

Classical linear model ^{*}

Jean-Marie Dufour [†]
McGill University

January 2024

^{*}This work was supported by the William Dow Chair in Political Economy (McGill University), the Bank of Canada (Research Fellowship), the Toulouse School of Economics (Pierre-de-Fermat Chair of excellence), the Universidad Carlos III de Madrid (Banco Santander de Madrid Chair of excellence), a Guggenheim Fellowship, a Konrad-Adenauer Fellowship (Alexander-von-Humboldt Foundation, Germany), the Canadian Network of Centres of Excellence [program on *Mathematics of Information Technology and Complex Systems* (MITACS)], the Natural Sciences and Engineering Research Council of Canada, the Social Sciences and Humanities Research Council of Canada, and the Fonds de recherche sur la société et la culture (Québec).

[†]William Dow Professor of Economics, McGill University, Centre interuniversitaire de recherche en analyse des organisations (CIRANO), and Centre interuniversitaire de recherche en économie quantitative (CIREQ). Mailing address: Department of Economics, McGill University, Leacock Building, Room 414, 855 Sherbrooke Street West, Montréal, Québec H3A 2T7, Canada. TEL: (1) 514 398 6071; FAX: (1) 514 398 4800; e-mail: jean-marie.dufour@mcgill.ca. Web page: <http://www.jeanmariedufour.com>

Contents

1. Model-free linear regression and ordinary least squares	1
1.1. Notations	1
1.2. The least squares problem	3
2. Classical linear model	6
3. Linear unbiased estimation	8
4. Prediction	18
5. Estimation with Gaussian errors	21
6. Confidence and prediction intervals	26
6.1. Confidence interval for the error variance	26
6.2. Confidence interval for a linear combination of regression coefficients	27
6.3. Confidence region for a regression coefficient vector	28
6.4. Prediction intervals	29
6.5. Confidence regions for several predictions	30
7. Hypothesis tests	32
8. Estimator optimal properties with Gaussian errors	38

1. Model-free linear regression and ordinary least squares

1.1. Notations

We wish to explain or predict a variable y through k other x_1, x_2, \dots, x_k . We T observations on each variable:

$$y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_T \end{pmatrix} : \text{dependent variable (to explain)}$$
$$x_i = \begin{pmatrix} x_{1i} \\ x_{2i} \\ \vdots \\ x_{Ti} \end{pmatrix}, \quad i = 1, \dots, k : \text{explanatory variables.}$$

Usually, the explanatory variables are represented by the $T \times k$ matrix

$$X = [x_1, x_2, \dots, x_k] = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1k} \\ x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & & \vdots \\ x_{T1} & x_{T2} & \cdots & x_{Tk} \end{bmatrix} = \begin{bmatrix} X'_1 \\ X'_2 \\ \vdots \\ X'_T \end{bmatrix},$$

where X_t is a $k \times 1$ vector:

$$X'_t = (x_{t1}, x_{t2}, \dots, x_{tk}), \quad t = 1, \dots, T.$$

We wish to represent each observation y_t as a function of x_{t1}, \dots, x_{tk} :

$$y_t = x_{t1}\beta_1 + x_{t2}\beta_2 + \cdots + x_{tk}\beta_k + \varepsilon_t, \quad t = 1, \dots, T \quad (1.1)$$

where ε_t is a “residual” which is left unexplained by the explanatory variables. This model can also be written in the following matrix form:

$$y = X\beta + \varepsilon \quad (1.2)$$

where $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_T)'$.

1.2. The least squares problem

1.2.1 In general, we cannot obtain a “perfect fit” ($\varepsilon_t = 0$, $t = 1, \dots, T$). In view of this, a natural approach (proposed by Gauss) consists in minimizing the sum of squared residuals:

$$\begin{aligned}\sum_{t=1}^T \varepsilon_t^2 &= \sum_{t=1}^T [y_t - x_{t1}\beta_1 - \dots - x_{tk}\beta_k]^2 \\ &= (y - X\beta)'(y - X\beta) \equiv S(\beta) .\end{aligned}$$

We consider the problem:

$$\text{Min}_{\beta} (y - X\beta)'(y - X\beta) .$$

Since

$$S(\beta) = (y' - \beta'X')(y - X\beta) = y'y - 2\beta'X'y + \beta'X'X\beta ,$$

we have:

$$\frac{\partial S(\beta)}{\partial \beta} = -2X'y + 2X'X\beta .$$

To compute the above, we use the following result on differentiation with respect to a vector x :

$$\frac{\partial (x'a)}{\partial x} = a , \tag{1.3}$$

$$\frac{\partial (x'Ax)}{\partial x} = (A + A')x . \tag{1.4}$$

For any point $\beta = \hat{\beta}$ such that $S(\beta)$ is a minimum, we must have:

$$\frac{\partial S(\beta)}{\partial \beta} \Big|_{\beta=\hat{\beta}} = -2X'y + 2X'X\hat{\beta} = 0$$

hence

$$(X'X)\hat{\beta} = X'y : \text{normal equations} .$$

1.2.2 When $\text{rank}(X) = k$, we must have $\text{rank}(X'X) = k$ so that $(X'X)^{-1}$ exists. In this case, the normal equations have a unique solution:

$$\hat{\beta} = (X'X)^{-1} X'y. \quad (1.5)$$

Once $\hat{\beta}$ is known, we can compute the “fitted values” and the “residuals” of the model.

1.2.3 The model fitted values are

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

where

$$\begin{aligned} P &= X(X'X)^{-1} X' && \text{(projection matrix)} \\ P' &= P, PP = P && \text{(symmetric idempotent matrix).} \end{aligned}$$

1.2.4 The model residuals are:

$$\hat{\varepsilon} = y - X\hat{\beta} = y - \hat{y} = y - Py = (I - P)y = My$$

where

$$PX = X, MX = 0, \quad (1.6)$$

$$PM = P(I - P) = 0, MP = 0. \quad (1.7)$$

1.2.5 Each column of M is orthogonal with each column of X :

$$\begin{aligned} X'M &= 0, \\ x'_i M &= 0, \quad i = 1, \dots, k. \end{aligned}$$

Residuals and regressors are orthogonal:

$$\begin{aligned} X'\hat{\varepsilon} &= X'My = 0 \\ \Rightarrow x'_i \hat{\varepsilon} &= 0, \quad i = 1, \dots, k \end{aligned}$$

$$\Rightarrow i_T' \hat{\varepsilon} = \sum_{t=1}^T \hat{\varepsilon}_t = 0, \quad \text{if the matrix } X \text{ contains a constant.}$$

where $\hat{\varepsilon} = (\hat{\varepsilon}_1, \hat{\varepsilon}_2, \dots, \hat{\varepsilon}_T)'$ et $i_T = (1, 1, \dots, 1)'$.

1.2.6 Fitted values and residuals are orthogonal:

$$\hat{y}' \hat{\varepsilon} = y' P M y = 0. \quad (1.8)$$

1.2.7 The vector y can be decomposed as the sum of two orthogonal vectors:

$$y = P y + (I - P) y = \hat{y} + \hat{\varepsilon}. \quad (1.9)$$

1.2.8 For any vector β ,

$$\begin{aligned} S(\beta) &\equiv (y - X\beta)' (y - X\beta) = (y - X\hat{\beta})' (y - X\hat{\beta}) + (\hat{\beta} - \beta)' X' X (\hat{\beta} - \beta) \\ &\geq (y - X\hat{\beta})' (y - X\hat{\beta}) = S(\hat{\beta}) \end{aligned}$$

for

$$\begin{aligned} (y - X\beta)' (y - X\beta) &= [y - X\hat{\beta} + X(\hat{\beta} - \beta)]' [y - X\hat{\beta} + X(\hat{\beta} - \beta)] \\ &= [\hat{\varepsilon} + X(\hat{\beta} - \beta)]' [\hat{\varepsilon} + X(\hat{\beta} - \beta)] \\ &= \hat{\varepsilon}' \hat{\varepsilon} + 2(\hat{\beta} - \beta)' X' \hat{\varepsilon} + (\hat{\beta} - \beta)' X' X (\hat{\beta} - \beta) \\ &= \hat{\varepsilon}' \hat{\varepsilon} + (\hat{\beta} - \beta)' X' X (\hat{\beta} - \beta). \end{aligned}$$

This directly verifies that $\beta = \hat{\beta}$ minimizes $S(\beta)$.

2. Classical linear model

In order to establish the statistical properties of $\hat{\beta}$, we need assumptions on X and ε . The following assumptions define the *classical linear model* (CLM).

2.1 Assumption $y = X\beta + \varepsilon$

where y is a $T \times 1$ vector of observations on a dependent variable ,

X is a $T \times k$ matrix of observations on explanatory variables,

β is a $k \times 1$ vector of fixed parameters,

ε is a $T \times 1$ vector of random disturbances.

2.2 Assumption $E(\varepsilon) = 0$.

2.3 Assumption $E[\varepsilon\varepsilon'] = \sigma^2 I_T$.

2.4 Assumption X is fixed (non-stochastic).

2.5 Assumption $\text{rank}(X) = k < T$.

From the assumption 2.1 - 2.4, we see that:

$$\begin{aligned} E(y) &= E(y | X) = X\beta = \begin{pmatrix} X'_1\beta \\ \vdots \\ X'_T\beta \end{pmatrix} \\ &= (x_1, x_2, \dots, x_k) \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{pmatrix} \\ &= x_1\beta_1 + x_2\beta_2 + \dots + x_k\beta_k, \\ V(y) &= V(y | X) = \sigma^2 I_T \\ &= \begin{bmatrix} \sigma^2 & 0 & \dots & 0 \\ 0 & \sigma^2 & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & \sigma^2 \end{bmatrix} = V(\varepsilon). \end{aligned}$$

If, furthermore, we add the assumption that ε follows a multinormal (or Gaussian) distribution, we get the normal classical linear model (NCLM).

2.6 Assumption ε follows a multinormal distribution.

3. Linear unbiased estimation

From the assumptions 2.1 - 2.5, we can make the following observations.

3.1 $\hat{\beta}$ is linear with respect to y .

PROOF $\hat{\beta}$ has the form $\hat{\beta} = Ay$, where $A = (X'X)^{-1}X'$ is a non-stochastic matrix. □

3.2 $\hat{\beta} = (X'X)^{-1}X'(X\beta + \varepsilon) = \beta + (X'X)^{-1}X'\varepsilon$.

3.3 $\hat{\beta}$ is an unbiased estimator of β .

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

3.4 $V(\hat{\beta}) = \sigma^2(X'X)^{-1}$.

PROOF

$$\begin{aligned} V(\hat{\beta}) &= E[(\hat{\beta} - \beta)(\hat{\beta} - \beta)'] \\ &= E[(X'X)^{-1}X'\varepsilon\varepsilon'X(X'X)^{-1}] \\ &= (X'X)^{-1}X'E(\varepsilon\varepsilon')X(X'X)^{-1} \\ &= \sigma^2(X'X)^{-1} \end{aligned}$$

where the last identity follows from Assumption 2.3. □

3.5 Theorem GAUSS-MARKOV THEOREM. $\hat{\beta}$ is the best estimator of β in the class of linear unbiased estimators (BLUE) of β , i.e. $V(\tilde{\beta}) - V(\hat{\beta})$ is a

positive semidefinite matrix for any linear unbiased estimator (LUE) $\tilde{\beta}$ of β . In particular, if $\tilde{\beta} = Cy$ and $D = C - (X'X)^{-1}X'$, then

$$V(\tilde{\beta}) = V(\hat{\beta}) + \sigma^2 DD' .$$

PROOF Since $\tilde{\beta}$ is unbiased and

$$C = D + (X'X)^{-1}X' ,$$

we have:

$$\begin{aligned} E(\tilde{\beta}) &= E \left\{ \left[D + (X'X)^{-1}X' \right] (X\beta + \varepsilon) \right\} \\ &= DX\beta + \beta \\ &= \beta , \end{aligned}$$

hence

$$DX = 0 \quad \text{and} \quad CX = I_k .$$

Consequently,

$$\tilde{\beta} = Cy = CX\beta + C\varepsilon = \beta + C\varepsilon$$

and

$$\tilde{\beta} - \beta = C\varepsilon ,$$

hence

$$\begin{aligned} V(\tilde{\beta}) &= E[(\tilde{\beta} - \beta)(\tilde{\beta} - \beta)'] = E[C\varepsilon\varepsilon'C'] = \sigma^2 CC' \\ &= \sigma^2 [D + (X'X)^{-1}X'] [D' + X(X'X)^{-1}] \\ &= \sigma^2 [DD' + (X'X)^{-1}] = \sigma^2 DD' + \sigma^2 (X'X)^{-1} \\ &= \sigma^2 DD' + V(\hat{\beta}) \end{aligned}$$

and

$$V(\tilde{\beta}) - V(\hat{\beta}) = \sigma^2 DD' \tag{3.1}$$

is a positive semidefinite matrix. □

3.6 Corollary *Let w be a $k \times 1$ vector of constants. Then,*

$$V(w'\tilde{\beta}) \geq V(w'\hat{\beta})$$

for any linear unbiased estimator $\tilde{\beta}$ of β .

PROOF Since $E(\tilde{\beta}) = E(\hat{\beta}) = \beta$, we have:

$$\begin{aligned} E(w'\tilde{\beta}) &= E(w'\hat{\beta}) = w'\beta, \\ V(w'\tilde{\beta}) &= w'V(\tilde{\beta})w = w'[\sigma^2DD' + V(\hat{\beta})]w \\ &= \sigma^2w'DD'w + w'V(\hat{\beta})w \\ &= \sigma^2w'DD'w + V(w'\hat{\beta}) \geq V(w'\hat{\beta}), \end{aligned}$$

for $w'DD'w \geq 0$. □

In particular, we must have:

$$V(\tilde{\beta}_i) \geq V(\hat{\beta}_i), \quad i = 1, \dots, k.$$

3.7 Theorem GENERALIZED GAUSS-MARKOV THEOREM. *Let L be a $r \times k$ fixed matrix and $\gamma = L\beta$. Then $\hat{\gamma} = L\hat{\beta}$ is the BLUE γ , i.e. $V(\tilde{\gamma}) - V(\hat{\gamma})$ is a positive semidefinite matrix for any linear unbiased estimator $\tilde{\gamma}$ of γ . In particular, if $\tilde{\gamma} = Cy$ and $D = C - L(X'X)^{-1}X'$, then*

$$V(\tilde{\gamma}) = V(\hat{\gamma}) + \sigma^2DD'$$

and

$$C(\tilde{\gamma} - \hat{\gamma}, \hat{\gamma}) = 0.$$

PROOF Since $\tilde{\gamma}$ is unbiased and

$$C = D + L(X'X)^{-1}X'$$

we have

$$\begin{aligned}
E(\tilde{\gamma}) &= E\{(D + L(X'X)^{-1}X')(X\beta + \varepsilon)\} \\
&= DX\beta + L\beta = DX\beta + \gamma \\
&= \gamma,
\end{aligned}$$

hence

$$DX = 0 \quad \text{and} \quad CX = L.$$

Consequently,

$$\begin{aligned}
\tilde{\gamma} &= Cy = CX\beta + C\varepsilon \\
&= L\beta + C\varepsilon = \gamma + C\varepsilon
\end{aligned}$$

and

$$\begin{aligned}
V(\tilde{\gamma}) &= E[(\tilde{\gamma} - \gamma)(\tilde{\gamma} - \gamma)'] = E[C\varepsilon\varepsilon'C'] = \sigma^2 CC' \\
&= \sigma^2 [D + L(X'X)^{-1}X'] [D' + X(X'X)^{-1}L'] \\
&= \sigma^2 [DD' + L(X'X)^{-1}L'] \\
&= \sigma^2 DD' + \sigma^2 L(X'X)^{-1}L' = \sigma^2 DD' + V(L\hat{\beta}) \\
&= \sigma^2 DD' + V(\hat{\gamma}),
\end{aligned}$$

so

$$V(\tilde{\gamma}) - V(\hat{\gamma}) = \sigma^2 DD' \tag{3.2}$$

is a positive semidefinite matrix, and

$$\begin{aligned}
C(\tilde{\gamma}, \hat{\gamma}) &= E[C\varepsilon\varepsilon'X(X'X)^{-1}L'] \\
&= \sigma^2 CX(X'X)^{-1}L' = \sigma^2 L(X'X)^{-1}L' = V(\hat{\gamma}),
\end{aligned}$$

$$C(\tilde{\gamma} - \hat{\gamma}, \hat{\gamma}) = C(\tilde{\gamma}, \hat{\gamma}) - C(\hat{\gamma}, \hat{\gamma}) = V(\hat{\gamma}) - V(\hat{\gamma}) = 0. \tag{3.3}$$

□

3.8 Corollary QUADRATIC GAUSS-MARKOV OPTIMALITY. Let Q be a $r \times r$ positive semidefinite fixed matrix and L a $r \times k$ fixed matrix, $\gamma = L\beta$ and $\hat{\gamma} = L\hat{\beta}$. Then

$$E[(\tilde{\gamma} - \gamma)' Q (\tilde{\gamma} - \gamma)] \geq E[(\hat{\gamma} - \gamma)' Q (\hat{\gamma} - \gamma)]$$

for any linear unbiased estimator $\tilde{\gamma}$ of γ .

PROOF Let $\tilde{\gamma} = C\gamma$ and $D = C - L(X'X)^{-1}X'$. Then

$$\begin{aligned} E[(\tilde{\gamma} - \gamma)' Q (\tilde{\gamma} - \gamma)] &= E[\text{tr} Q (\tilde{\gamma} - \gamma) (\tilde{\gamma} - \gamma)'] \\ &= \text{tr} Q E[(\tilde{\gamma} - \gamma) (\tilde{\gamma} - \gamma)'] \\ &= \text{tr} Q [\sigma^2 D D' + V(\hat{\gamma})] \\ &= \sigma^2 \text{tr} (Q D D') + \text{tr} [Q V(\hat{\gamma})] \\ &= \sigma^2 \text{tr} (D' Q D) + \text{tr} Q E[(\hat{\gamma} - \gamma) (\hat{\gamma} - \gamma)'] \\ &= \sigma^2 \text{tr} (D' Q D) + E[\text{tr} (\hat{\gamma} - \gamma)' Q (\hat{\gamma} - \gamma)] \\ &= \sigma^2 \text{tr} (D' Q D) + E[(\hat{\gamma} - \gamma)' Q (\hat{\gamma} - \gamma)] \\ &\geq E[(\hat{\gamma} - \gamma)' Q (\hat{\gamma} - \gamma)] \end{aligned}$$

since Q is p.s.d. $\Rightarrow D' Q D$ is p.s.d. $\Rightarrow \text{tr} D' Q D \geq 0$. □

3.9 Corollary For any LUE of $\tilde{\gamma}$ of $\gamma = L\beta$,

$$\text{tr} V(\tilde{\gamma}) \geq \text{tr} V(\hat{\gamma}) .$$

PROOF

$$\begin{aligned} \text{tr} V(\tilde{\gamma}) &= \text{tr} E[(\tilde{\gamma} - \gamma) (\tilde{\gamma} - \gamma)'] = E[\text{tr} (\tilde{\gamma} - \gamma) (\tilde{\gamma} - \gamma)'] \\ &= E[(\tilde{\gamma} - \gamma)' (\tilde{\gamma} - \gamma)] \geq E[(\hat{\gamma} - \gamma)' (\hat{\gamma} - \gamma)] = \text{tr} V(\hat{\gamma}) \end{aligned}$$

by Corollary 3.8 with $Q = I$. □

3.10 Lemma PROPERTIES OF MATRIX DOMINANCE. *If $A = B + C$ where B is a p.d. matrix and C is a p.s.d. matrix, then*

- (a) A is p.d.,
- (b) $|B| \leq |A|$,
- (c) $B^{-1} - A^{-1}$ is p.s.d.

3.11 Corollary *Let L be an $r \times k$ fixed matrix, $\gamma = L\beta$ and $\hat{\gamma} = L\hat{\beta}$. Then*

$$|V(\tilde{\gamma})| \geq |V(\hat{\gamma})|$$

for any LUE $\tilde{\gamma}$ of γ .

PROOF Since $\hat{\gamma}$ is the BLUE of γ (by the generalized Gauss-Markov theorem), we have:

$$V(\tilde{\gamma}) = V(\hat{\gamma}) + C \tag{3.4}$$

where C is p.s.d. If $|V(\hat{\gamma})| = 0$, then $|V(\hat{\gamma})| \leq |V(\tilde{\gamma})|$, for $\text{car } |V(\tilde{\gamma})| \geq 0$. If $|V(\hat{\gamma})| > 0$, then $V(\hat{\gamma})$ is p.d. This entails that $V(\tilde{\gamma})$ is also p.d. and $|V(\hat{\gamma})| \leq |V(\tilde{\gamma})|$. \square

3.12 $\hat{y} = X\beta + P\varepsilon$, $\hat{\varepsilon} = My = M\varepsilon$.

PROOF

$$\begin{aligned} \hat{y} &= Py = P[X\beta + \varepsilon] = X\beta + P\varepsilon , \quad \text{car } PX = X , \\ \hat{\varepsilon} &= My = M[X\beta + \varepsilon] = M\varepsilon , \quad \text{car } MX = 0 . \end{aligned}$$

\square

3.13 $E(\hat{y}) = X\beta$, $E(\hat{\varepsilon}) = 0$.

PROOF

$$\begin{aligned} E(\hat{\gamma}) &= E[X\beta + P\varepsilon] = X\beta + PE(\varepsilon) = X\beta, \\ E(\hat{\varepsilon}) &= E(y - \hat{y}) = X\beta - X\beta = 0. \end{aligned}$$

□

3.14 $V(\hat{y}) = \sigma^2 P$, $V(\hat{\varepsilon}) = \sigma^2 M$.

PROOF

$$\begin{aligned} V(\hat{y}) &= V(X\hat{\beta}) = XV(\hat{\beta})X' = \sigma^2 X(X'X)^{-1}X' = \sigma^2 P, \\ V(\hat{\varepsilon}) &= V(My) = MV(y)M' = \sigma^2 M. \end{aligned}$$

□

3.15 \hat{y} is the best linear unbiased estimator of $X\beta$.

PROOF This follows directly on taking $L = X$ in the generalized Gauss-Markov theorem. □

3.16 $\hat{\varepsilon}$ is the best linear unbiased estimator (BLUE) of ε , in the sense that $E(\hat{\varepsilon} - \varepsilon) = 0$ and

$$V(\tilde{\varepsilon} - \varepsilon) - V(\hat{\varepsilon} - \varepsilon) \text{ is a p.s.d. matrix}$$

for for LUE $\tilde{\varepsilon}$ of ε .

PROOF Since $\tilde{\varepsilon}$ is a LUE of ε , we must have:

$$\tilde{\varepsilon} = Ay \quad \text{and} \quad E(\tilde{\varepsilon} - \varepsilon) = 0.$$

Consequently,

$$\begin{aligned} E(\tilde{\varepsilon}) &= E(Ay) \\ &= E[A(X\beta + \varepsilon)] = AX\beta = 0, \forall \beta, \end{aligned}$$

which entails that

$$\begin{aligned} AX &= 0, \\ \tilde{\varepsilon} &= A(X\beta + \varepsilon) = A\varepsilon. \end{aligned}$$

Let

$$B = A - M \quad \text{where} \quad M = I - X(X'X)^{-1}X'.$$

Then

$$AX = [B + M]X = BX = 0, \quad \text{since} \quad MX = 0,$$

hence

$$\begin{aligned} V(\tilde{\varepsilon} - \varepsilon) &= V[A\varepsilon - \varepsilon] \\ &= V[(B + M)\varepsilon - \varepsilon] = V[(B + M - I)\varepsilon] \\ &= E[(B + M - I)\varepsilon\varepsilon'(B' + M - I)] \\ &= \sigma^2[B - X(X'X)^{-1}X'][B' - X(X'X)^{-1}X'] \\ &= \sigma^2[BB' + X(X'X)^{-1}X'], \end{aligned}$$

and

$$\begin{aligned} V(\hat{\varepsilon} - \varepsilon) &= E[(M - I)\varepsilon\varepsilon'(M - I)] \\ &= \sigma^2(I - M) = \sigma^2X(X'X)^{-1}X', \end{aligned}$$

so that

$$V(\tilde{\varepsilon} - \varepsilon) = \sigma^2BB' + V(\hat{\varepsilon} - \varepsilon).$$

Thus

$$V(\tilde{\varepsilon} - \varepsilon) - V(\hat{\varepsilon} - \varepsilon) = \sigma^2BB'$$

a p.s.d. matrix. □

$$\mathbf{3.17} \quad C(\hat{\beta}, \hat{\varepsilon}) = C(\hat{\beta}, y - X\hat{\beta}) = 0.$$

PROOF

$$\begin{aligned} C(\hat{\beta}, \hat{\varepsilon}) &= E[(\hat{\beta} - \beta)\hat{\varepsilon}'] = E[(X'X)^{-1}X'\varepsilon\varepsilon'M] \\ &= \sigma^2(X'X)^{-1}X'M = 0. \end{aligned}$$

□

$$\mathbf{3.18} \quad C(\hat{y}, \hat{\varepsilon}) = 0.$$

PROOF

$$\begin{aligned} C(\hat{y}, \hat{\varepsilon}) &= E[(X\hat{\beta} - X\beta)\hat{\varepsilon}'] \\ &= XE[(\hat{\beta} - \beta)\hat{\varepsilon}'] = XC(\hat{\beta}, \hat{\varepsilon}) = 0. \end{aligned}$$

□

3.19 Estimation of σ^2 . Since $\sigma^2 = E(\varepsilon_t^2)$, $t = 1, \dots, T$, it is natural to consider the residuals of the regression which can be viewed as estimations of the error terms ε_t :

$$\begin{aligned} \hat{\varepsilon} &= y - X\hat{\beta} = My = M(X\beta + \varepsilon) = M\varepsilon, \\ \sum_{t=1}^T \hat{\varepsilon}_t^2 &= \hat{\varepsilon}'\hat{\varepsilon} = \varepsilon'M'M\varepsilon = \varepsilon'M\varepsilon, \end{aligned}$$

hence

$$E[\hat{\varepsilon}'\hat{\varepsilon}] = E[\varepsilon'M\varepsilon] = E[\text{tr}(\varepsilon'M\varepsilon)]$$

$$\begin{aligned}
&= E[\text{tr}(M\varepsilon\varepsilon')] = \text{tr}[ME(\varepsilon\varepsilon')] \\
&= \sigma^2 \text{tr}M,
\end{aligned}$$

where

$$\begin{aligned}
\text{tr}M &= \text{tr}[I_T - X(X'X)^{-1}X'] = \text{tr}I_T - \text{tr}[X(X'X)^{-1}X'] \\
&= \text{tr}I_T - \text{tr}[X'X(X'X)^{-1}] = \text{tr}I_T - \text{tr}I_k \\
&= T - k.
\end{aligned}$$

Thus,

$$\begin{aligned}
E(\hat{\varepsilon}'\hat{\varepsilon}) &= \sigma^2(T - k) \\
E\left[\frac{\hat{\varepsilon}'\hat{\varepsilon}}{T - k}\right] &= \sigma^2.
\end{aligned}$$

3.20 The statistic

$$s^2 = \hat{\varepsilon}'\hat{\varepsilon} / (T - k) = y'My / (T - k)$$

is an unbiased estimator of σ^2 , and $s^2(X'X)^{-1}$ is an unbiased estimator of $V(\hat{\beta}) = \sigma^2(X'X)^{-1}$:

$$\begin{aligned}
E(s^2) &= \sigma^2, \\
E\left[s^2(X'X)^{-1}\right] &= \sigma^2(X'X)^{-1}.
\end{aligned}$$

4. Prediction

In the previous section, we studied how one can estimate β in the linear regression model. Suppose now we know the matrix X_0 of explanatory variables for m additional periods (or observations). We wish to predict the corresponding values of y :

$$y_0 = X_0\beta + \varepsilon_0$$

where

$$E(\varepsilon_0) = 0, V(\varepsilon_0) = \sigma^2 I_m, E(\varepsilon\varepsilon'_0) = 0.$$

The natural “predictor” in this case is:

$$\hat{y}_0 = X_0\hat{\beta} = X_0(X'X)^{-1}X'y. \quad (4.1)$$

We can then show the following properties.

4.1 \hat{y}_0 is an unbiased estimator of $X_0\beta$:

$$E(\hat{y}_0) = X_0\beta = E(y_0), \quad E(\hat{y}_0 - y_0) = 0.$$

4.2 $V(\hat{y}_0) = V(X_0\hat{\beta}) = X_0V(\hat{\beta})X'_0 = \sigma^2X_0(X'X)^{-1}X'_0$.

4.3 $C(y_0, \hat{y}_0) = 0$.

PROOF

$$\begin{aligned} C(y_0, \hat{y}_0) &= E \left[(y_0 - X_0\beta) (X_0\hat{\beta} - X_0\beta)' \right] \\ &= E \left[\varepsilon_0 (\hat{\beta} - \beta)' X'_0 \right] = E \left[\varepsilon_0 \varepsilon' X (X'X)^{-1} X'_0 \right] = 0. \end{aligned}$$

□

4.4 \hat{y}_0 is best linear unbiased estimator of $X_0\beta$, in the sense that $V(\tilde{y}_0) - V(\hat{y}_0)$ is a p.s.d. matrix for any linear unbiased estimator \tilde{y}_0 of $X_0\beta$. In particular, if

$\tilde{y}_0 = Cy$ and $D = C - X_0(X'X)^{-1}X'$, then

$$V(\tilde{y}_0) = V(\hat{y}_0) + \sigma^2 DD' .$$

PROOF This follows directly from the generalized Gauss-Markov theorem. \square

The “prediction errors” are given by:

$$\begin{aligned} \hat{e}_0 &= y_0 - \hat{y}_0 = y_0 - X_0\hat{\beta} \\ &= X_0\beta + \varepsilon_0 - X_0\hat{\beta} = \varepsilon_0 + X_0(\beta - \hat{\beta}) . \end{aligned}$$

4.5 \hat{y}_0 is a linear unbiased predictor (LUP) of y_0 :

$$E[\hat{e}_0] = 0 .$$

PROOF $\hat{y}_0 = X_0\hat{\beta}$ and

$$E[\hat{e}_0] = E[y_0 - \hat{y}_0] = X_0\beta - X_0\beta = 0 .$$

\square

4.6 $V(\hat{e}_0) = \sigma^2 [I_m + X_0(X'X)^{-1}X_0']$.

PROOF

$$\begin{aligned} V(y_0 - \hat{y}_0) &= V(y_0) + V(\hat{y}_0) - C(y_0, \hat{y}_0) - C(\hat{y}_0, y_0) \\ &= \sigma^2 I_m + \sigma^2 X_0(X'X)^{-1}X_0' \\ &= \sigma^2 [I_m + X_0(X'X)^{-1}X_0'] . \end{aligned}$$

\square

4.7 Theorem \hat{y}_0 is the best linear unbiased predictor (BLUP) of y_0 , in the sense that $V(y_0 - \tilde{y}_0) - V(y_0 - \hat{y}_0)$ is a p.s.d. matrix for any LUP \tilde{y}_0 of y_0 . In particular, if $\tilde{y}_0 = Cy$ and $D = C - X_0(X'X)^{-1}X'$, then

$$V(y_0 - \tilde{y}_0) = V(y_0 - \hat{y}_0) + \sigma^2 DD' .$$

PROOF

$$V(y_0 - \tilde{y}_0) = V(y_0) + V(\tilde{y}_0) - C(y_0, \tilde{y}_0) - C(\tilde{y}_0, y_0)$$

where

$$C(y_0, \tilde{y}_0) = E[\varepsilon_0 \varepsilon' C'] = 0$$

for, by the generalized Gauss-Markov theorem,

$$E[\tilde{y}_0] = X_0 \beta \Rightarrow CX = X_0 \Rightarrow \tilde{y}_0 = C(X\beta + \varepsilon) = X_0 \beta + C\varepsilon .$$

Further, $V(\tilde{y}_0) = V(\hat{y}_0) + \sigma^2 DD'$ and $V(y_0) = \sigma^2 I_m$. Consequently,

$$\begin{aligned} V(y_0 - \tilde{y}_0) &= \sigma^2 I_m + V(\hat{y}_0) + \sigma^2 DD' \\ &= \left[\sigma^2 I_m + \sigma^2 X_0 (X'X)^{-1} X_0' \right] + \sigma^2 DD' \\ &= V(y_0 - \hat{y}_0) + \sigma^2 DD' . \end{aligned}$$

□

5. Estimation with Gaussian errors

If we wish to build confidence intervals and perform hypothesis tests, we need a more complete specification of the error distribution. The standard hypothesis for this is to assume that the errors follow a Gaussian distribution.

5.1 Assumption $\varepsilon \sim N_T [0, \sigma^2 I_T]$.

This means that the errors ε_t are i.i.d. $N [0, \sigma^2]$. We can now completely establish the distribution of the least squares estimator.

5.2 $y \sim N [X\beta, \sigma^2 I_T]$, since $y = X\beta + \varepsilon$.

5.3 $\hat{\beta} \sim N [\beta, \sigma^2 (X'X)^{-1}]$, since $\hat{\beta} = (X'X)^{-1} X'y$.

The probability density function of y is given by:

$$L(y; X\beta, \sigma^2 I_T) = \frac{1}{(2\pi\sigma^2)^{T/2}} \exp \left\{ -\frac{1}{2} \frac{(y - X\beta)'(y - X\beta)}{\sigma^2} \right\}.$$

5.4 $\hat{\beta} = (X'X)^{-1} X'y$ and $\hat{\sigma}^2 = \hat{\varepsilon}'\hat{\varepsilon}/T$ are the maximum likelihood estimators of β and σ^2 respectively.

PROOF To maximize L is equivalent to maximizing $\ln(L)$. Since

$$\begin{aligned} \ln(L) &= -\frac{T}{2} \ln(2\pi) - \frac{T}{2} \ln(\sigma^2) - \frac{1}{2\sigma^2} (y - X\beta)'(y - X\beta) \\ &= -\frac{T}{2} \ln(2\pi) - \frac{T}{2} \ln(\sigma^2) - \frac{1}{2\sigma^2} [y'y - 2y'X\beta + \beta'X'X\beta], \end{aligned}$$

the first-order conditions (which are necessary) for a maximum is:

$$\begin{aligned} \frac{\partial (\ln(L))}{\partial \beta} &= -\frac{1}{2\sigma^2} [-2X'y + 2(X'X)\beta] = 0, \\ \frac{\partial (\ln(L))}{\partial \sigma^2} &= -\frac{T}{2} \frac{1}{\sigma^2} + \frac{1}{2\sigma^4} (y - X\beta)'(y - X\beta) = 0, \end{aligned}$$

hence

$$\begin{aligned}(X'X)\hat{\beta} &= X'y, \hat{\beta} = (X'X)^{-1}X'y, \\ \hat{\sigma}^2 &= (y - X\hat{\beta})'(y - X\hat{\beta})/T.\end{aligned}$$

Further the second-order derivative of $\ln(L)$ is:

$$\frac{\partial(\ln(L))}{\partial\beta'\partial\beta} = -\frac{1}{\sigma^2}(X'X) \quad (5.1)$$

which is negative semidefinite as required for a maximum. \square

5.5 $\hat{y} = X\hat{\beta} \sim N_T [X\beta, \sigma^2P]$.

5.6 $\hat{\varepsilon} = M\varepsilon \sim N_T [0, \sigma^2M]$.

5.7 $\hat{\varepsilon}$ and $\hat{\beta}$ are independent, because $\hat{\varepsilon}$ et $\hat{\beta}$ are multinormal and $C(\hat{\beta}, \hat{\varepsilon}) = 0$.

5.8 $\hat{\varepsilon}$ and \hat{y} are independent, because $\hat{\varepsilon}$ and \hat{y} are multinormal and $C(\hat{y}, \hat{\varepsilon}) = 0$.

5.9 Lemma DISTRIBUTION OF AN IDEMPOTENT QUADRATIC FORM IN I.I.D. GAUSSIAN VARIABLES. *Let Q be a $T \times T$ symmetric idempotent matrix of rank $q \leq T$. If $\varepsilon \sim N_T [0, \sigma^2I_T]$, then*

$$\varepsilon'Q\varepsilon/\sigma^2 \sim \chi^2(q).$$

PROOF Since Q is a symmetric idempotent matrix, there is a $T \times T$ orthogonal matrix C , i.e. $CC' = C'C = I_T$, such that

$$CQC' = \begin{pmatrix} I_q & 0 \\ 0 & 0 \end{pmatrix},$$

hence

$$\varepsilon'Q\varepsilon = \varepsilon'C'CQC'C\varepsilon = (C\varepsilon)'(CQC')(C\varepsilon).$$

Further,

$$\begin{aligned}\boldsymbol{\varepsilon} &\sim N[0, \sigma^2 I_T] \Rightarrow C\boldsymbol{\varepsilon} \sim N[0, \sigma^2 C I_T C'] \\ &\Rightarrow C\boldsymbol{\varepsilon} \sim N[0, \sigma^2 I_T] .\end{aligned}$$

Let $\boldsymbol{v} = C\boldsymbol{\varepsilon} = (v_1, v_2, \dots, v_T)'$. Then

$$v_1, v_2, \dots, v_T \text{ are i.i.d. } N[0, \sigma^2]$$

and

$$\begin{aligned}\boldsymbol{\varepsilon}' Q \boldsymbol{\varepsilon} &= \boldsymbol{v}' (C Q C') \boldsymbol{v} \\ &= (v_1, v_2, \dots, v_T) \begin{pmatrix} I_q & 0 \\ 0 & 0 \end{pmatrix} \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_T \end{bmatrix} \\ &= v_1^2 + v_2^2 + \dots + v_q^2 + 0 \cdot v_{q+1}^2 \dots + 0 \cdot v_T^2 \\ &= \sum_{t=1}^q v_t^2 .\end{aligned}$$

This entails

$$\begin{aligned}\frac{\boldsymbol{\varepsilon}' Q \boldsymbol{\varepsilon}}{\sigma^2} &= \sum_{t=1}^q \left(\frac{v_t}{\sigma} \right)^2 , \\ \text{where } \frac{v_t}{\sigma} &\stackrel{\text{ind}}{\sim} N[0, 1] , \quad t = 1, \dots, T ,\end{aligned}$$

and

$$\boldsymbol{\varepsilon}' Q \boldsymbol{\varepsilon} / \sigma^2 \sim \chi^2(q) .$$

□

5.10

$$\frac{S(\hat{\beta})}{\sigma^2} = \frac{\hat{\varepsilon}'\hat{\varepsilon}}{\sigma^2} \sim \chi^2(T - k) .$$

PROOF This follows directly on applying Lemma 5.9 with $Q = M$ and the fact that $\text{tr}(M) = T - k$. \square

5.11 Let R be a $q \times k$ fixed matrix. Then,

$$R\hat{\beta} \sim N_q \left[R\beta, \sigma^2 R(X'X)^{-1} R' \right] . \quad (5.2)$$

Further $R\hat{\beta}$ and s^2 are independent.

PROOF $\hat{\beta} \sim N \left[\beta, \sigma^2 (X'X)^{-1} \right]$ entails $R\hat{\beta} \sim N \left[R\beta, \sigma^2 R(X'X)^{-1} R' \right]$. Since $\hat{\beta}$ and $\hat{\varepsilon}$ are independent, $R\hat{\beta}$ and $\hat{\varepsilon}'\hat{\varepsilon}$ are also independent, so that $R\hat{\beta}$ and $s^2 = \hat{\varepsilon}'\hat{\varepsilon} / (T - k)$ are independent. \square

5.12 Let R be a $q \times k$ fixed matrix of rank q , $r = R\beta$ and

$$S(R, \hat{\beta}) = [R\hat{\beta} - r]' \left[R(X'X)^{-1} R' \right]^{-1} [R\hat{\beta} - r] .$$

Then

$$S(R, \hat{\beta}) / \sigma^2 \sim \chi^2(q) . \quad (5.3)$$

Further, $S(R, \hat{\beta})$ and s^2 are independent.

PROOF

$$R\hat{\beta} - r = R(\hat{\beta} - \beta)$$

and

$$R(\hat{\beta} - \beta) \sim N_q \left[0, \sigma^2 R(X'X)^{-1} R' \right] .$$

Thus,

$$\begin{aligned} S(R, \hat{\beta})/\sigma^2 &= \left[R(\hat{\beta} - \beta) \right]' \left[\sigma^2 R(X'X)^{-1} R' \right]^{-1} \left[R(\hat{\beta} - \beta) \right] \\ &\sim \chi^2(q) . \end{aligned}$$

□

6. Confidence and prediction intervals

6.1. Confidence interval for the error variance

In the normal classical linear model, we have:

$$\hat{\varepsilon}'\hat{\varepsilon}/\sigma^2 = (T - k) s^2 / \sigma^2 \sim \chi^2(T - k) .$$

Thus, we can find a and b such that

$$\begin{aligned} \mathrm{P} [\chi^2(T - k) > b] &= \frac{\alpha}{2}, \\ \mathrm{P} [\chi^2(T - k) < a] &= \frac{\alpha}{2}, \\ \mathrm{P} [a \leq \chi^2(T - k) \leq b] &= 1 - \left(\frac{\alpha}{2} + \frac{\alpha}{2} \right) = 1 - \alpha, \end{aligned}$$

which entails that

$$\begin{aligned} \mathrm{P} \left[a \leq \frac{(T - k) s^2}{\sigma^2} \leq b \right] &= 1 - \alpha \\ \mathrm{P} \left[\frac{1}{b} \leq \frac{\sigma^2}{(T - k) s^2} \leq \frac{1}{a} \right] &= 1 - \alpha \\ \mathrm{P} \left[\frac{(T - k) s^2}{b} \leq \sigma^2 \leq \frac{(T - k) s^2}{a} \right] &= 1 - \alpha . \end{aligned}$$

It is important to note this is not the smallest confidence interval for σ^2 .

6.2. Confidence interval for a linear combination of regression coefficients

Consider now the linear combination $w'\beta$. Then

$$w'\hat{\beta} - w'\beta \sim N \left[0, \sigma^2 w' (X'X)^{-1} w \right],$$

hence

$$\frac{w'\hat{\beta} - w'\beta}{\sigma\Delta} \sim N[0, 1]$$

where $\Delta = \sqrt{w' (X'X)^{-1} w}$. Since σ is unknown, consider:

$$\begin{aligned} t &= \frac{w'\hat{\beta} - w'\beta}{s\Delta} \\ &= \frac{w'\hat{\beta} - w'\beta}{\Delta\sigma\sqrt{\frac{s^2}{\sigma^2}}} = \frac{w'\hat{\beta} - w'\beta}{\sigma\Delta} / \sqrt{\frac{(T-k)s^2}{\sigma^2(T-k)}} \\ &= Y / \sqrt{\frac{X}{T-k}} \end{aligned}$$

where X and Y are independent, $Y \sim N[0, 1]$ and $X \sim \chi^2(T-k)$. Thus, t follows a Student t distribution with $T-k$ degrees of freedom:

$$t \sim t(T-k)$$

hence

$$P \left[-t_{\alpha/2} \leq t(T-k) \leq t_{\alpha/2} \right] = 1 - \alpha$$

where $P \left[t(T-k) > t_{\alpha/2} \right] = \alpha/2$ and

$$P \left[w'\hat{\beta} - t_{\alpha/2}s\Delta \leq w'\beta \leq w'\hat{\beta} + t_{\alpha/2}s\Delta \right] = 1 - \alpha.$$

6.3. Confidence region for a regression coefficient vector

We now wish to build a confidence region for a vector $R\beta$ of linear combinations of the elements of β , where $R : q \times k$ and has rank q . Then

$$S(R, \hat{\beta})/\sigma^2 = (R\hat{\beta} - R\beta)' [R(X'X)^{-1}R']^{-1} (R\hat{\beta} - R\beta)/\sigma^2 \sim \chi^2(q) .$$

Since σ is unknown, let us consider:

$$F = S(R, \hat{\beta})/qs^2 = \frac{S(R, \hat{\beta})/q\sigma^2}{(T-k)s^2/\sigma^2} = \frac{X_1/q}{X_2/(T-k)}$$

where X_1 and X_2 are independent,

$$\begin{aligned} X_1 &= S(R, \hat{\beta})/\sigma^2 \sim \chi^2(q) , \\ X_2 &= (T-k)s^2/\sigma^2 \sim \chi^2(T-k) . \end{aligned}$$

Thus F follows a Fisher distribution with $(q, T-k)$ degrees of freedom:

$$F \sim F(q, T-k) .$$

If we define F_α by

$$P[F(q, T-k) > F_\alpha] = \alpha ,$$

the set of all vectors $R\beta$ such that $F \leq F_\alpha$:

$$(R\hat{\beta} - R\beta)' [R(X'X)^{-1}R']^{-1} (R\hat{\beta} - R\beta)/qs^2 \leq F_\alpha .$$

is a confidence region with level $1 - \alpha$ for $R\beta$. This set is a an ellipsoid (*confidence ellipsoid*).

6.4. Prediction intervals

$$y_0 = x_0' \beta + \varepsilon_0$$

where

$$\begin{pmatrix} \varepsilon \\ \varepsilon_0 \end{pmatrix} \sim N[0, \sigma^2 I_{T+1}] .$$

Further

$$\begin{aligned} \hat{y}_0 &= x_0' \hat{\beta}, \quad \hat{\beta} = (X'X)^{-1} X'y, \\ \hat{y}_0 - y_0 &= x_0'(\hat{\beta} - \beta) - \varepsilon_0 \sim N\{0, \sigma^2[1 + x_0'(X'X)^{-1}x_0]\}. \end{aligned}$$

hence

$$\frac{\hat{y}_0 - y_0}{\sigma \Delta_1} \sim N[0, 1] ,$$

where $\Delta_1 = \left[1 + x_0'(X'X)^{-1}x_0\right]^{1/2}$, and

$$\frac{\hat{y}_0 - y_0}{s \Delta_1} \sim t(T - k)$$

where $t_{\alpha/2}$ satisfies

$$P \left[\hat{y}_0 - t_{\alpha/2} s \Delta_1 \leq y_0 \leq \hat{y}_0 + t_{\alpha/2} s \Delta_1 \right] = 1 - \alpha .$$

6.5. Confidence regions for several predictions

We now consider the problem of predicting a vector of observations y_0 generated according to the same model independently of y :

$$y_0 = X_0\beta + \varepsilon_0 ,$$

$$\begin{pmatrix} \varepsilon \\ \varepsilon_0 \end{pmatrix} \sim N [0, \sigma^2 I_{T+m}] ,$$

where X_0 is known but y_0 is not observed. For predicting y_0 , let us define:

$$\hat{y}_0 = X_0\hat{\beta} ,$$

$$\hat{e}_0 = y_0 - \hat{y}_0 = \varepsilon_0 - X_0(\hat{\beta} - \beta) ,$$

where

$$E(\hat{e}_0) = 0 ,$$

$$V(\hat{e}_0) = \sigma^2 [I_m + X_0(X'X)^{-1}X_0'] = \sigma^2 D_0 ,$$

$$\hat{e}_0 \sim N [0, \sigma^2 [I_m + X_0(X'X)^{-1}X_0']] .$$

Consequently,

$$\hat{e}_0' V(\hat{e}_0)^{-1} \hat{e}_0 \sim \chi^2(m) ,$$

$$\hat{e}_0' D_0^{-1} \hat{e}_0 / \sigma^2 \sim \chi^2(m) .$$

Since σ^2 is unknown, we replace it by s^2 :

$$(T - k) s^2 / \sigma^2 \sim \chi^2(T - k) .$$

Further, since s^2 is independent of y_0 and $\hat{y}_0 = X\hat{\beta}$, s^2 is independent of \hat{e}_0 ,

$$F = \frac{\hat{e}_0' D_0^{-1} \hat{e}_0}{ms^2} = \frac{\hat{e}_0' D_0^{-1} \hat{e}_0 / \sigma^2 m}{(T - k) s^2 / \sigma^2 (T - k)} \sim F(m, T - k) ,$$

$$F = (y_0 - \hat{y}_0)' [I_m + X_0(X'X)^{-1}X_0']^{-1} (y_0 - \hat{y}_0) / ms^2 \sim F(m, T - k) .$$

Then the set of vectors y_0 such that

$$F \leq F_\alpha(m, T - k)$$

is a confidence region for y_0 with level $1 - \alpha$.

7. Hypothesis tests

7.0.1 Let us now consider the problem of testing an hypothesis of the form

$$H_0 : w' \beta = w_0 \quad (7.1)$$

where w be a $k \times 1$ vector of constants. To test H_0 , it is natural to consider the difference:

$$w' \hat{\beta} - w_0 = w' (\hat{\beta} - \beta) \sim N \left[0, \sigma^2 w' (X'X)^{-1} w \right].$$

Under the assumptions of the Gaussian classical linear model, we then have:

$$\frac{w' \hat{\beta} - w_0}{\sigma \Delta} \sim N[0, 1], \Delta = \left[w' (X'X)^{-1} w \right]^{1/2},$$

$$t = \frac{w' \hat{\beta} - w_0}{s \Delta} \sim t(T - k).$$

This suggests the following tests of H_0 :

reject H_0 at level α against $w' \beta - w_0 \neq 0$ when $|t| \geq t_{\alpha/2}$ (two-sided test) (7.2)

reject H_0 at level α against $w' \beta - w_0 > 0$ when $t \geq t_\alpha$ (one-sided test) (7.3)

reject H_0 at level α against $w' \beta - w_0 < 0$ when $t \leq -t_\alpha$ (one-sided test). (7.4)

An important special case of the above problem consists in testing the value of any given component of β :

$$H_0(\beta_{i_0}) : \beta_i = \beta_{i_0}$$

where β_i is an element of β .

Let us now consider the more general hypothesis which consists in testing the

value of a general vector linear transformation of β :

$$H_0 : R\beta = r = \begin{bmatrix} w'_1 \\ w'_2 \\ \vdots \\ w'_q \end{bmatrix} \beta = \begin{bmatrix} w'_1\beta \\ w'_2\beta \\ \vdots \\ w'_q\beta \end{bmatrix} \quad (7.5)$$

where R is a $q \times k$ fixed matrix with full row rank [$\text{rank}(R) = q$].

7.0.2 Wald-type test. A natural approach then consists in estimating $R\beta$ by $R\hat{\beta}$, and then to examine the difference $R\hat{\beta} - r$. Under H_0 ,

$$R\hat{\beta} \sim N[r, \Sigma_R], \quad \text{where} \quad \Sigma_R = \sigma^2 R(X'X)^{-1} R'.$$

We need a concept of distance between $R\hat{\beta}$ and r . By (5.3),

$$W = (R\hat{\beta} - r)' \Sigma_R^{-1} (R\hat{\beta} - r) \sim \chi^2(q) \quad \text{under } H_0.$$

We tend to reject H_0 when W is too large ($W \geq c$). However, σ^2 and Σ_R are unknown. It is then natural to replace σ^2 by the estimate s^2 , and Σ_R by

$$\hat{\Sigma}_R = s^2 R(X'X)^{-1} R'.$$

This yields a Wald-type criterion:

$$\begin{aligned} \hat{W} &= (R\hat{\beta} - r)' \hat{\Sigma}_R^{-1} (R\hat{\beta} - r) \\ &= (R\hat{\beta} - r)' \left[s^2 R(X'X)^{-1} R' \right]^{-1} (R\hat{\beta} - r) \\ &= (R\hat{\beta} - r)' \left[R(X'X)^{-1} R' \right]^{-1} (R\hat{\beta} - r) / s^2 \\ &= S(R, \hat{\beta}) / s^2. \end{aligned}$$

Since

$$F = \hat{W} / q = S(R, \hat{\beta}) / qs^2 \sim F(q, T - k),$$

we reject H_0 at level α when

$$F > F_\alpha(q, T - k) . \quad (7.6)$$

7.0.3 Likelihood ratio test. Another approach to testing H_0 consists in looking for a likelihood ratio test. This test is based on focusing on the likelihood function:

$$L(y; X\beta, \sigma^2 I_T) = \frac{1}{(2\pi\sigma^2)^{T/2}} \exp \left\{ -\frac{1}{2} \frac{(y - X\beta)'(y - X\beta)}{\sigma^2} \right\} . \quad (7.7)$$

Let

$$L(\hat{\Omega}) = \max_{\beta, \sigma^2} L = \max_{(\beta, \sigma^2) \in \Omega} L \quad (7.8)$$

i.e. we find values of β and σ^2 which maximize “the probability of the observed sample”, and

$$L(\hat{\omega}) = \max_{\substack{\beta, \sigma^2 \\ R\beta=r}} L = \max_{(\beta, \sigma^2) \in \omega} L \quad (7.9)$$

i.e. we find values of β and σ^2 which maximize “the probability of the observed sample” and satisfy H_0 , where

$$\begin{aligned} \Omega &= \{ (\beta, \sigma^2) : -\infty < \beta_i < +\infty, i = 1, \dots, k, 0 < \sigma^2 < +\infty \} , \\ \omega &= \{ (\beta, \sigma^2) \in \Omega : R\beta = r \} . \end{aligned}$$

We see easily that

$$0 \leq L(\hat{\omega}) \leq L(\hat{\Omega}) ,$$

hence

$$\begin{aligned} 0 &\leq \frac{L(\hat{\omega})}{L(\hat{\Omega})} \leq 1 , \\ \frac{L(\hat{\Omega})}{L(\hat{\omega})} &\geq 1 . \end{aligned}$$

We reject H_0 when

$$LR(y) \equiv \frac{L(\hat{\Omega})}{L(\hat{\omega})} \geq \lambda_\alpha ,$$

where λ_α depends on the level of the test:

$$P[LR(y) \geq \lambda_\alpha] = \alpha .$$

7.0.4 $L(\hat{\Omega})$ is achieved when $\beta = \hat{\beta}$ and $\sigma^2 = \hat{\sigma}^2$:

$$\begin{aligned} L(\hat{\Omega}) &= \frac{1}{(2\pi\hat{\sigma}^2)^{T/2}} \exp \left\{ -\frac{1}{2} \frac{(y - X\hat{\beta})' (y - X\hat{\beta})}{\hat{\sigma}^2} \right\} = \frac{1}{(2\pi\hat{\sigma}^2)^{T/2}} \exp \left\{ -\frac{T}{2} \right\} \\ &= \frac{e^{-T/2}}{[2\pi\hat{\sigma}^2]^{T/2}} = \frac{T^{T/2} e^{-T/2}}{(2\pi)^{T/2} \left[(y - X\hat{\beta})' (y - X\hat{\beta}) \right]^{T/2}} \\ &= \frac{T^{T/2} e^{-T/2}}{(2\pi)^{T/2} S_\Omega^{T/2}} , \end{aligned}$$

where $S_\Omega = (y - X\hat{\beta})' (y - X\hat{\beta})$.

7.0.5 To find $L(\hat{\omega})$, it is equivalent to maximize

$$\ln(L) = -\frac{T}{2} \ln(2\pi) - \frac{T}{2} \ln(\sigma^2) - \frac{1}{2\sigma^2} (y - X\beta)' (y - X\beta)$$

under the constraint $R\beta = r$. Consider σ^2 as given. It is then sufficient to solve the problem:

$$\text{Min}_\beta (y - X\beta)' (y - X\beta)$$

with restriction $r - R\beta = 0$. To do this, we consider the Lagrangian function:

$$\mathcal{L} = (y - X\beta)' (y - X\beta) - \lambda' [r - R\beta] .$$

The optimum $\beta = \tilde{\beta}$ must satisfy the first-order conditions:

$$\frac{\partial \mathcal{L}}{\partial \beta} = -2X'y + 2(X'X)\tilde{\beta} + R'\lambda = 0 \quad (7.10)$$

$$\frac{\partial \mathcal{L}}{\partial \lambda} = r - R\tilde{\beta} = 0. \quad (7.11)$$

On multiplying by (7.10) by $R(X'X)^{-1}$, we get:

$$\begin{aligned} -2R(X'X)^{-1}X'y + 2R\tilde{\beta} + R(X'X)^{-1}R'\lambda &= 0 \\ R(X'X)^{-1}R'\lambda &= 2R(X'X)^{-1}X'y - 2r = 2[R\hat{\beta} - r] \\ \lambda &= 2[R(X'X)^{-1}R']^{-1}[R\hat{\beta} - r]. \end{aligned}$$

By (7.10),

$$2(X'X)\tilde{\beta} = 2X'y - R'\lambda \quad (7.12)$$

$$= 2X'y - 2R'[R(X'X)^{-1}R']^{-1}[R\hat{\beta} - r] \quad (7.13)$$

hence

$$\begin{aligned} \tilde{\beta} &= (X'X)^{-1}X'y - (X'X)^{-1}R'[R(X'X)^{-1}R']^{-1}[R\hat{\beta} - r] \\ &= \hat{\beta} + (X'X)^{-1}R'[R(X'X)^{-1}R']^{-1}[r - R\hat{\beta}]. \end{aligned}$$

We see that $\tilde{\beta}$ does not depend on σ^2 . Substituting $\tilde{\beta}$ in $\ln(L)$, we see that

$$\ln(L) = -\frac{T}{2}\ln(2\pi) - \frac{T}{2}\ln\sigma^2 - \frac{1}{2\sigma^2}S_\omega$$

where $S_\omega = (y - X\tilde{\beta})'(y - X\tilde{\beta})$, from which we get

$$\frac{\partial \ln(L)}{\partial \sigma^2} = -\frac{T}{2\sigma^2} + \frac{S_\omega}{2\sigma^4} = 0$$

at the optimum, hence

$$\begin{aligned}\tilde{\sigma}^2 &= S_\omega/T = (y - X\tilde{\beta})' (y - X\tilde{\beta}) / T, \\ L(\hat{\omega}) &= \frac{T^{T/2} e^{-T/2}}{(2\pi)^{T/2} S_\omega^{T/2}},\end{aligned}$$

The likelihood ratio test is given by the critical region:

$$\frac{L(\hat{\Omega})}{L(\hat{\omega})} = \left(\frac{S_\omega}{S_\Omega} \right)^{T/2} \geq \lambda_\alpha$$

or, equivalently,

$$\frac{S_\omega}{S_\Omega} \geq \lambda_\alpha^{2/T}. \quad (7.14)$$

Since

$$\begin{aligned}S_\omega &= (y - X\tilde{\beta})'(y - X\tilde{\beta}) \\ &= (y - X\hat{\beta})'(y - X\hat{\beta}) + (\hat{\beta} - \tilde{\beta})'(X'X)(\hat{\beta} - \tilde{\beta}) \\ &= S_\Omega + (\hat{\beta} - \tilde{\beta})'(X'X)(\hat{\beta} - \tilde{\beta}),\end{aligned}$$

we also see that

$$\begin{aligned}S_\omega - S_\Omega &= (r - R\hat{\beta})' \left[R(X'X)^{-1} R' \right]^{-1} R(X'X)^{-1} (X'X)(X'X)^{-1} \\ &\quad R' \left[R(X'X)^{-1} R' \right]^{-1} [r - R\hat{\beta}] \\ &= (r - R\hat{\beta})' \left[R(X'X)^{-1} R' \right]^{-1} [r - R\hat{\beta}] \\ &= (R\hat{\beta} - r)' \left[R(X'X)^{-1} R' \right]^{-1} (R\hat{\beta} - r) = S(R, \hat{\beta}) \\ &= (qs^2) F,\end{aligned}$$

hence

$$F = \frac{S_\omega - S_\Omega}{qs^2} = \frac{(S_\omega - S_\Omega)/q}{S_\Omega/(T - k)}$$

and

$$\frac{S_\omega}{S_\Omega} = \frac{S_\Omega + (qs^2)F}{S_\Omega} = 1 + \frac{(qs^2)F}{(T-k)s^2} = 1 + \frac{q}{T-k}F \geq \lambda_\alpha^{2/T}$$

$$\iff F \geq \frac{T-k}{q} \left(\lambda_\alpha^{2/T} - 1 \right) = F_\alpha .$$

The likelihood ratio test of $H_0 : R\beta = r$ has the critical region

$$F \equiv \frac{(S_\omega - S_\Omega)/q}{S_\Omega/(T-k)} \geq F_\alpha(q, T-k)$$

where

$$F \sim F(q, T-k) .$$

This is an easy method for testing $H_0 : R\beta = r$. Note also that:

$$LR = \left(\frac{S_\omega}{S_\Omega} \right)^{T/2} = \left(1 + \frac{q}{T-k}F \right)^{T/2} ,$$

$$F = \frac{T-k}{q} \left(LR^{2/T} - 1 \right) .$$

8. Estimator optimal properties with Gaussian errors

When errors are Gaussian, the OLS estimators $\hat{\beta}_i, i = 1, \dots, k$ and $s^2 = \left(y - X\hat{\beta} \right)' \left(y - X\hat{\beta} \right) / (T-k)$ have minimum variance in the class of all unbiased estimators of $\beta_i, i = 1, \dots, k$, and σ^2 respectively [see Rao (1973, section 5a)].

References

Maddala, G. S. (1977), *Econometrics*, McGraw-Hill, New York.

Rao, C. R. (1973), *Linear Statistical Inference and its Applications*, second edn, John Wiley & Sons, New York.

Schmidt, P. (1976), *Econometrics*, Marcel Dekker, New York.

Theil, H. (1971), *Principles of Econometrics*, John Wiley & Sons, New York.