Classical linear model *

Jean-Marie Dufour †
McGill University

First version: April 1982
Revised: April 1995, February 2002, July 2011
This version: July 2011
Compiled: October 18, 2011, 21:25

* This work was supported by the William Dow Chair in Political Economy (McGill University), the Canada Research Chair Program (Chair in Econometrics, Université de Montréal), the Bank of Canada (Research Fellowship), a Guggenheim Fellowship, a Konrad-Adenauer Fellowship (Alexander-von-Humboldt Foundation, Germany), the Institut de finance mathématique de Montréal (IFM2), 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 519, 855 Sherbrooke Street West, Montréal, Québec H3A 2T7, Canada. TEL: (1) 514 398 8879; FAX: (1) 514 398 4938; 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. Notations

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

$$y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_T \end{pmatrix} : \text{dependent variable (to explain)}$$

$$x_i = \begin{pmatrix} x_{1i} \\ x_{2i} \\ \vdots \\ x_{Ti} \end{pmatrix}, \quad i = 1, \ldots, k : \text{explanatory variables.}$$

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

$$X = [x_1, x_2, \ldots, x_k] = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1k} \\ x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & & \vdots \\ x_{T1} & x_{T2} & \cdots & x_{Tk} \end{bmatrix} = \begin{bmatrix} X_1' \\ X_2' \\ \vdots \\ X_T' \end{bmatrix},$$

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

$$X_t' = (x_{t1}, x_{t2}, \ldots, x_{tk}), \quad t = 1, \ldots, T .$$

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

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

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:

$$y = X\beta + \varepsilon \tag{1.2}$$

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

1.2. The least squares problem

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

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

We consider the problem:

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

Since

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

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

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

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

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

1.2.4 The model residuals are:

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

where

$$PX = X, \; MX = 0, \tag{1.6}$$

$$PM = P(I - P) = 0, \; MP = 0. \tag{1.7}$$

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

$$X'M = 0,$$
$$x_i'M = 0, \quad i = 1, \ldots, k.$$

Residuals and regressors are orthogonal:

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

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

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

1.2.6 Fitted values and residuals are orthogonal:

$$\hat{y}'\hat{\varepsilon} = y'PMy = 0. \tag{1.8}$$

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

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

1.2.8 For any vector β,

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

for

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

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 $\operatorname{rank}(X) = k < T$.

From the assumption 2.1 - 2.4, we see that:

$$
\begin{aligned}
\mathsf{E}(y) &= \mathsf{E}(y \mid X) = X\beta = \begin{pmatrix} X_1'\beta \\ \vdots \\ X_T'\beta \end{pmatrix} \\
&= (x_1, x_2, \ldots, x_k) \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{pmatrix} \\
&= x_1\beta_1 + x_2\beta_2 + \cdots + x_k\beta_k, \\
\mathsf{V}(y) &= \mathsf{V}(y \mid X) = \sigma^2 I_T \\
&= \begin{bmatrix} \sigma^2 & 0 & \cdots & 0 \\ 0 & \sigma^2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \sigma^2 \end{bmatrix} = \mathsf{V}(\varepsilon) \ .
\end{aligned}
$$

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

PROOF

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

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 linear unbiased estimators (BLUE) of β, i.e. $\mathsf{V}(\tilde{\beta}) - \mathsf{V}(\hat{\beta})$ 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*

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

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

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

we have:

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

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

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

and

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

is a positive semidefinite matrix. □

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

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

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

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

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

for $w'DD'w \geq 0$. □

In particular, we must have:

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

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. $\mathsf{V}(\tilde{\gamma}) - \mathsf{V}(\hat{\gamma})$ 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*

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

and

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

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

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

we have

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

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

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

so

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

is a positive semidefinite matrix, and

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

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

□

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*

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

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

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

$$\begin{aligned}
\mathsf{E}\left[(\tilde{\gamma}-\gamma)' Q (\tilde{\gamma}-\gamma)\right] &= \mathsf{E}\left[\operatorname{tr} Q(\tilde{\gamma}-\gamma)(\tilde{\gamma}-\gamma)'\right] \\
&= \operatorname{tr} Q \mathsf{E}\left[(\tilde{\gamma}-\gamma)(\tilde{\gamma}-\gamma)'\right] \\
&= \operatorname{tr} Q\left[\sigma^2 DD' + \mathsf{V}(\hat{\gamma})\right] \\
&= \sigma^2 \operatorname{tr}(QDD') + \operatorname{tr}\left[Q\mathsf{V}(\hat{\gamma})\right] \\
&= \sigma^2 \operatorname{tr}(D'QD) + \operatorname{tr} Q\mathsf{E}\left[(\hat{\gamma}-\gamma)(\hat{\gamma}-\gamma)'\right] \\
&= \sigma^2 \operatorname{tr}(D'QD) + \mathsf{E}\left[\operatorname{tr}(\hat{\gamma}-\gamma)' Q (\hat{\gamma}-\gamma)\right] \\
&= \sigma^2 \operatorname{tr}(D'QD) + \mathsf{E}\left[(\hat{\gamma}-\gamma)' Q (\hat{\gamma}-\gamma)\right] \\
&\geq \mathsf{E}\left[(\hat{\gamma}-\gamma)' Q (\hat{\gamma}-\gamma)\right]
\end{aligned}$$

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

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

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

PROOF

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

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*

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

for any LUE $\tilde{\gamma}$ of γ.

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

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

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

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

PROOF

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

□

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

PROOF

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

□

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

PROOF

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

□

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

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

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

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

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

Consequently,

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

which entails that

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

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

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

and

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

so that

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

Thus

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

a p.s.d. matrix. □

3.17 $\mathsf{C}\left(\hat{\beta},\hat{\varepsilon}\right)=\mathsf{C}\left(\hat{\beta},y-X\hat{\beta}\right)=0$.

PROOF

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

□

3.18 $\mathsf{C}\left(\hat{y},\hat{\varepsilon}\right)=0$.

PROOF

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

□

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

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

hence

$$\mathsf{E}\left[\hat{\varepsilon}'\hat{\varepsilon}\right] = \mathsf{E}\left[\varepsilon'M\varepsilon\right]=\mathsf{E}\left[\mathrm{tr}\left(\varepsilon'M\varepsilon\right)\right]$$

$$
\begin{aligned}
&= \mathsf{E}\left[\operatorname{tr}\left(M\varepsilon\varepsilon'\right)\right] = \operatorname{tr}\left[M\mathsf{E}\left(\varepsilon\varepsilon'\right)\right] \\
&= \sigma^2 \operatorname{tr} M \, ,
\end{aligned}
$$

where

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

Thus,

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

3.20 The statistic

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

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

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

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

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

The natural "predictor" in this case is:

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

We can then show the following properties.

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

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

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

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

PROOF

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

□

4.4 $\hat{y}_0$ is best linear unbiased estimator of $X_0\beta$, in the sense that $\mathsf{V}(\tilde{y}_0) - \mathsf{V}(\hat{y}_0)$ 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

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

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

The "prediction errors" are given by:

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

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

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

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

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

□

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

PROOF

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

□

4.7 Theorem $\hat{y}_0$ *is the best linear unbiased predictor (BLUP) of* y_0*, in the sense that* $\mathsf{V}(y_0 - \tilde{y}_0) - \mathsf{V}(y_0 - \hat{y}_0)$ *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*

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

PROOF

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

where

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

for, by the generalized Gauss-Markov theorem,

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

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

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

□

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 $\varepsilon \sim N_T\left[0, \sigma^2 I_T\right]$.

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

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

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

The probability density function of y is given by:

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

5.4 $\hat{\beta} = \left(X'X\right)^{-1} X'y$ 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

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

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

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

hence

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

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

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

which is negative semidefinite as required for a maximum. □

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

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

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

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

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 $\varepsilon \sim N_T\left[0, \sigma^2 I_T\right]$, 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,

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

Let $v = C\varepsilon = (v_1, v_2, \ldots, v_T)'$. Then

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

and

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

This entails

$$\frac{\varepsilon' Q \varepsilon}{\sigma^2} = \sum_{t=1}^{q} \left(\frac{v_t}{\sigma}\right)^2 \ ,$$

$$\text{where } \frac{v_t}{\sigma} \overset{ind}{\sim} N[0,1] \ , \quad t = 1, \ldots, T \ ,$$

and

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

□

5.10

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

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

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

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

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

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

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

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

Then

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

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

PROOF

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

and

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

Thus,

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

□

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

$$\mathsf{P}\left[\chi^2\,(T-k) > b\right] = \frac{\alpha}{2}\,,$$
$$\mathsf{P}\left[\chi^2\,(T-k) < a\right] = \frac{\alpha}{2}\,,$$
$$\mathsf{P}\left[a \le \chi^2\,(T-k) \le b\right] = 1 - \left(\frac{\alpha}{2} + \frac{\alpha}{2}\right) = 1-\alpha\,,$$

which entails that

$$\mathsf{P}\left[a \le \frac{(T-k)\,s^2}{\sigma^2} \le b\right] = 1-\alpha$$
$$\mathsf{P}\left[\frac{1}{b} \le \frac{\sigma^2}{(T-k)\,s^2} \le \frac{1}{a}\right] = 1-\alpha$$
$$\mathsf{P}\left[\frac{(T-k)\,s^2}{b} \le \sigma^2 \le \frac{(T-k)\,s^2}{a}\right] = 1-\alpha\ .$$

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

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

hence

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

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

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

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

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

hence

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

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

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

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

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

Since σ is unknown, let us consider:

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

where X_1 and X_2 are independent,

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

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

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

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

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

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

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

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

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

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

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

where $t_{\alpha/2}$ satisfies

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

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 :

$$y_0 = X_0\beta + \varepsilon_0 \ ,$$
$$\begin{pmatrix} \varepsilon \\ \varepsilon_0 \end{pmatrix} \sim N\left[0, \sigma^2 I_{T+m}\right] \ ,$$

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

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

where

$$\begin{aligned} \mathsf{E}\,(\hat{e}_0) &= 0 \, , \\ \mathsf{V}\,(\hat{e}_0) &= \sigma^2 \left[I_m + X_0\,(X'X)^{-1} X_0'\right] = \sigma^2 D_0 \, , \\ \hat{e}_0 &\sim N\left[0, \, \sigma^2 [I_m + X_0\,(X'X)^{-1} X_0']\right] \ . \end{aligned}$$

Consequently,

$$\begin{aligned} \hat{e}_0' \mathsf{V}\,(\hat{e}_0)^{-1}\,\hat{e}_0 &\sim \chi^2\,(m) \ , \\ \hat{e}_0' D_0^{-1} \hat{e}_0 / \sigma^2 &\sim \chi^2\,(m) \ . \end{aligned}$$

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 $\hat{e}_0$,

$$\begin{aligned} F &= \frac{\hat{e}_0' D_0^{-1} \hat{e}_0}{m s^2} = \frac{\hat{e}_0' D_0^{-1} \hat{e}_0 / \sigma^2 m}{(T-k)\,s^2/\sigma^2\,(T-k)} \sim F\,(m, \, T-k) \ , \\ F &= (y_0 - \hat{y}_0)' \left[I_m + X_0\,(X'X)^{-1} X_0'\right]^{-1} (y_0 - \hat{y}_0) \,/ m s^2 \sim F\,(m, \, T-k) \ . \end{aligned}$$

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

$$H_0 : w'\beta = w_0 \tag{7.1}$$

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

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

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

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

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

This suggests the following tests of H_0 :

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

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

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

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

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

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 :

$$H_0 : R\beta = r = \begin{bmatrix} w_1' \\ w_2' \\ \vdots \\ w_q' \end{bmatrix} \beta = \begin{bmatrix} w_1'\beta \\ w_2'\beta \\ \vdots \\ w_q'\beta \end{bmatrix} \tag{7.5}$$

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

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 tom 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

$$F > F_{\alpha}(q, T-k) \ . \tag{7.6}$$

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:

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

Let

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

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

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

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

$$\Omega = \left\{\left(\beta, \sigma^2\right) : -\infty < \beta_i < +\infty,\ i = 1, \ldots, k,\ 0 < \sigma^2 < +\infty\right\} \ ,$$
$$\omega = \left\{\left(\beta, \sigma^2\right) \in \Omega : R\beta = r\right\} \ .$$

We see easily that

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

hence

$$0 \le \frac{L(\hat{\omega})}{L(\hat{\Omega})} \le 1 \, ,$$
$$\frac{L(\hat{\Omega})}{L(\hat{\omega})} \ge 1 \ .$$

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:

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

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

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

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

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

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

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

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

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

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

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

$$\frac{\partial \mathscr{L}}{\partial \beta} = -2X'y + 2\left(X'X\right)\tilde{\beta} + R'\lambda = 0 \qquad (7.10)$$

$$\frac{\partial \mathscr{L}}{\partial \lambda} = r - R\tilde{\beta} = 0\,. \qquad (7.11)$$

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

$$-2R\left(X'X\right)^{-1}X'y + 2R\tilde{\beta} + R\left(X'X\right)^{-1}R'\lambda = 0$$

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

$$\lambda = 2\left[R\left(X'X\right)^{-1}R'\right]^{-1}\left[R\hat{\beta} - r\right]\ .$$

By (7.10),

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

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

hence

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

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

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 $S_\omega = \left(y - X\tilde{\beta}\right)'\left(y - X\tilde{\beta}\right)$, 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

$$\tilde{\sigma}^2 = S_\omega / T = \left(y - X\tilde{\beta}\right)' \left(y - X\tilde{\beta}\right) / T \ ,$$

$$L(\hat{\omega}) = \frac{T^{T/2} e^{-T/2}}{(2\pi)^{T/2} S_\omega^{T/2}} \ ,$$

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,

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

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

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

hence

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

and

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

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 $\hat{\beta}_i, i = 1, \ldots, k$ 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 β_i, $i = 1, \ldots, k$, 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.