan.

Secara umum, katakanlah response Y bergantung dari k variabel regressor, model (2.8) dapat ditulis :

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k + \varepsilon \tag{2,9}$$

Disebut model regresi linier multiple k regressor dengan parameter β_j, j = 0, 1, 2 … k disebut koefisien regresi. Parameter β_j menunjukkan perubahan rata-rata response Y perunit. Perubahan dalam Xj ketika semua variabel regressor regressor X_i ($i \neq j$) adalah konstan. Selanjutnya parameter β_j, j = 1,2, … k disebut koefisien regresi parsial.

Model (2.9) dapat digunakan untuk menganalisa hubungan variabel regressor X_k dengan variabel respons Y yang lebih komplek.

Sebagai contoh model regresi kubik :

$$Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \beta_3 X^3 + \varepsilon \qquad (2.10a)$$

Jika diambil $X_1 = X$, $X_2 = X^2$ dan $X_3 = X^3$, maka model (2.10a) dapat dinyatakan sebagai :

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \varepsilon \qquad (2.10b)$$

Yang merupakan model regresi linier multiple dengan tiga variabel regressor. Sehingga dalam penganalisaannya dapat digunakan prinsip dari model regresi berganda (multiple).

Model (2.9) dapat pula diterapkan dari model yang mengandung efek interaksi.

Sebagai contoh :

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_{12} X_1 X_2 + \varepsilon \qquad (2.11a)$$

Jika diambil $X_3 = X_1 X_2$ dan $\beta_{12} = \beta_3$, maka model (2,11a) dapat ditulis sebagai :

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X^3 + \varepsilon \qquad (2.11b)$$

yang merupakan model regresi linier multiple dengan tiga variabel regressor. Model dengan efek interaksi ini dijumpai pada percobaan faktorial.

2.2.1 Estimasi Parameter Model

Metode OLS dapat digunakan untuk mengestimasi koefisien regresi (2, 9).

Ambil Y_1 sebagai response ke-i dan X_{ij} menunjukkan observasi ke-i pada level regresi X_j. Asumsi yang dipakai yaitu komponen error (ε) dalam model mempunyai E (ε) = 0, V (ε) = σ^2 dan residu tidak berkorelasi, Cov (ε_i , ε_j) = 0, $i \neq j$.

Model (2.9) dapat dinyatakan kembali sebagai

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \ldots + \beta_k X_{ik} + \varepsilon_I$$

$$= \beta_0 + \sum_{j=1}^{k} \beta_j X_{ij} + \varepsilon_i, \quad i = 1,2 \ldots n \qquad (2.12)$$

Dengan fungsi kuadrat terkecil (least square)nya yaitu :

$$S(\beta_0, \beta_1, \ldots \beta_k) = \sum_{i=1}^{n} \varepsilon_i^2$$

$$= \sum_{i=1}^{n} (Y_i - \beta_0 - \sum_{i=1}^{k} \beta_j X_{ij})^2 \qquad (2.13)$$

Parameter β_0, $\beta_1 \ldots \beta_k$ dapat diestimasi dari (2.13) dengan mengambil turunan pertama S terhadap β_j dan menyamakannya dengan nol sehingga diperoleh :

$\frac{\partial S}{\partial \beta_j} = 0$, sehingga :

$$\left.\frac{\partial S}{\partial \beta_0}\right|_{\hat{\beta}_0, \hat{\beta}_1 \ldots \hat{\beta}_k} = -2\sum_{i=1}^{n}(Y_i - \hat{\beta}_0 - \sum_{i=1}^{k}\hat{\beta}_j X_{ij}) = 0 \qquad (2.14a)$$

dan

$$\left.\frac{\partial S}{\partial \beta_j}\right|_{\hat{\beta}_0, \hat{\beta}_1 \ldots \hat{\beta}_k} = -2\sum_{i=1}^{n}(Y_i - \hat{\beta}_0 - \sum_{i=1}^{k}\hat{\beta}_j X_{ij})X_{ij} = 0$$

$$j = 1,2, \ldots k \qquad (2.14b)$$

Secara umum (2.14a) dan (2.14b) dapat dituliskan dalam persamaan normal kuadrat terkecil, sebagai berikut :

$$n\hat{\beta}_0 + \hat{\beta}_1\sum_{i=1}^{n}X_{i1} + \hat{\beta}_2\sum_{i=1}^{n}X_{i2} + \ldots + \hat{\beta}_k\sum_{i=1}^{n}X_{ik} = \sum_{i=1}^{n}Y_i$$

$$\hat{\beta}_0\sum_{i=1}^{n}X_{i1} + \hat{\beta}_1\sum_{i=1}^{n}X_{i1}^2 + \hat{\beta}_2\sum_{i=1}^{n}X_{i1}X_{i2} + \ldots + \hat{\beta}_k\sum_{i=1}^{n}X_{i1}X_{ik} = \sum_{i=1}^{n}X_{i1}Y_i$$

$$\vdots \qquad \vdots \qquad \vdots \qquad \vdots \qquad \vdots$$

$$\hat{\beta}_0\sum_{i=1}^{n}X_{ik} + \hat{\beta}_1\sum_{i=1}^{n}X_{ik}X_{i1} + \hat{\beta}_2\sum_{i=1}^{n}X_{ik}X_{i2} + \ldots + \hat{\beta}_k\sum_{i=1}^{n}X_{ik}^2 = \sum_{i=1}^{n}X_{ik}Y_i$$

.......(2.21)

Jadi terdapat p = k + 1 persamaan normal, solusi persamaan normal merupakan estimator OLS $\hat{\beta}_0, \hat{\beta}_1 \ldots \hat{\beta}_k$.

Dalam praktek, (2,15) dituliskan dalam bentuk matrik, yaitu :

$$\underset{\sim}{Y} = X\underset{\sim}{\beta} + \underset{\sim}{\varepsilon} \qquad (2.16)$$

Dengan : $\underset{\sim}{Y} = \begin{bmatrix} Y_1 \\ Y_1 \\ \vdots \\ Y_n \end{bmatrix}$ $\qquad X = \begin{bmatrix} 1 & X_{11} & X_{12} & \ldots X_{1k} \\ 1 & X_{21} & X_{22} & \ldots X_{2k} \\ \vdots & & & \\ 1 & X_{n1} & X_{n2} & \ldots X_{nk} \end{bmatrix}$

$\underset{\sim}{\beta} = \begin{bmatrix} \beta_0 \\ \beta_0 \\ \vdots \\ \beta k \end{bmatrix}$ $\qquad \underset{\sim}{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$

dalam hal ini Y = vektor observasi berukuran n x 1

X = matrik variabel regresor berukuran n x p

$\underset{\sim}{\beta}$ = vektor koefisien regresi berukuran p x 1

$\underset{\sim}{\varepsilon}$ = vektor error random berukuran n x 1

Kita dapat menemukan vektor estimasi parameter, $\hat{\beta}$ dengan meminimumkan :

$$S(\beta) = \sum_{i=1}^{n} \varepsilon_i^2 = \varepsilon'\varepsilon = (Y - X\beta)'(Y - X\beta)$$

atau

$$S(\beta) = Y'Y - \beta' X' Y - Y' X\beta + \beta' X' X\beta$$

$$= Y'Y - 2\beta' X' Y + \beta' X' X\beta \qquad (2.17)$$

karena β'X'Y adalah sebuah matrik berukuran 1 x 1, atau merupakan suatu skalar, maka transpose (β'X'Y)' = Y'Xβ adalah skalar yang sama.

Estimator β dengan OLS dapat dinyatakan dengan menurunkan S terhadap terhadap β dan menyamakan dengan 0,

$$\frac{\delta S}{\delta \beta_j}\bigg|\hat{\beta} = -2X'Y + 2X'X\,\beta = 0$$

atau

$$X'X\hat{\beta} = X'Y \qquad (2.18)$$

Persamaan (2.18) disebut persamaan normal yang analog dengan (2.17)

Nilai parameter β dalam (2.18) dapat diperoleh dengan mengalikan kedua ruas dengan invers dari X'X. Sedemikian sehingga estimator β dengan OLS adalah :

$$\hat{\beta} = (X'X)^{-1}\; X'Y \qquad (2.19a)$$

Dari sini nampak bahwa untuk mengestimasi parameter dengan metode OLS selalu harus dihitung terlebih dahulu matrik $(X'X)^{-1}$ yang tentunya menyulitkan untuk ukuran yang besar, sehingga dalam bab III akan dikembangkan "Metode Doolittle dipersingkat" yang dalam pengestimasian parameternya tidak mutlak harus dihitung terlebih dahulu matrik $(X'X)^{-1}$.

Kalau diperhatikan persamaan normal (2.18) adalah identik dengan bentuk skalar (2.15) sehingga dapat ditulis :

$$\begin{bmatrix} n & \sum_{l=1}^{n} X_{i1} \ldots \sum_{l=1}^{n} X_{ik} \\ \sum_{l=1}^{n} X_{i1} & \sum_{l=1}^{n} X_{i1} \ldots \sum_{l=1}^{n} X_{i1} X_{ik} \\ \vdots & \vdots \\ \sum_{l=1}^{n} X_{ik} & \sum_{l=1}^{n} X_{ik} X_{i1} \ldots \sum_{l=1}^{n} X_{ik}{}^2 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta k \end{bmatrix} = \begin{bmatrix} \sum Y_1 \\ \sum X_i Y_i \\ \vdots \\ \sum X_{ik} Y_i \end{bmatrix} \qquad (2.19b)$$

Terlihat bahwa X'X adalah matrik simetri berukuran (pXp) dan X'Y adalah vektor kolom berukuran (pX1). Struktur khusus dari X'X adalah elemen diagonal dari X'X adalah jumlah kuadrat dari elemen kolom X, selain diagonal adalah jumlah hasil kali silang dari elemen dalam kolom X. Disisi lain elemen X'Y adalah jumlahan hasil kali silang dari kolom-kolom X dengan observasi Y_i.

Taksiran model regresi yang berkaitan dengan level-level variabel regresor $X' = (1, X_1, X_2, \ldots X_k)$ adalah

$$\hat{Y} = X'\hat{\beta}$$

$$= \hat{\beta}_0 + \sum_{j=1}^{n} \hat{\beta}_j X_j$$

Hubungan antara Y_i dengan Y_i adalah

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

$$\hat{Y} = H\,Y \qquad (2.20)$$

Matrik $H = X(X'X)^{-1} X'$ yang berukuran n x n ini disebut matrik "Hessian" yaitu merupakan pemetaan dari nilai vektor observasi ke dalam vektor taksiran.

Sedangkan perbedaan nilai observasi Y_i dengan taksiran Y_i disebut eror $\varepsilon_i = Y_i - \hat{Y}_i$ yang dalam notasi matrik bila ditulis sebagai :

$$\varepsilon = Y - \hat{Y} \tag{2.21a}$$

atau

$$\varepsilon = Y - X\beta$$

$$= Y - HY \tag{2.21b}$$

$$= (I - H)Y \tag{2.21c}$$

2.2.2. Aturan dari Estimator OLS

Aturan dari estimator β dapat ditunjukkan

a) Unbias

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

$$= E\{(X'X)^{-1} X' (X\beta + \varepsilon)\}$$

$$= E\{(X'X)^{-1} X'X\beta + (X'X)^{-1} X'\varepsilon\}$$

$$= \beta + 0 \quad \text{karena } E(\varepsilon) = 0 \text{ dan } (X'X)^{-1} X'X = I$$

$$E(\hat{\beta}) = \beta$$

Jadi estimator $\hat{\beta}$ unbias terhadap β

b) Aturan Variansi

Variansi estimator $\hat{\beta}$ diekspresikan dengan matrik kovariansi :

$$\text{Cov}(\hat{\beta}) = E\{(\hat{\beta} - E(\hat{\beta}))\}\{\hat{\beta} - E(\hat{\beta})'\} \quad (2.22)$$

Yang merupakan matrik simetri berukuran p x p dengan elemen diagonal ke-j adalah variansi dari $\hat{\beta}_j$ dan elemen bukan diagonal ke (ij) adalah kovariansi antara $\hat{\beta}_I$ dan $\hat{\beta}_j$.

Matrik Kov $\hat{\beta}$ adalah Cov $(\hat{\beta}) = \hat{\sigma}^2 (X'X)^{-1}$, Jika kita nyatakan $(X'X)^{-1}$ dengan C, maka variansi dari $\hat{\beta}_j$ adalah $\hat{\sigma}^2 C_{jj}$ dan kovariansi antara $\hat{\beta}_i$, dengan $\hat{\beta}_j$ adalah $\hat{\sigma}^2 C_{ij}$

2.2.3. Estimasi dari σ^2

Sebagaimana dalam regresi linier sederhana, estimator dari σ^2 diperoleh dari jumlah kuadrat residual

$$JKG = \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2$$

$$= \sum_{i=1}^{n} e_i^2 = e_i' e_i \quad (2.23a)$$

dengan substitusi $e = Y - X\hat{\beta}$ didapat :

$$JKG = (Y - X\hat{\beta})'(Y - X\hat{\beta})$$

$$= Y'Y - \hat{\beta}'X'Y - Y'X\hat{\beta} + \hat{\beta}'X'X\hat{\beta}$$

$$= Y'Y - 2\hat{\beta}'X'Y + \hat{\beta}'X'X\hat{\beta}$$

Karena $X'X\hat{\beta} = X'Y$, persamaan menjadi :

$$JKG = Y'Y - \hat{\beta}' X\beta \qquad (2.23b)$$

d.b = n – p

p = banyaknya parameter

Rata-rata kuadrat dari (2.21) adalah :

$$KTG = \frac{JKG}{n-p} \qquad (2.23c)$$

Kita dapat menunjukkan bahwa nilai dari KTG adalah σ^2, sehingga estimator unbias dari σ^2 adalah :

$$\hat{\sigma}^2 = KTG \qquad (2.23d)$$

2.2.3. Interval Kepercayaan (Konfidensi) dari Parameter Model Regresi Berganda

Penentuan Interval konfidensi dalam model regresi sangat penting. Semakin lebar interval semakin jelek estimasi yang dibuat. Dalam bagian ini akan diuraikan interval konfidensi untuk koefisien regresi (β_j).

Untuk menyusun interval konfidensi dari estimator β_j kita harus mengasumsikan bahwa residu (ε_i) berdistribusi normal dan saling independen dengan mean 0 dan variansi σ^2.

Demikian pula observasi Y_i berdistribusi normal dengan mean $\beta_o + \sum_{j=1}^{k} \beta_j X_{ij}$ dan variansi σ^2.

Karena OLS $\hat{\beta}$ adalah kombinasi linier dari observasi, maka hal ini menunjukkan bahwa $\hat{\beta}$ berdistribusi normal dengan mean vektor β dan matrik kovariansi $\hat{\sigma}^2$ $(X'X)^{-1}$. Hal ini mengindikasikan bahwa distribusi marginal dari koefisien regresi β_j adalah normal dengan mean β_j dan variansi $\hat{\sigma}^2 C_{jj}$. dengan C_{jj} adalah elemen diagonal ke-j dari matrik $(X'X)^{-1}$ sesuai dengan β_j yang sedang diestimasi.

Akibatnya setiap koefisien regresi dapat diuji dengan statistik:

$$\frac{\beta_j - \hat{\beta}_j}{\sqrt{\hat{\sigma}^2 C_{jj}}} \quad j = 0,1, \ldots k \tag{2.24}$$

Yang berdistribusi t dengan d.b = n-p dengan $\hat{\sigma}^2$ merupakan estimasi variansi galat.

100 $(1-\alpha)$% Interval konfidensi dari koefisien regresi β_j, $j = 0,1, \ldots k$ adalah :

$$\hat{\beta}_j - t_{\alpha/2,(n-p)}\sqrt{\hat{\sigma}^2 C_{jj}} < \beta_j < \hat{\beta}_j + t_{\alpha/2,\, n-p}\sqrt{\hat{\sigma}^2 C_{jj}} \tag{2.23}$$

Dengan Se $(\hat{\beta}_j)$ dan $\sqrt{\hat{\sigma}^2 C_{jj}}$ yaitu standar error dari koefisien regresi $\hat{\beta}_j$