Covariance matrices and multiple linear regression between random variables *

Jean-Marie Dufour †
McGill University

First version: July 2021
Revised: January 2022, January 2024
This version: January 2024
Compiled: January 17, 2024, 19:37

* 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 Universitad Carlos III de Madrid (Banco Santander de Madrid Chair of excellence), 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

List of Definitions, Assumptions, Propositions and Theorems

1. Covariance matrices

Consider now k random variables $X_1, X_2, \ldots, X_k$ such that

$$\mathbb{E}(X_i) = \mu_i\,, \;\; i = 1, \ldots, k\,, \tag{1.1}$$

$$\mathsf{C}(X_i, X_j) = \sigma_{ij}\,, \;\; i, j = 1, \ldots, k\,. \tag{1.2}$$

We often wish to compute the mean and variance of a linear combination of $X_1, \ldots, X_k$:

$$\sum_{i=1}^{k} a_i X_i = a_1 X_1 + a_2 X_2 + \cdots + a_k X_k\,. \tag{1.3}$$

It is easy to verify that

$$\mathbb{E}\left[\sum_{i=1}^{k} a_i X_i\right] = \sum_{i=1}^{k} a_i \mu_i \tag{1.4}$$

and

$$\begin{aligned} \mathbb{V}\left[\sum_{i=1}^{k} a_i X_i\right] &= \mathbb{E}\left\{\left[\sum_{i=1}^{k} a_i (X_i - \mu_i)\right]\left[\sum_{j=1}^{k} a_j \left(X_j - \mu_j\right)\right]\right\} \\ &= \sum_{i=1}^{k}\sum_{j=1}^{k} a_i a_j \sigma_{ij}\,. \end{aligned} \tag{1.5}$$

Since such formulae may often become cumbersome, it will be convenient to use vector and matrix notation

We define a random vector $\boldsymbol{X}$ and its mean value $\mathbb{E}(\boldsymbol{X})$ by:

$$\boldsymbol{X} = \begin{pmatrix} X_1 \\ \vdots \\ X_k \end{pmatrix} \;,\; \mathbb{E}(\boldsymbol{X}) = \begin{pmatrix} \mathbb{E}(X_1) \\ \vdots \\ \mathbb{E}(X_k) \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_k \end{pmatrix} \equiv \mu_X\,. \tag{1.6}$$

Similarly, we define a random matrix M and its mean value $\mathbb{E}(M)$ by:

$$\boldsymbol{M} = \begin{bmatrix} X_{11} & X_{12} & \ldots & X_{1n} \\ X_{21} & X_{22} & \ldots & X_{2n} \\ \vdots & \vdots & & \vdots \\ X_{m1} & X_{m2} & \ldots & X_{mn} \end{bmatrix} , \mathbb{E}(\boldsymbol{M}) = \begin{bmatrix} \mathbb{E}(X_{11}) & \mathbb{E}(X_{12}) & \ldots & \mathbb{E}(X_{1n}) \\ \mathbb{E}(X_{21}) & \mathbb{E}(X_{22}) & \ldots & \mathbb{E}(X_{2n}) \\ \vdots & \vdots & & \vdots \\ \mathbb{E}(X_{m1}) & \mathbb{E}(X_{m2}) & \ldots & \mathbb{E}(X_{mn}) \end{bmatrix} \tag{1.7}$$

where the X_{ij} are random variables. To a random vector $\boldsymbol{X}$, we can associate a *covariance matrix* $\mathbb{V}(\boldsymbol{X})$:

$$\boldsymbol{\Sigma}(\boldsymbol{X}) := \mathbb{V}(\boldsymbol{X}) := \mathbb{E}\left\{[\boldsymbol{X} - \mathbb{E}(\boldsymbol{X})]\,[\boldsymbol{X} - \mathbb{E}(\boldsymbol{X})]'\right\} = \mathbb{E}\{[\boldsymbol{X} - \mu_X]\,[\boldsymbol{X} - \mu_X]'\}$$

$$
= \mathbb{E}\left\{\begin{bmatrix} (X_1-\mu_1)(X_1-\mu_1) & (X_1-\mu_1)(X_2-\mu_2) & \dots & (X_1-\mu_1)(X_k-\mu_k) \\ \vdots & \vdots & & \vdots \\ (X_k-\mu_k)(X_1-\mu_1) & (X_k-\mu_k)(X_2-\mu_2) & \dots & (X_k-\mu_k)(X_k-\mu_k) \end{bmatrix}\right\}
$$

$$
= \begin{bmatrix} \sigma_{11} & \sigma_{12} & \dots & \sigma_{1k} \\ \vdots & \vdots & & \vdots \\ \sigma_{k1} & \sigma_{k2} & \dots & \sigma_{kk} \end{bmatrix} = \boldsymbol{\Sigma}\,. \tag{1.8}
$$

Similarly, we define the *uncentered covariance matrix* of $\boldsymbol{X}$:

$$
\bar{\boldsymbol{\Sigma}}(\boldsymbol{X}) := \mathbb{E}(\boldsymbol{X}\boldsymbol{X}')\,. \tag{1.9}
$$

If $\mathbf{a} = (a_1, \dots, a_k)'$, we see that:

$$
\sum_{i=1}^{k} a_i X_i = \mathbf{a}'\boldsymbol{X}\,. \tag{1.10}
$$

Basic properties of $\mathbb{E}(\boldsymbol{X})$ and $\mathbb{V}(\boldsymbol{X})$ are summarized by the following proposition.

Proposition 1.1 *Let $\boldsymbol{X} = (X_1, \dots, X_k)'$ a $k \times 1$ random vector, α a scalar, $\mathbf{a}$ and $\mathbf{b}$ fixed $k \times 1$ vectors, and $\boldsymbol{A}$ a fixed $g \times k$ matrix. Then, provided the moments considered are finite, we have the following properties*:

(*a*) $\mathbb{E}(\boldsymbol{X} + \mathbf{a}) = \mathbb{E}(\boldsymbol{X}) + \mathbf{a}$;

(*b*) $\mathbb{E}(\alpha\boldsymbol{X}) = \alpha\mathbb{E}(\boldsymbol{X})$;

(*c*) $\mathbb{E}(\mathbf{a}'\boldsymbol{X}) = \mathbf{a}'\mathbb{E}(\boldsymbol{X})$, $\mathbb{E}(\boldsymbol{A}\boldsymbol{X}) = \boldsymbol{A}\mathbb{E}(\boldsymbol{X})$;

(*d*) $\mathbb{V}(\boldsymbol{X} + \mathbf{a}) = \mathbb{V}(\boldsymbol{X})$;

(*e*) $\mathbb{V}(\alpha\boldsymbol{X}) = \alpha^2\mathbb{V}(\boldsymbol{X})$;

(*f*) $\mathbb{V}(\mathbf{a}'\boldsymbol{X}) = \mathbf{a}'\mathbb{V}(\boldsymbol{X})\mathbf{a}$, $\mathbb{V}(\boldsymbol{A}\boldsymbol{X}) = \boldsymbol{A}\mathbb{V}(\boldsymbol{X})\boldsymbol{A}'$;

(*g*) $\mathsf{C}(\mathbf{a}'\boldsymbol{X}, \mathbf{b}'\boldsymbol{X}) = \mathbf{a}'\mathbb{V}(\boldsymbol{X})\mathbf{b} = \mathbf{b}'\mathbb{V}(\boldsymbol{X})\mathbf{a}$.

Definition 1.1 *Let $\boldsymbol{X} = (X_1, \dots, X_k)'$ a $k \times 1$ random vector with finite second moments.* $\det[\mathbb{V}(\boldsymbol{X})]$ *is called the* generalized variance *of $\boldsymbol{X}$.*

Theorem 1.2 *Let $\boldsymbol{X} = (X_1, \dots, X_k)'$ be a random vector with covariance matrix $\mathbb{V}(\boldsymbol{X}) = \boldsymbol{\Sigma}$. Then the following properties hold*:

(*a*) $\boldsymbol{\Sigma}' = \boldsymbol{\Sigma}$;

(*b*) *$\boldsymbol{\Sigma}$ is a positive semidefinite matrix*;

(*c*) *$\boldsymbol{\Sigma}$ is positive definite $\Leftrightarrow$ $\boldsymbol{\Sigma}$ is nonsingular* ;

(*d*) $0 \leq \det(\boldsymbol{\Sigma}) \leq \sigma_1^2 \sigma_2^2 \cdots \sigma_k^2$ *where* $\sigma_i^2 = \mathbb{V}(X_i)\,, \; i = 1, \ldots, k$;

(*e*) $\det(\boldsymbol{\Sigma}) = 0 \Leftrightarrow$ *there is at least one linear relation between the random variables* $X_1, \ldots, X_k$, *i.e., we can find constants* $a_1, \ldots, a_k$, b *not all equal to zero such that* $a_1X_1 + \cdots + a_kX_k = b$ *with probability 1;*

(*f*) $\operatorname{rank}(\boldsymbol{\Sigma}) = r < k \Leftrightarrow \boldsymbol{X}$ *can be expressed in the form*

$$\boldsymbol{X} = \boldsymbol{B}\boldsymbol{Y} + \mathbf{c} \tag{1.11}$$

where $\boldsymbol{Y}$ *is a random vector of dimension r whose covariance matrix is* $\boldsymbol{I}_r$, $\boldsymbol{B}$ *is a* $k \times r$ *matrix of rank r, and* $\mathbf{c}$ *is a* $k \times 1$ *constant vector.*

Definition 1.2 *Let* $\boldsymbol{X}_1$ *and* $\boldsymbol{X}_2$ *two random vectors of dimensions* $k_1 \times 1$ *and* $k_2 \times 1$ *respectively with finite second moments. The covariance matrix between* $\boldsymbol{X}_1$ *and* $\boldsymbol{X}_2$ *is defined by:*

$$\mathsf{C}(\boldsymbol{X}_1, \boldsymbol{X}_2) = \mathbb{E}\left\{[\boldsymbol{X}_1 - \mathbb{E}(\boldsymbol{X}_1)]\,[\boldsymbol{X}_2 - \mathbb{E}(\boldsymbol{X}_2)]'\right\} . \tag{1.12}$$

If $k_1 = k_2$, $\det[\mathsf{C}(\boldsymbol{X}_1, \boldsymbol{X}_2)]$ *is called the* generalized covariance *between* $\boldsymbol{X}_1$ *and* $\boldsymbol{X}_2$.

The following proposition summarizes some basic properties of $\mathsf{C}(\boldsymbol{X}_1, \boldsymbol{X}_2)$.

Proposition 1.3 *Let* $\boldsymbol{X}_1$ *and* $\boldsymbol{X}_2$ *two random vectors of dimensions* $k_1 \times 1$ *and* $k_2 \times 1$ *respectively. Then, provided the moments considered are finite we have the following properties:*

(*a*) $\mathsf{C}(\boldsymbol{X}_1, \boldsymbol{X}_2) = \mathbb{E}[\boldsymbol{X}_1\boldsymbol{X}_2'] - \mathbb{E}(\boldsymbol{X}_1)\,\mathbb{E}(\boldsymbol{X}_2)'$;

(*b*) $\mathsf{C}(\boldsymbol{X}_1, \boldsymbol{X}_2) = \mathsf{C}(\boldsymbol{X}_2, \boldsymbol{X}_1)'$;

(*c*) $\mathsf{C}(\boldsymbol{X}_1, \boldsymbol{X}_1) = \mathbb{V}(\boldsymbol{X}_1)\,, \quad \mathsf{C}(\boldsymbol{X}_2, \boldsymbol{X}_2) = \mathbb{V}(\boldsymbol{X}_2)$;

(*d*) *if* $\mathbf{a}$ *and* $\mathbf{b}$ *are fixed vectors of dimensions* $k_1 \times 1$ *and* $k_2 \times 1$ *respectively, then*

$$\mathsf{C}(\boldsymbol{X}_1 + \mathbf{a}, \boldsymbol{X}_2 + \mathbf{b}) = \mathsf{C}(\boldsymbol{X}_1, \boldsymbol{X}_2) \; ; \tag{1.13}$$

(*e*) *if* α *and* β *are two scalar constants,then*

$$\mathsf{C}(\alpha\boldsymbol{X}_1, \beta\boldsymbol{X}_2) = \alpha\beta\,\mathsf{C}(\boldsymbol{X}_1, \boldsymbol{X}_2) \; ; \tag{1.14}$$

(*f*) *if* $\mathbf{a}$ *and* $\mathbf{b}$ *are fixed* $k_1 \times 1$ *and* $k_2 \times 1$ *vectors, then*

$$\mathsf{C}\left(\mathbf{a}'\boldsymbol{X}_1, \mathbf{b}'\boldsymbol{X}_2\right) = \mathbf{a}'\,\mathsf{C}(\boldsymbol{X}_1, \boldsymbol{X}_2)\,\mathbf{b} \; ; \tag{1.15}$$

(*g*) *if* $\boldsymbol{A}$ *and B are fixed matrices matrices with dimensions* $g_1 \times k_1$ *and* $g_2 \times k_2$ *respectively, then*

$$\mathsf{C}(\boldsymbol{A}\boldsymbol{X}_1, \boldsymbol{B}\boldsymbol{X}_2) = \boldsymbol{A}\,\mathsf{C}(\boldsymbol{X}_1, \boldsymbol{X}_2)\,\mathbf{B}' \; ; \tag{1.16}$$

(h) *if* $k_1 = k_2$ *and* $\boldsymbol{X}_3$ *is a* $k \times 1$ *random vector, then*

$$\mathsf{C}(\boldsymbol{X}_1 + \boldsymbol{X}_2, \boldsymbol{X}_3) = \mathsf{C}(\boldsymbol{X}_1, \boldsymbol{X}_3) + \mathsf{C}(\boldsymbol{X}_2, \boldsymbol{X}_3) \; ; \tag{1.17}$$

(i) *if* $k_1 = k_2$, *then*

$$\mathbb{V}(\boldsymbol{X}_1 + \boldsymbol{X}_2) = \mathbb{V}(\boldsymbol{X}_1) + \mathbb{V}(\boldsymbol{X}_2) + \mathsf{C}(\boldsymbol{X}_1, \boldsymbol{X}_2) + \mathsf{C}(\boldsymbol{X}_2, \boldsymbol{X}_1) \; , \tag{1.18}$$

$$\mathbb{V}(\boldsymbol{X}_1 - \boldsymbol{X}_2) = \mathbb{V}(\boldsymbol{X}_1) + \mathbb{V}(\boldsymbol{X}_2) - \mathsf{C}(\boldsymbol{X}_1, \boldsymbol{X}_2) - \mathsf{C}(\boldsymbol{X}_2, \boldsymbol{X}_1) \; . \tag{1.19}$$

Proposition 1.4 *Let*

$$\boldsymbol{X} = \begin{pmatrix} \boldsymbol{X}_1 \\ \boldsymbol{X}_2 \end{pmatrix} \tag{1.20}$$

be a $k \times 1$ *random vector with finite second moments, where* $\boldsymbol{X}_1$ *and* $\boldsymbol{X}_2$ *two random vectors of dimensions* $k_1 \times 1$ *and* $k_2 \times 1$ *respectively, with*

$$\boldsymbol{\Sigma} := \mathbb{V}(\boldsymbol{X}) = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix}, \tag{1.21}$$

$$\boldsymbol{\Sigma}_{11} := \mathbb{V}(\boldsymbol{X}_1) \, , \quad \boldsymbol{\Sigma}_{22} := \mathbb{V}(\boldsymbol{X}_2) \, , \quad \boldsymbol{\Sigma}_{12} := \mathsf{C}(\boldsymbol{X}_1, \boldsymbol{X}_2) \, . \tag{1.22}$$

Then, the following conditions are equivalent:

(a) $\boldsymbol{\Sigma}$ *is nonsingular* ;

(b) $\boldsymbol{\Sigma}_{11}$ *and* $\boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}$ *are nonsingular* ;

(c) $\det(\boldsymbol{\Sigma}_{11}) > 0$ *and* $\det(\boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}) > 0$;

(d) $\det(\boldsymbol{\Sigma}) = \det(\boldsymbol{\Sigma}_{11})\det(\boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}) > 0$.

Further, if $\boldsymbol{\Sigma}_{11}$ *is nonsingular, then*

$$\det(\boldsymbol{\Sigma}) = \det(\boldsymbol{\Sigma}_{11})\det(\boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}) \, . \tag{1.23}$$

2. Multinormal distribution

Consider two random vectors $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ with dimensions $k_1 \times 1$ and $k_2 \times 1$ respectively. If $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ are independent, then

$$\mathsf{C}(\boldsymbol{X}_1, \boldsymbol{X}_2) \equiv \mathbb{E}[(\boldsymbol{X}_1 - \mu_{X_1})(\boldsymbol{X}_2 - \mu_{X_2})'] = 0 \tag{2.1}$$

The reverse implication is not true in general, except in special cases. One such case is the one where the random vector $\boldsymbol{X} = (\boldsymbol{X}_1^{'}, \boldsymbol{X}_2^{'})'$ follows a multinormal distribution.

Definition 2.1 *We say that the* $k \times 1$ *random vector* $\boldsymbol{X}$ *follows a multinormal distribution with mean* μ *and covariance matrix* $\boldsymbol{\Sigma}$*, denoted* $\boldsymbol{X} \sim \mathrm{N}_k[\mu, \boldsymbol{\Sigma}]$*, if the characteristic function of* $\boldsymbol{x}$ *has the form:*

$$\mathbb{E}\left[e^{i\mathbf{t}'\boldsymbol{X}}\right] = \exp[i\mu'\mathbf{t} - \mathbf{t}'\boldsymbol{\Sigma}\mathbf{t}/2], \quad \mathbf{t} \in \mathbb{R}^k,\ i = \sqrt{-1}\,. \tag{2.2}$$

When $|\boldsymbol{\Sigma}| \neq 0$, the vector $\boldsymbol{X}$ has a density function of the form:

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{k/2}|\boldsymbol{\Sigma}|^{\frac{1}{2}}} \exp\left[-\frac{1}{2}(\mathbf{x}-\mu)'\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\mu)\right] \tag{2.3}$$

If $k = 1$, then $\boldsymbol{\Sigma} = \sigma^2$ and

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2}(x-\mu)\frac{1}{\sigma^2}(x-\mu)\right] = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right]. \tag{2.4}$$

Some important properties of the multinormal distribution are summarized in the following theorem.

Theorem 2.1 *If* $\boldsymbol{X} \sim \mathrm{N}_k[\mu, \boldsymbol{\Sigma}]$*, then*

(a) $\boldsymbol{X} + \mathbf{c} \sim \mathrm{N}_k[\mu + \mathbf{c}, \boldsymbol{\Sigma}]$ *, for any fixed* $k \times 1$ *vector* $\mathbf{c}$;

(b) $\mathbf{a}'\boldsymbol{X} \sim \mathrm{N}_1[\mathbf{a}'\mu, \mathbf{a}'\boldsymbol{\Sigma}\mathbf{a}]$ *, for any fixed* $k \times 1$ *vector* $\mathbf{a}$;

(c) $\boldsymbol{A}\boldsymbol{X} \sim \mathrm{N}_g[\boldsymbol{A}\mu, \boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{A}']$ *, for any fixed* $g \times k$ *matrix* $\boldsymbol{A}$;

(d) *if*

$$\boldsymbol{X} = \begin{pmatrix}\boldsymbol{X}_1\\ \boldsymbol{X}_2\end{pmatrix} \sim \mathrm{N}_k\left[\begin{pmatrix}\mu_1\\ \mu_2\end{pmatrix}, \begin{pmatrix}\boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12}\\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22}\end{pmatrix}\right], \tag{2.5}$$

where $\boldsymbol{X}_1$ *and* $\boldsymbol{X}_2$ *are vectors of dimensions* $k_1 \times 1$ *and* $k_2 \times 1$,

$$\mu_1 = \mathbb{E}(\boldsymbol{X}_1),\ \mu_2 = \mathbb{E}(\boldsymbol{X}_2),\ \boldsymbol{\Sigma}_{11} = \mathsf{C}(\boldsymbol{X}_1, \boldsymbol{X}_1),\ \boldsymbol{\Sigma}_{22} = \mathsf{C}(\boldsymbol{X}_2, \boldsymbol{X}_2), \tag{2.6}$$

$$\boldsymbol{\Sigma}_{12} = \mathsf{C}(\boldsymbol{X}_1, \boldsymbol{X}_2) = \boldsymbol{\Sigma}_{21}', \tag{2.7}$$

then

(i) $\boldsymbol{X}_1 \sim \mathrm{N}_{k_1}[\mu_1, \boldsymbol{\Sigma}_{11}],\ \boldsymbol{X}_2 \sim \mathrm{N}_{k_2}[\mu_2, \boldsymbol{\Sigma}_{22}]$;

(ii) $\boldsymbol{X}_1$ *and* $\boldsymbol{X}_2$ *are independent* $\Leftrightarrow \boldsymbol{\Sigma}_{12} = 0$;

(iii) *the conditional distribution of* $\boldsymbol{X}_2$ *given* $\boldsymbol{X}_1$ *is normal with mean and et variance*

$$\mathbb{E}[\boldsymbol{X}_2 \,|\, \boldsymbol{X}_1] = \mu_2 + \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}(\boldsymbol{X}_1 - \mu_1), \tag{2.8}$$

$$\mathbb{V}[\boldsymbol{X}_2 \,|\, \boldsymbol{X}_1] = \boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}, \tag{2.9}$$

i.e.

$$\boldsymbol{X}_2 \,|\, \boldsymbol{X}_1 \sim \mathrm{N}_{k_2}\left[\mu_2 + \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}(\boldsymbol{X}_1 - \mu_1), \boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}\right]. \tag{2.10}$$

Theorem 2.2 *If* $\boldsymbol{X} \sim \mathrm{N}_k[\mu,\ \boldsymbol{\Sigma}]$ *with* $|\boldsymbol{\Sigma}| \neq 0$, *then*

$$(\boldsymbol{X}-\mu)'\,\boldsymbol{\Sigma}^{-1}(\boldsymbol{X}-\mu) \sim \chi^2(k)\ . \tag{2.11}$$

PROOF Since $\boldsymbol{\Sigma}$ is a positive definite matrix ($|\boldsymbol{\Sigma}| \neq 0$), there exists a nonsingular matrix $\boldsymbol{P}$ such that

$$\boldsymbol{P}\boldsymbol{\Sigma}\boldsymbol{P}' = \boldsymbol{I}_k \tag{2.12}$$

hence

$$\boldsymbol{\Sigma} = \boldsymbol{P}^{-1}\left(\boldsymbol{P}'\right)^{-1} = \left(\boldsymbol{P}'\boldsymbol{P}\right)^{-1}\ , \tag{2.13}$$

$$\boldsymbol{\Sigma}^{-1} = \boldsymbol{P}'\boldsymbol{P}\ . \tag{2.14}$$

Consequently,

$$\begin{aligned}(\boldsymbol{X}-\mu)'\,\boldsymbol{\Sigma}^{-1}(\boldsymbol{X}-\mu) &= (\boldsymbol{X}-\mu)'\,\boldsymbol{P}'\boldsymbol{P}(\boldsymbol{X}-\mu)\\ &= [\boldsymbol{P}(\boldsymbol{X}-\mu)]'[\boldsymbol{P}(\boldsymbol{X}-\mu)]\\ &= \mathbf{v}'\mathbf{v} = \sum_{i=1}^{k} v_i^2 \end{aligned} \tag{2.15}$$

where

$$\mathbf{v} \equiv \boldsymbol{P}[\boldsymbol{X}-\mu] = (v_1, v_2, \ldots, v_k)'\ . \tag{2.16}$$

Since $\boldsymbol{X} \sim \mathrm{N}[\mu,\ \boldsymbol{\Sigma}]$, we have $\boldsymbol{X}-\mu \sim \mathrm{N}[\mathbf{0},\ \boldsymbol{\Sigma}]$, hence

$$\boldsymbol{P}[\boldsymbol{X}-\mu] \sim \mathrm{N}\left[\mathbf{0}, P\Sigma P'\right]\ , \tag{2.17}$$

and

$$\mathbf{v} = \boldsymbol{P}[\boldsymbol{X}-\mu] \sim \mathrm{N}[\mathbf{0}, I_k]\ . \tag{2.18}$$

Thus $v_1, \ldots, v_k$ are i.i.d. $\mathrm{N}[0,\ 1]$ and

$$(\boldsymbol{X}-\mu)'\,\boldsymbol{\Sigma}^{-1}(\boldsymbol{X}-\mu) = \sum_{i=1}^{k} v_i^2 \sim \chi^2(k)\ . \tag{2.19}$$

□

3. Multiple linear regression

Consider the problem of finding a $k \times 1$ vector $\boldsymbol{b}$ such that

$$S(\boldsymbol{b}) := \mathbb{E}[(Y-\boldsymbol{X}'\boldsymbol{b})^2] \tag{3.1}$$

is minimal.

Let β be any $k \times 1$ vector such that

$$\mathbb{E}[\boldsymbol{X}(Y - \boldsymbol{X}'\beta)] = \boldsymbol{0}. \tag{3.2}$$

Then, for any $\boldsymbol{b} \in \mathbb{R}^k$,

$$\begin{aligned} S(\boldsymbol{b}) &= \mathbb{E}\{[(Y - \boldsymbol{X}'\beta) + (\boldsymbol{X}'\beta - \boldsymbol{X}'\boldsymbol{b})]^2\} \\ &= \mathbb{E}[(Y - \boldsymbol{X}'\beta)^2] + \mathbb{E}[(\boldsymbol{X}'\beta - \boldsymbol{X}'\boldsymbol{b})^2] + 2\mathbb{E}[(\boldsymbol{X}'\beta - \boldsymbol{X}'\boldsymbol{b})(Y - \boldsymbol{X}'\beta)] \\ &= \mathbb{E}[(Y - \boldsymbol{X}'\beta)^2] + \mathbb{E}[(\boldsymbol{X}'\beta - \boldsymbol{X}'\boldsymbol{b})^2] + 2\mathbb{E}[(\beta - \boldsymbol{b})'\boldsymbol{X}(Y - \boldsymbol{X}'\beta)] \\ &= \mathbb{E}[(Y - \boldsymbol{X}'\beta)^2] + \mathbb{E}[(\boldsymbol{X}'\beta - \boldsymbol{X}'\boldsymbol{b})^2] + (\beta - \boldsymbol{b})'\mathbb{E}[\boldsymbol{X}(Y - \boldsymbol{X}'\beta)] \\ &= S(\beta) + \mathbb{E}[(\boldsymbol{X}'\beta - \boldsymbol{X}'\boldsymbol{b})^2] \geq S(\beta) \end{aligned} \tag{3.3}$$

so that β minimizes $S(\boldsymbol{b})$. If β and β^* are two such solutions, *i.e.*

$$S(\beta) = S(\beta^*), \tag{3.4}$$

we must have

$$S(\beta^*) = S(\beta) + \mathbb{E}[(\boldsymbol{X}'\beta - \boldsymbol{X}'\beta^*)^2] \tag{3.5}$$

hence

$$\mathbb{E}[(\boldsymbol{X}'\beta - \boldsymbol{X}'\beta^*)^2] = S(\beta^*) - S(\beta) = 0 \tag{3.6}$$

and

$$\boldsymbol{X}'\beta = \boldsymbol{X}'\beta^* \quad \text{a.s.} \tag{3.7}$$

Even if β is not unique, $S(\beta)$ and $\boldsymbol{X}'\beta$ are unique. Consequently, there is a unique approximation (or "fitted value")

$$\boldsymbol{P}(Y; \boldsymbol{X}) := \boldsymbol{X}'\beta \tag{3.8}$$

and a unique residual

$$U(Y; \boldsymbol{X}) := Y - P(Y; \boldsymbol{X}) = Y - \boldsymbol{X}'\beta \tag{3.9}$$

where β is any vector such that

$$\mathbb{E}[\boldsymbol{X}(Y - \boldsymbol{X}'\beta)] = \boldsymbol{0}. \tag{3.10}$$

This yields the following unique decomposition:

$$Y = P(Y; \boldsymbol{X}) + U(Y; \boldsymbol{X}) = \boldsymbol{X}'\beta + U(Y; \boldsymbol{X}) \tag{3.11}$$

where

$$\mathbb{E}[\boldsymbol{X}\,U(Y; \boldsymbol{X})] = \boldsymbol{0}. \tag{3.12}$$

This entails:

$$\begin{aligned} \mathbb{E}[P(Y; \boldsymbol{X})U(Y; \boldsymbol{X})] &= \mathbb{E}[\beta'\boldsymbol{X}\,U(Y; \boldsymbol{X})] \\ &= \beta'\mathbb{E}[\boldsymbol{X}\,U(Y; \boldsymbol{X})] = 0, \end{aligned} \tag{3.13}$$

$$\mathbb{E}(Y^2) = \mathbb{E}[P(Y;\boldsymbol{X})^2] + \mathbb{E}[U(Y;\boldsymbol{X})^2]\,. \tag{3.14}$$

We call the ratio

$$R_0^2(Y;\boldsymbol{X}) := \frac{\mathbb{E}[P(Y;\boldsymbol{X})^2]}{\mathbb{E}(Y^2)} \tag{3.15}$$

the *uncentered R-square* of Y on $\boldsymbol{X}$. Clearly,

$$\frac{\mathbb{E}[U(Y;\boldsymbol{X})^2]}{\mathbb{E}(Y^2)} = 1 - R_0^2(Y;\boldsymbol{X})\,. \tag{3.16}$$

β is any solution of the equation

$$\mathbb{E}(\boldsymbol{X}\boldsymbol{X}')\beta = \mathbb{E}(\boldsymbol{X}Y) \tag{3.17}$$

where

$$\mathbb{E}(\boldsymbol{X}\boldsymbol{X}') = [\bar{\sigma}_{ij}] := \bar{\boldsymbol{\Sigma}}(\boldsymbol{X})\,, \quad \mathbb{E}(\boldsymbol{X}Y) := \bar{\mathsf{C}}(\boldsymbol{X}, Y)\,, \tag{3.18}$$

$$\bar{\sigma}_{ij} = \mathbb{E}[X_i X_j] = \bar{\rho}_{ij}\,\bar{\sigma}_i\,\bar{\sigma}_j, \quad \bar{\sigma}_{ii} = [\mathbb{E}(X_i^2)]^{1/2} = \bar{\sigma}_i^2 = \bar{\sigma}(X_i)^2, \tag{3.19}$$

$$\bar{\rho}_{ij} = \frac{\mathbb{E}[X_i X_j]}{\bar{\sigma}(X_i)\bar{\sigma}(X_j)} = \frac{\bar{\sigma}_{ij}}{\bar{\sigma}_i\,\bar{\sigma}_i}\,, \tag{3.20}$$

for $i,\ j = 1, \ldots, k$. $\bar{\sigma}_{ij}$ is called the *uncentered covariance* between X_i and X_j, and $\bar{\rho}_{ij}$ *uncentered correlation* between X_i and X_j. Equation (3.17) is called the *uncentered normal equation* for the linear regression of Y on $\boldsymbol{X}$. For any β that satisfies (3.17),

$$\mathbb{E}[(\boldsymbol{X}'\beta)^2] = \beta'\mathbb{E}(\boldsymbol{X}\boldsymbol{X}')\beta = \beta'\mathbb{E}(\boldsymbol{X}Y)\,, \tag{3.21}$$

$$\begin{aligned}
\mathbb{E}[U(Y;\boldsymbol{X})^2] &= \mathbb{E}[U(Y;\boldsymbol{X})(Y - \boldsymbol{X}'\beta)] = \mathbb{E}[U(Y;\boldsymbol{X})(Y - \boldsymbol{X}'\beta)] \\
&= \mathbb{E}[U(Y;\boldsymbol{X})Y] = \mathbb{E}[(Y - \boldsymbol{X}'\beta)Y] = \mathbb{E}[Y^2] - \mathbb{E}[(\boldsymbol{X}'\beta)Y] \\
&= \mathbb{E}[Y^2] - \beta'\mathbb{E}(\boldsymbol{X}Y) \\
&= \mathbb{E}[Y^2] - \beta'\mathbb{E}(\boldsymbol{X}\boldsymbol{X}')\beta \\
&= \mathbb{E}[Y^2] - \mathbb{E}[(\boldsymbol{X}'\beta)^2]\,.
\end{aligned} \tag{3.22}$$

Due to the unicity of $\boldsymbol{X}'\beta$ and $U(Y;\boldsymbol{X})$, $\mathbb{E}[(\boldsymbol{X}'\beta)^2]$ and $\mathbb{E}[U(Y;\boldsymbol{X})^2]$ are also uniquely defined irrespective of the solution β of the normal equation.

If the matrix $\mathbb{E}(\boldsymbol{X}\boldsymbol{X}')$ is invertible, then β is unique with

$$\beta = [\mathbb{E}(\boldsymbol{X}\boldsymbol{X}')]^{-1}\mathbb{E}(\boldsymbol{X}Y)\,. \tag{3.23}$$

In this case,

$$\mathbb{E}[(\boldsymbol{X}'\beta)^2] = \mathbb{E}(\boldsymbol{X}Y)'[\mathbb{E}(\boldsymbol{X}\boldsymbol{X}')]^{-1}\mathbb{E}(\boldsymbol{X}Y)\,, \tag{3.24}$$

$$\mathbb{E}[U(Y;\boldsymbol{X})^2] = \mathbb{E}[Y^2] - \mathbb{E}(\boldsymbol{X}Y)'[\mathbb{E}(\boldsymbol{X}\boldsymbol{X}')]^{-1}\mathbb{E}(\boldsymbol{X}Y)\,. \tag{3.25}$$

Theorem 3.1 *Let $X_1, \ldots, X_k$, Y be random variables with finite second moments, let $\boldsymbol{X}$ be defined as in* (1.6)*, and set*

$$\operatorname{argmin} S(b) := \left\{\beta \in \mathbb{R}^k : S(\beta) = \min_{\boldsymbol{b}\in\mathbb{R}^k} S(\boldsymbol{b})\right\}. \tag{3.26}$$

Then, there exists a vector $\beta \in \mathbb{R}^k$ such that

$$S(\beta) = \min_{\boldsymbol{b}\in\mathbb{R}^k} S(\boldsymbol{b}). \tag{3.27}$$

Further,

$$\{\beta \in \operatorname{argmin} S(b)\} \Leftrightarrow \{\mathbb{E}[\boldsymbol{X}(Y - \boldsymbol{X}'\beta)] = \boldsymbol{0}\}, \tag{3.28}$$

$$\{\beta, \beta^* \in \operatorname{argmin} S(b)\} \Rightarrow \{\boldsymbol{X}'\beta = \boldsymbol{X}'\beta^*\}. \tag{3.29}$$

Proposition 3.2 IDENTIFICATION OF LINEAR REGRESSION BY MOMENT EQUATIONS. *Let $X_1, \ldots, X_k$, Y be random variables with finite second moments, let $\boldsymbol{X}$ be defined as in* (1.6)*, $\boldsymbol{Z} := (\boldsymbol{X}', Y)'$, and $\boldsymbol{a} = (a_1, a_2, \ldots, a_{k+1})'$ be a $(k+1)\times 1$ fixed vector. If*

$$\mathbb{E}[\boldsymbol{X}(\boldsymbol{a}'\boldsymbol{Z})] = \mathbf{o} \text{ and } \mathbb{E}[(\boldsymbol{a}'\boldsymbol{Z})Y] = 1, \tag{3.30}$$

then $a_{k+1} \neq 0$, and for

$$\beta = -\frac{1}{a_{k+1}}(a_1, a_2, \ldots, a_k)', \tag{3.31}$$

we have:

$$\mathbb{E}[\boldsymbol{X}(Y - \boldsymbol{X}'\beta)] = \boldsymbol{0}, \tag{3.32}$$

$$\mathbb{E}[(Y - \boldsymbol{X}'\beta)^2] = \mathbb{E}[Y^2] - \beta'\mathbb{E}(\boldsymbol{X}\boldsymbol{X}')\beta > 0, \tag{3.33}$$

$$a_{k+1} = \frac{1}{\mathbb{E}[(Y - \boldsymbol{X}'\beta)^2]}. \tag{3.34}$$

If $\mathbb{E}(\boldsymbol{X}\boldsymbol{X}')$ is invertible, then

$$(a_1, a_2, \ldots, a_k)' = -a_{k+1}[\mathbb{E}(\boldsymbol{X}\boldsymbol{X}')]^{-1}\mathbb{E}(\boldsymbol{X}Y), \tag{3.35}$$

$$a_{k+1} = \frac{1}{\mathbb{E}[Y^2] - \mathbb{E}(\boldsymbol{X}Y)'[\mathbb{E}(\boldsymbol{X}\boldsymbol{X}')]^{-1}\mathbb{E}(\boldsymbol{X}Y)}. \tag{3.36}$$

It follows from Proposition 3.2, that $1/a_{k+1}$ is the residual variance from the linear regression of Y on $\boldsymbol{X}$, while each coefficient

$$\beta_i = -\frac{a_i}{a_{k+1}} \tag{3.37}$$

is the coefficient of X_i in this regression ($1 \le i \le k$), in the sense that $\beta = (\beta_1, \ldots, \beta_k)'$ provides a solution of the normal equation (3.17). This holds irrespective of the rank of $\mathbb{E}(\boldsymbol{X}\boldsymbol{X}')$. When $\mathbb{E}(\boldsymbol{X}\boldsymbol{X}')$ is singular, the normal equation has other solutions. a_{k+1} is unique in all cases. There exists a vector $\boldsymbol{a}$ such that (3.30) holds as soon as

$$\text{rank}[\mathbb{E}(\boldsymbol{X}\boldsymbol{X}')] \geq 1\,. \tag{3.38}$$

4. Sources and additional references

Good overviews of various notions associated with covariances, correlations and regression may be found in Hannan (1970, Chapter 1), Theil (1971, Chapter 4), Kendall and Stuart (1979, Chapters 26-28), Rao (1973, Section 4g), Drouet Mari and Kotz (2001), and Anderson (2003, Chapter 1). See also Lehmann (1966).

References

ANDERSON, T. W. (2003): *An Introduction to Multivariate Statistical Analysis*. John Wiley & Sons, New York, third edn.

DROUET MARI, D., AND S. KOTZ (2001): *Correlation and Dependence*. World Scientific, River Edge, New Jersey.

HANNAN, E. J. (1970): *Multiple Time Series*. John Wiley & Sons, New York.

KENDALL, M., AND A. STUART (1979): *The Advanced Theory of Statistics. Volume 2: Inference and Relationship*. Macmillan, New York, fourth edn.

LEHMANN, E. L. (1966): "Some Concepts of Dependence," *The Annals of Mathematical Statistics*, 37, 1137–1153.

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

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