Covariance, correlation and linear regression between random variables *

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 Universitad 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

List of Definitions, Assumptions, Propositions and Theorems

1. Random variables

In general, economic theory specifies exact relations between economic variables. Even a superficial examination of economic data indicates it is not (almost never) possible to find such relationships in actual data. Instead, we have relations of the form:

$$C_t = \alpha + \beta Y_t + \varepsilon_t$$

where ε_t can be interpreted as a "random variable".

Definition 1.1 *A random variable* (r.v.) *X is a variable whose behavior can be described by a "probability law". If X takes its values in the real numbers, the probability law of X can be described by a "distribution function":*

$$F_X(x) = \mathbb{P}[X \leq x]$$

If X is continuous, there is a "density function" $f_X(x)$ such that

$$F_X(x) = \int_{-\infty}^{x} f_X(x)\, dx\,.$$

The mean and variance of X are given by:

$$\mu_X = \mathbb{E}(X) = \int_{-\infty}^{+\infty} x\, dF_X(x) \qquad \text{(general case)}$$

$$= \int_{-\infty}^{+\infty} x\, f_X(x)\; dx \qquad \text{(continuous case)}$$

$$\mathbb{V}(X) = \sigma_X^2 = \mathbb{E}\left[(X-\mu_X)^2\right] = \int_{-\infty}^{+\infty} (x-\mu_X)^2\, dF_X(x) \qquad \text{(general case)}$$

$$= \int_{-\infty}^{+\infty} (x-\mu_X)^2\, F_X(x)\, dx \qquad \text{(continuous case)}$$

$$= \mathbb{E}\left(X^2\right) - [\mathbb{E}(X)]^2$$

It is easy to characterize relations between two non-random variables x and y :

$$g(x, y) = 0$$

or (in certain cases)

$$y = f(x)\,.$$

How does one characterize the links or relations between random variables? The behavior of a pair $(X,Y)'$ is described by a joint distribution function:

$$\begin{aligned} F(x,y) &= \mathbb{P}[X \le x,\, Y \le y] \\ &= \int_{-\infty}^{y}\int_{-\infty}^{x} f(x,y)\,dx\,dy \qquad \text{(continuous case.)} \end{aligned}$$

We call $f(x,y)$ the joint density function of $(X,Y)'$. More generally, if we consider k r.v.'s $X_1, X_2, \dots, X_k$, their behavior can be described through a k-dimensional distribution function:

$$\begin{aligned} F(x_1,x_2,\dots,x_k) &= \mathbb{P}[X_1 \le x_1, X_2 \le x_2, \dots, X_k \le x_k] \\ &= \int_{-\infty}^{x_k}\cdots\int_{-\infty}^{x_2}\int_{-\infty}^{x_1} f(x_1,x_2,\dots,x_k)\,dx_1 dx_2\cdots dx_k \qquad \text{(continuous case)} \end{aligned}$$

where $f(x_1,x_2,\dots,x_k)$ is the joint density function of $X_1,X_2,\dots,X_k$.

2. Covariances and correlations

We often wish to have a simple measure of association between two random variables X and Y. The notions of "covariance" and "correlation" provide such measures of association. Let X and Y be two *r.v.*'s with means

$$\mu_X := \mathbb{E}(X), \quad \mu_Y := \mathbb{E}(Y), \tag{2.1}$$

and finite second moments

$$\bar{\sigma}_X^2 := \mathbb{E}(X^2), \quad \bar{\sigma}_Y^2 := \mathbb{E}(Y^2). \tag{2.2}$$

Then, X and Y have finite variances:

$$\sigma_X^2 := \mathbb{V}(X) := \mathbb{E}[(X-\mu_X)^2] = \mathbb{E}(X^2) - \mu_X^2 = \bar{\sigma}_X^2 - \mu_X^2, \tag{2.3}$$

$$\sigma_Y^2 := \mathbb{V}(Y) := \mathbb{E}[(Y-\mu_Y)^2] = \mathbb{E}(Y^2) - \mu_Y^2 = \bar{\sigma}_Y^2 - \mu_Y^2. \tag{2.4}$$

We also denote:

$$\bar{\sigma}(X) := \bar{\sigma}_X = [\mathbb{E}(X^2)]^{1/2}, \quad \bar{\sigma}(Y) := \bar{\sigma}_Y = [\mathbb{E}(Y^2)]^{1/2}, \tag{2.5}$$

$$\sigma(X) := \sigma_X, \quad \sigma(Y) := \sigma_Y, \tag{2.6}$$

where $\bar{\sigma}(X) \ge 0$, $\bar{\sigma}(Y) \ge 0$, $\sigma(X) \ge 0$ and $\sigma(Y) \ge 0$, so that

$$\sigma(X)^2 = \mathbb{V}(X), \quad \sigma(Y)^2 = \mathbb{V}(Y). \tag{2.7}$$

Below *a.s.* means "almost surely" (with probability 1). In particular, we have:

$$\mathbb{E}(X^2) = 0 \Leftrightarrow [X = 0 \quad \text{a.s.}] \Leftrightarrow \mathbb{P}[X = 0] = 1, \tag{2.8}$$

$$\mathbb{V}(X) = 0 \Leftrightarrow [X = \mathbb{E}(X) \quad \text{a.s.}] \Leftrightarrow \mathbb{P}[X = \mathbb{E}(X)] = 1. \tag{2.9}$$

Definition 2.1 COVARIANCE. *The* covariance *between X and Y is defined by*

$$\mathsf{C}(X, Y) := \sigma_{XY} := \mathbb{E}[(X - \mu_X)(Y - \mu_Y)] \,. \tag{2.10}$$

When $\mathsf{C}(X, Y) = 0$, we say that X and Y are orthogonal.

Definition 2.2 CORRELATION. *The* correlation *between X and Y is defined by*

$$\rho(X, Y) := \rho_{XY} := \frac{\mathsf{C}(X, Y)}{\sigma(X)\sigma(Y)} \tag{2.11}$$

where we set $\rho(X, Y) := 0$ when $\sigma(X)\sigma(Y) = 0$.

When X or Y is degenerate, we have $\mathsf{C}(X, Y) = \sigma(X)\sigma(Y) = 0$. The convention $\rho(X, Y) := 0$ when $\sigma(X)\sigma(Y) = 0$ is motivated by the fact that $\mathsf{C}(X, Y) = 0$ in this case.

Theorem 2.1 BASIC PROPERTIES OF COVARIANCES AND CORRELATIONS. *Let (X, Y) be a pair of random variables with finite second moments. The covariance and correlation between X and Y satisfy the following properties:*

(*a*) $\mathsf{C}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$;

(*b*) $\mathsf{C}(a_1 + b_1X, a_2 + b_2Y) = b_1b_2\mathsf{C}(X, Y)$ *for any constants a_1, a_2, b_1, b_2*;

(*c*) $\rho(a_1 + b_1X, a_2 + b_2Y) = \rho(X, Y)$ *for any constants a_1, a_2, b_1, b_2 such that $b_1b_2 \neq 0$*;

(*d*) $\mathsf{C}(X, Y) = \mathsf{C}(Y, X)$ *and* $\rho(X, Y) = \rho(Y, X)$;

(*e*) $\mathsf{C}(X,X) = \mathbb{V}(X)$;

(*f*) $\rho(X,X) = 1$ *if* $\mathbb{V}(X) > 0$;

(*g*) $\mathsf{C}(X, Y)^2 \leq \mathbb{V}(X)\mathbb{V}(Y)$; (*Cauchy-Schwarz inequality*)

(*h*) $-1 \leq \rho(X, Y) \leq 1$;

(*i*) *X and Y are independent* $\Rightarrow \mathsf{C}(X,Y)=0 \Rightarrow \rho(X,Y)=0$;

(*j*) *if* $\sigma(X)\sigma(Y)\neq 0$, *then*

$$\begin{aligned}[\rho(X,Y)^2=1] &\Leftrightarrow \left[\exists \text{ two constants } a \text{ and } b \text{ such that } b\neq 0 \text{ and } Y=a+bX \text{ a.s.}\right] \\ &\Leftrightarrow \left[Y=a+bX \text{ a.s. with } b=\beta(X \rightharpoondown Y) \text{ and } a=\mathbb{E}(Y)-b\,\mathbb{E}(X)\right], \end{aligned} \tag{2.12}$$

$$[\rho(X,Y)=1] \Leftrightarrow \left[Y=a+bX \;\text{ a.s. with } b=\beta(X \rightharpoondown Y)>0 \text{ and } a=\mathbb{E}(Y)-b\,\mathbb{E}(X)\right], \tag{2.13}$$

$$[\rho(X,Y)=-1] \Leftrightarrow \left[Y=a+bX \;\text{ a.s. with } b=\beta(X \rightharpoondown Y)<0 \text{ and } a=\mathbb{E}(Y)-b\,\mathbb{E}(X)\right]. \tag{2.14}$$

PROOF (a)

$$\begin{aligned}\mathsf{C}(X,Y) &= \mathbb{E}\left[(X-\mu_X)(Y-\mu_Y)\right] \\ &= \mathbb{E}\left[XY-\mu_X Y - X\mu_Y+\mu_X\mu_Y\right] \\ &= \mathbb{E}(XY)-\mu_X\mathbb{E}(Y)-\mathbb{E}(X)\mu_Y+\mu_X\mu_Y \\ &= \mathbb{E}(XY)-\mu_X\mu_Y-\mu_X\mu_Y+\mu_X\mu_y \\ &= \mathbb{E}(XY)-\mathbb{E}(X)\mathbb{E}(Y)\ . \end{aligned} \tag{2.15}$$

(b), (c), (d), (e) and (f) are immediate.
(g) To get (g), we observe that

$$\begin{aligned}\mathbb{E}\left\{\left[Y-\mu_Y-\lambda(X-\mu_X)\right]^2\right\} &= \mathbb{E}\left\{\left[(Y-\mu_Y)-\lambda(X-\mu_X)\right]^2\right\} \\ &= \mathbb{E}\left\{(Y-\mu_Y)^2-2\lambda(X-\mu_X)(Y-\mu_Y)+\lambda^2(X-\mu_X)^2\right\} \\ &= \sigma_Y^2-2\lambda\sigma_{XY}+\lambda^2\sigma_X^2\geq 0 \end{aligned} \tag{2.16}$$

for any arbitrary constant λ. In other words, the second-order polynomial

$$g(\lambda)=\sigma_Y^2-2\lambda\sigma_{XY}+\lambda^2\sigma_X^2 \tag{2.17}$$

cannot take negative values. This can happen only if the equation

$$\lambda^2\sigma_X^2-2\lambda\sigma_{XY}+\sigma_Y^2=0 \tag{2.18}$$

does not have two distinct real roots, *i.e.* the roots are either complex or identical. The roots of equation (2.18) are:

$$\lambda = \frac{2\sigma_{XY} \pm \sqrt{4\sigma_{XY}^2 - 4\sigma_X^2\sigma_Y^2}}{2\sigma_X^2} = \frac{\sigma_{XY} \pm \sqrt{\sigma_{XY}^2 - \sigma_X^2\sigma_Y^2}}{\sigma_X^2} . \tag{2.19}$$

Distinct real roots are excluded when $\sigma_{XY}^2 - \sigma_X^2\sigma_Y^2 \leq 0$, hence

$$\sigma_{XY}^2 \leq \sigma_X^2\sigma_Y^2 . \tag{2.20}$$

(h)

$$\begin{aligned} \sigma_{XY}^2 \leq \sigma_X^2\sigma_Y^2 \quad &\Rightarrow \quad -\sigma_X\sigma_Y \leq \sigma_{XY} \leq \sigma_X\sigma_Y \\ &\Rightarrow \quad -1 \leq \rho_{XY} \leq 1 . \end{aligned} \tag{2.21}$$

(i) If X and Y are independent, we have:

$$\begin{aligned} \sigma_{XY} \quad &= \quad \mathbb{E}\{(X-\mu_X)(Y-\mu_Y)\} = \mathbb{E}(X-\mu_X)\mathbb{E}(Y-\mu_Y) \\ &= \quad [\mathbb{E}(X)-\mu_X][\mathbb{E}(Y)-\mu_Y] = 0 , \end{aligned} \tag{2.22}$$

$$\rho_{XY} = \sigma_{XY} \big/ \sigma_X\sigma_Y = 0 . \tag{2.23}$$

Note the reverse implication does not hold in general, *i.e.*,

$$\rho_{XY} = 0 \neq> X \text{ and } Y \text{ are independent} . \tag{2.24}$$

(j) (a) Necessity of the condition. If $Y = aX + b$, then

$$\mathbb{E}(Y) = a\mathbb{E}(X) + b = a\mu_X + b \ , \ \ \sigma_Y^2 = a^2\sigma_X^2 \ , \tag{2.25}$$

and

$$\sigma_{XY} = \mathbb{E}[(Y-\mu_Y)(X-\mu_X)] = \mathbb{E}[a(X-\mu_X)(X-\mu_X)] = a\sigma_X^2 . \tag{2.26}$$

Consequently,

$$\rho_{XY}^2 = \frac{a^2\sigma_X^4}{a^2\sigma_X^2\sigma_X^2} = 1 . \tag{2.27}$$

(b) Sufficiency of the condition. If $\rho_{XY}^2 = 1$, then

$$\sigma_{XY}^2 - \sigma_X^2\sigma_Y^2 = 0. \tag{2.28}$$

In this case, the equation

$$\mathbb{E}\left\{[(Y-\mu_Y) - \lambda(X-\mu_X)]^2\right\} = \sigma_Y^2 - 2\lambda\sigma_{XY} + \lambda^2\sigma_X^2 = 0 \tag{2.29}$$

has one and only one root

$$\lambda = \frac{2\sigma_{XY}}{2\sigma_X^2} = \sigma_{XY}/\sigma_X^2\ , \tag{2.30}$$

so that

$$\mathbb{E}\{[\left(Y\sigma_Y^2 - \mu_Y\right) - (\sigma_{XY}/\sigma_X^2)\,(X - \mu_X)]^2\} = 0 \tag{2.31}$$

and

$$\mathbb{P}\left[\,(Y - \mu_Y) - (\sigma_{XY}/\sigma_X^2)\,(X - \mu_X) = 0\right] = \mathbb{P}\left[Y = \left(\mu_Y - (\sigma_{XY}/\sigma_X^2)\mu_X\right) + (\sigma_{XY}/\sigma_X^2)X\right] = 1 \tag{2.32}$$

We can thus write:

$$Y = a + bX \text{ with probability } 1 \tag{2.33}$$

where $b = \sigma_{XY}/\sigma_X^2$ and $a = \mu_Y - (\sigma_{XY}/\sigma_X^2)\mu_X$. This establishes (2.12). (2.13) follows on observing that, for $b = \sigma_{XY}/\sigma_X^2$ and $a = \mu_Y - (\sigma_{XY}/\sigma_X^2)\mu_X$,

$$\begin{aligned}
[\rho\,(X,Y) = 1] \quad &\Leftrightarrow \quad \left\{\rho\,(X,Y)^2 = 1 \text{ and } \rho\,(X,Y) = \frac{\sigma_{XY}}{\sigma_X\sigma_Y} > 0\right\} \\
&\Leftrightarrow \quad \left\{\mathbb{P}[Y = a + bX]\right] = 1 \text{ and } \rho\,(X,Y) = \frac{\sigma_{XY}}{\sigma_X\sigma_Y} > 0\right\} \\
&\Leftrightarrow \quad \left\{\mathbb{P}[Y = a + bX] = 1 \text{ and } \rho\,(X,Y) = \frac{b\sigma_X^2}{\sigma_X\sigma_Y} > 0\right\} \\
&\Leftrightarrow \quad \left\{\mathbb{P}[Y = a + bX] = 1 \text{ and } b > 0\right\}.
\end{aligned} \tag{2.34}$$

The proof for (2.14) is similar. □

A basic problem in this context consists in considering the case where

$$Y = a + bX \quad a.s. \tag{2.35}$$

and find whether a and b can be determined (or "identified") from the first and second moments of X and Y. The following theorem shows that a and b are uniquely determined if only if $\mathbb{V}(X) > 0$.

Theorem 2.2 IDENTIFICATION OF LINEAR TRANSFORMATION OF A RANDOM VARIABLE. *Suppose X and Y satisfy the linear equation* (2.35). *If* $\mathbb{V}(X) > 0$, *then*

$$\{\mathbb{P}[Y = a_1 + b_1 X] = 1\} \Rightarrow [a_1 = a \text{ and } b_1 = b]\,. \tag{2.36}$$

If $\mathbb{V}(X) = 0$, *then, for all* $b_1 \in \mathbb{R}$,

$$\mathbb{P}[Y = a^* + b_1 X] = 1 \tag{2.37}$$

where $a^* = \mathbb{E}(Y) - b_1\mathbb{E}(X)$.

PROOF By (2.35), we have

$$\mathbb{E}(Y) = a + b\mathbb{E}(X) . \tag{2.38}$$

Suppose $\mathbb{P}[Y = a_1 + b_1 X] = 1$ holds. Then

$$Y = a_1 + b_1 X = a + bX \quad \text{a.s.} \tag{2.39}$$

hence

$$(a_1 - a) + (b_1 - b)X = 0 \quad \text{a.s.} \tag{2.40}$$

$$\mathbb{V}[(a_1 - a) + (b_1 - b)X] = \mathbb{V}[(b_1 - b)X] = (b_1 - b)^2\mathbb{V}(X) = 0 . \tag{2.41}$$

If $\mathbb{V}(X) > 0$, this entails $b_1 = b$, which in turn implies

$$Y = a_1 + b_1 X = a_1 + bX = a + bX \tag{2.42}$$

hence $a_1 = a$. If $\mathbb{V}(X) = 0$, then

$$X - \mathbb{E}(X) = 0 \quad \text{a.s.}, \tag{2.43}$$

hence, for any $b_1 \in \mathbb{R}$,

$$b[X - \mathbb{E}(X)] = b_1[X - \mathbb{E}(X)] = 0 \text{ a.s.}, \tag{2.44}$$

and

$$\begin{aligned} Y &= a + bX \\ &= a + b\mathbb{E}(X) + b[X - \mathbb{E}(X)] \\ &= \mathbb{E}(Y) + b_1[X - \mathbb{E}(X)] \\ &= [\mathbb{E}(Y) - b_1\mathbb{E}(X)] + b_1 X \\ &= a^* + b_1 X \quad \text{a.s.} \end{aligned} \tag{2.45}$$

where $a^* := [\mathbb{E}(Y) - b_1\mathbb{E}(X)]$. □

If $\mathbb{V}(X) > 0$, there is only one pair (a, b) which satisfies (2.35). If $\mathbb{V}(X) = 0$, Y has several representations of the form $a + bX$: the values a and b are not "identified". But they are not completely undetermined. Once b is specified, a is determined by the equation

$$a = \mathbb{E}(Y) - b\mathbb{E}(X) . \tag{2.46}$$

Indeed, if (2.39) holds, we must have

$$(b_1 - b)\mathbb{E}(X) = a - a_1 . \tag{2.47}$$

Corollary 2.3 *Under the assumptions of Theorem* 2.1*,*

$$[\rho(X,Y)^2 = 1] \Leftrightarrow \left[\exists \textit{ two unique constants } a \textit{ and } b \textit{ such that } b \neq 0 \textit{ and } Y = a + bX \textit{ a.s.}\right].$$

3. Regression coefficients between two variables

Definition 3.1 LINEAR REGRESSION COEFFICIENT. *The* linear regression coefficient *of Y on X is defined by*

$$\beta(X \rightharpoondown Y) := \frac{\mathsf{C}(X,Y)}{\mathbb{V}(X)} \tag{3.1}$$

where we set $\beta(X \rightharpoondown Y) := 0$ when $\mathbb{V}(X) = 0$. By convention,

$$\beta(Y \leftharpoondown X) = \beta(X \rightharpoondown Y) . \tag{3.2}$$

The "harpoon" symbols $\rightharpoondown$ and $\leftharpoondown$ represent a statistical "dependence" or "predictability" relation; for example, $X \rightharpoondown Y$ and $Y \leftharpoondown X$ represent dependence of Y on X. The relation $X \rightharpoondown Y$ is typically asymmetric: $Y \leftharpoondown X$ represents a different relation. It does not necessarily correspond to a "causal" relation. From the above definitions, we can write:

$$\mathsf{C}(X,Y) = \rho(X,Y)\,\sigma(X)\,\sigma(Y) \tag{3.3}$$

which holds in all cases [including when $\sigma(X) = 0$ or $\sigma(Y) = 0$]. When $\sigma(X) > 0$, we also have:

$$\beta(X \rightharpoondown Y) = \frac{\rho(X,Y)\,\sigma(X)\sigma(Y)}{\sigma(X)^2} = \rho(X,Y)\,\frac{\sigma(Y)}{\sigma(X)} . \tag{3.4}$$

When $\sigma(X) > 0$, we have [by (3.4) and Theorem 2.1(h)]:

$$-\frac{\sigma(Y)}{\sigma(X)} \leq \beta(X \rightharpoondown Y) = \rho(X,Y)\,\frac{\sigma(Y)}{\sigma(X)} \leq \frac{\sigma(Y)}{\sigma(X)} \tag{3.5}$$

so that the regression coefficient can be bounded the variance ratio $\sigma(Y)/\sigma(X)$. More generally, if $\sigma(X) > 0$ and

$$\rho_L \leq \rho(X,Y) \leq \rho_U , \tag{3.6}$$

we have

$$\rho_L \frac{\sigma(Y)}{\sigma(X)} \leq \beta(X \dashv Y) \leq \rho_U \frac{\sigma(Y)}{\sigma(X)} . \tag{3.7}$$

4. Uncentered covariances, correlations and regression coefficients

Definition 4.1 UNCENTERED COVARIANCE. *The* uncentered covariance *between X and Y is defined by*

$$\bar{\mathsf{C}}(X,Y) := \bar{\sigma}_{XY} := \mathbb{E}[XY] . \tag{4.1}$$

When $\bar{\mathsf{C}}(X,Y) = 0$, we say that X and Y are orthogonal with respect to zero.

Definition 4.2 UNCENTERED CORRELATION. *The* uncentered correlation *between X and Y is defined by*

$$\bar{\rho}(X,Y) := \bar{\rho}_{XY} := \frac{\bar{\mathsf{C}}(X,Y)}{\bar{\sigma}(X)\bar{\sigma}(Y)} \tag{4.2}$$

where we set $\rho(X,Y) := 0$ when $\bar{\sigma}(X)\bar{\sigma}(Y) = 0$.

Definition 4.3 UNCENTERED LINEAR REGRESSION COEFFICIENT. *The* uncentered linear regression coefficient *of Y on X is defined by*

$$\bar{\beta}(X \dashv Y) := \frac{\bar{\mathsf{C}}(X,Y)}{\bar{\sigma}(X)} \tag{4.3}$$

where we set $\bar{\beta}(X \dashv Y) := 0$ when $\bar{\sigma}(X) = 0$.

5. Difference and sum of two correlated random variables

Highly correlated random variables tend to be “close”. This feature can be explicated in different ways:

1. by looking at the distribution of the difference $Y - X$;
2. by looking at the difference of two variances (polarization identity);
3. through a “decoupling” representation of covariances and correlations;

4. Hoeffding identity;
5. by looking at the linear regression of Y on X;

5.1. Uncentered second moments

Let us look the difference and the sum of two random variables X and Y:

$$\mathbb{E}[(Y-X)^2] = \mathbb{E}(X^2+Y^2-2XY) = \mathbb{E}(X^2)+\mathbb{E}(Y^2)-2\mathbb{E}(XY)\,. \tag{5.1}$$

$$\mathbb{E}[(Y+X)^2] = \mathbb{E}(X^2+Y^2+2XY) = \mathbb{E}(X^2)+\mathbb{E}(Y^2)+2\mathbb{E}(XY)\,. \tag{5.2}$$

From these, we see that:

$$\mathbb{E}(XY) = \frac{1}{2}\{[\mathbb{E}(X^2)+\mathbb{E}(Y^2)]-\mathbb{E}[(Y-X)^2]\}\,, \tag{5.3}$$

$$\mathbb{E}(XY) = \frac{1}{2}\{\mathbb{E}[(Y+X)^2]-[\mathbb{E}(X^2)+\mathbb{E}(Y^2)]\,. \tag{5.4}$$

The cross second moment $\mathbb{E}(XY)$ can be interpreted in two ways in terms of (uncentered) second moments:

1. $\mathbb{E}(XY)$ is equal to half the difference between the sum of the second moments X and Y and the second moment of $Y-X$;
2. $\mathbb{E}(XY)$ is equal to half the difference between the second moment of $Y+X$ and the sum of the second moments of X and Y.

5.2. Covariances

We now consider similar expressions for the covariance $\sigma_{XY} = \mathbb{E}[(Y-\mu_Y)-(X-\mu_X)]$. It is easy to see that

$$\begin{aligned}
\mathbb{E}[(Y-X)^2] &= \mathbb{E}\left\{\left([(Y-\mu_Y)-(X-\mu_X)]+(\mu_Y-\mu_X)\right)^2\right\} \\
&= \mathbb{E}\{[(Y-\mu_Y)-(X-\mu_X)]^2\}+(\mu_Y-\mu_X)^2 \\
&= \sigma_Y^2+\sigma_X^2-2\sigma_{XY}+(\mu_Y-\mu_X)^2 \\
&= \sigma_Y^2+\sigma_X^2-2\rho_{XY}\sigma_X\sigma_Y+(\mu_Y-\mu_X)^2\,.
\end{aligned} \tag{5.5}$$

$\mathbb{E}[(Y-X)^2]$ has three components:

1. a *variance component* $\sigma_Y^2+\sigma_X^2$;
2. a *covariance component* $-2\sigma_{XY}$;

3. a *mean component* $(\mu_Y - \mu_X)^2$.

Equation (5.5) shows clearly that $\mathbb{E}[(Y-X)^2]$ tends to be large, when Y and X very different means or variances. Similarly,

$$\begin{aligned}
\mathbb{E}[(Y+X)^2] &= \mathbb{E}\left\{\left([(Y-\mu_Y)+(X-\mu_X)]+(\mu_Y+\mu_X)\right)^2\right\} \\
&= \mathbb{E}\left\{[(Y-\mu_Y)+(X-\mu_X)]^2\right\}+(\mu_Y+\mu_X)^2 \\
&= \sigma_Y^2+\sigma_X^2+2\sigma_{XY}+(\mu_Y+\mu_X)^2 \\
&= \sigma_Y^2+\sigma_X^2+2\rho_{XY}\sigma_X\sigma_Y+(\mu_Y+\mu_X)^2 . \qquad (5.6)
\end{aligned}$$

From (5.5), we see that

$$\begin{aligned}
\sigma_{XY} &= \frac{1}{2}\{(\sigma_Y^2+\sigma_X^2)-\mathbb{E}[(Y-X)^2]+(\mu_Y-\mu_X)^2\} \\
&= \frac{1}{2}\left[(\sigma_Y^2+\sigma_X^2)-\mathbb{E}\{[(Y-\mu_Y)-(X-\mu_X)]^2\}\right] \\
&= \frac{1}{2}\left[(\sigma_Y^2+\sigma_X^2)-\mathbb{V}(Y-X)\right] \\
&= \frac{1}{2}\left[\mathbb{V}(Y)+\mathbb{V}(X)-\mathbb{V}(Y-X)\right]. \qquad (5.7)
\end{aligned}$$

σ_{XY} represents the difference between the sum of the variances of X and Y and the variance of $Y-X$. In particular, if $\mu_Y=\mu_X$,

$$\begin{aligned}
\sigma_{XY} &= \frac{1}{2}\{\sigma_Y^2+\sigma_X^2-\mathbb{E}[(Y-X)^2]\}. \\
&= \frac{1}{2}\{\mathbb{V}(Y)+\mathbb{V}(X)-\mathbb{E}[(Y-X)^2]\} . \qquad (5.8)
\end{aligned}$$

In this case, σ_{XY} represents the difference between the sum of the variances of X and Y and the mean square difference $\mathbb{E}[(Y-X)^2$.

Similarly, by (5.6), we have:

$$\begin{aligned}
\sigma_{XY} &= \frac{1}{2}\{\mathbb{E}[(Y+X)^2]-(\sigma_Y^2+\sigma_X^2)-(\mu_Y+\mu_X)^2\} \\
&= \frac{1}{2}\left[\mathbb{E}\{[(Y-\mu_Y)+(X-\mu_X)]^2\}-(\sigma_Y^2+\sigma_X^2)\right] \\
&= \frac{1}{2}\left[\mathbb{V}(Y+X)-(\sigma_Y^2+\sigma_X^2)\right]
\end{aligned}$$

$$= \frac{1}{2}\left[\mathbb{V}(Y+X) - [\mathbb{V}(Y)+\mathbb{V}(X)]\right]. \tag{5.9}$$

σ_{XY} represents the difference between the variance of $Y+X$ and the sum of the variances of X and Y. In particular, if $\mu_Y = \mu_X = 0$,

$$\begin{aligned} \sigma_{XY} &= \frac{1}{2}\{\mathbb{E}[(Y+X)^2] - (\sigma_Y^2 + \sigma_X^2)\} \\ &= \frac{1}{2}\{\mathbb{E}[(Y+X)^2] - [\mathbb{V}(Y)+\mathbb{V}(X)]\} \\ &= \frac{1}{2}\{\mathbb{E}[(Y+X)^2] - [\mathbb{E}(Y^2)+\mathbb{E}(X^2)]\}. \end{aligned} \tag{5.10}$$

In this case, σ_{XY} represents the difference between the sum of the variances of Y and X and the mean square difference $\mathbb{E}[(Y-X)^2]$.

In general, we thus have:

$$\begin{aligned} \sigma_{XY} &= \frac{1}{2}\{[\mathbb{V}(Y)+\mathbb{V}(X)] - \mathbb{V}(Y-X)\} \\ &= \frac{1}{2}\{\mathbb{V}(Y+X) - [\mathbb{V}(Y)+\mathbb{V}(X)]\}. \end{aligned} \tag{5.11}$$

If $\mu_Y = \mu_X$,

$$\sigma_{XY} = \frac{1}{2}\{[\mathbb{V}(Y)+\mathbb{V}(X)] - \mathbb{E}[(Y-X)^2]\} \tag{5.12}$$

and, if $\mu_Y = \mu_X = 0$,

$$\begin{aligned} \sigma_{XY} &= \frac{1}{2}\{[\mathbb{E}(Y^2)+\mathbb{E}(X^2)] - \mathbb{E}[(Y-X)^2]\} \\ &= \frac{1}{2}\{\mathbb{E}[(Y+X)^2] - [\mathbb{E}(Y^2)+\mathbb{E}(X^2)]\}. \end{aligned} \tag{5.13}$$

5.3. Correlations

From (5.5), it is also easy to see that

$$\mathbb{E}\left[\left(\frac{Y}{\sigma_Y} - \frac{X}{\sigma_X}\right)^2\right] = 2(1-\rho_{XY}) + \left(\frac{\mu_Y}{\sigma_Y} - \frac{\mu_X}{\sigma_X}\right)^2, \tag{5.14}$$

$$\mathbb{E}\left[\left(\frac{Y}{\sigma_Y} + \frac{X}{\sigma_X}\right)^2\right] = 2(1+\rho_{XY}) + \left(\frac{\mu_Y}{\sigma_Y} + \frac{\mu_X}{\sigma_X}\right)^2. \tag{5.15}$$

Consider the normalized values of X and Y :

$$\tilde{X} = \frac{X-\mu_X}{\sigma_X}, \quad \tilde{Y} = \frac{Y-\mu_Y}{\sigma_Y}, \quad \rho(\tilde{X},\tilde{Y}) = \rho(X,Y) := \rho_{XY}, \tag{5.16}$$

where we set $\tilde{X}=0$ if $\sigma_X=0$, and $\tilde{Y}=0$ if $\sigma_Y=0$. We then have:

$$\mathbb{E}(\tilde{X}) = \mathbb{E}(\tilde{Y}) = 0, \quad \mathbb{V}(\tilde{X}) = \mathbb{V}(\tilde{Y}) = 1, \tag{5.17}$$

and

$$\mathbb{E}[(\tilde{Y}-\tilde{X})^2] = 2(1-\rho_{XY}), \tag{5.18}$$

$$\rho_{XY} = 1-\frac{1}{2}\mathbb{E}[(\tilde{Y}-\tilde{X})^2]. \tag{5.19}$$

The correlation $\rho(X,Y)$ is inversely related to the mean-square distance $\mathbb{E}[(\tilde{Y}-\tilde{X})^2]$ between $\tilde{X}$ and $\tilde{Y}$. (5.19) is a general form of the standard formula for Spearman's rank correlation coefficient.

Similarly,

$$\mathbb{E}[(\tilde{Y}+\tilde{X})^2] = 2(1+\rho_{XY}), \tag{5.20}$$

$$\rho_{XY} = \frac{1}{2}\mathbb{E}[(\tilde{Y}+\tilde{X})^2]-1. \tag{5.21}$$

The correlation $\rho(X,Y)$ measures the mean square $\mathbb{E}[(\tilde{Y}+\tilde{X})^2]$ of the sum of $\tilde{X}+\tilde{Y}$. The above formulae can also be rewritten in terms of the arithmetic mean of $\tilde{X}$ and $\tilde{Y}$:

$$\mathbb{E}\{\big[\frac{1}{2}(\tilde{Y}+\tilde{X})\big]^2\} = \frac{1}{2}(1+\rho_{XY}), \tag{5.22}$$

$$\rho_{XY} = 2\mathbb{E}\{\big[\frac{1}{2}(\tilde{Y}+\tilde{X})\big]^2\}-1 \tag{5.23}$$

5.4. Inequalities

Since $|\rho_{XY}| \leq 1$, it is interesting to observe that

$$(\sigma_Y-\sigma_X)^2+(\mu_Y-\mu_X)^2 \leq \mathbb{E}[(Y-X)^2] \leq (\sigma_Y+\sigma_X)^2+(\mu_Y-\mu_X)^2, \tag{5.24}$$

and

$$\mathbb{E}[(Y-X)^2] \leq \sigma_Y^2+\sigma_X^2+(\mu_Y-\mu_X)^2 \leq (\sigma_Y+\sigma_X)^2+(\mu_Y-\mu_X)^2, \text{ if } \rho_{XY} \geq 0, \tag{5.25}$$

$$\mathbb{E}[(Y-X)^2] \geq \sigma_Y^2+\sigma_X^2+(\mu_Y-\mu_X)^2 \geq (\sigma_Y-\sigma_X)^2+(\mu_Y-\mu_X)^2, \text{ if } \rho_{XY} \leq 0, \tag{5.26}$$

$$\mathbb{E}[(Y-X)^2] = \sigma_Y^2+\sigma_X^2+(\mu_Y-\mu_X)^2, \text{ if } \rho_{XY} = 0. \tag{5.27}$$

$\mathbb{E}[(Y-X)^2]$ reaches its minimum value when $\rho_{XY}=1$, and its maximal value when $\rho_{XY}=-1$:

$$\mathbb{E}[(Y-X)^2] = (\sigma_Y - \sigma_X)^2 + (\mu_Y - \mu_X)^2\,, \quad \text{if } \rho_{XY} = 1\,, \tag{5.28}$$

-

$$\mathbb{E}[(Y-X)^2] = (\sigma_Y + \sigma_X)^2 + (\mu_Y - \mu_X)^2\,, \quad \text{if } \rho_{XY} = -1\,. \tag{5.29}$$

If $\sigma_Y^2 > 0$, we can also write:

$$\left(1-\frac{\sigma_X}{\sigma_Y}\right)^2 + \left(\frac{\mu_Y-\mu_X}{\sigma_Y}\right)^2 \le \frac{\mathbb{E}[(Y-X)^2]}{\sigma_Y^2} \le \left(1+\frac{\sigma_X}{\sigma_Y}\right)^2 + \left(\frac{\mu_Y-\mu_X}{\sigma_Y}\right)^2\,. \tag{5.30}$$

The inequalities (5.24) - (5.27) also entail similar properties for $X+Y$:

$$(\sigma_X - \sigma_Y)^2 + (\mu_X + \mu_Y)^2 \le \mathbb{E}[(X+Y)^2] \le (\sigma_X + \sigma_Y)^2 + (\mu_X + \mu_Y)^2\,, \tag{5.31}$$

$$\mathbb{E}[(X+Y)^2] \le \sigma_X^2 + \sigma_Y^2 + (\mu_X + \mu_Y)^2 \le (\sigma_Y + \sigma_X)^2 + (\mu_X + \mu_Y)^2\,, \text{ if } \rho_{XY} \le 0\,, \tag{5.32}$$

$$\mathbb{E}[(X+Y)^2] \ge \sigma_X^2 + \sigma_Y^2 + (\mu_X + \mu_Y)^2 \ge (\sigma_X - \sigma_Y)^2 + (\mu_X + \mu_Y)^2\,, \text{ if } \rho_{XY} \ge 0\,, \tag{5.33}$$

$$\mathbb{E}[(Y+X)^2] = \sigma_X^2 + \sigma_Y^2 + (\mu_X + \mu_Y)^2\,, \text{ if } \rho_{XY} = 0\,. \tag{5.34}$$

By (5.18), we have:

$$0 \le \mathbb{E}[(\tilde{Y}-\tilde{X})^2] \le 4\,, \tag{5.35}$$

$$0 \le \mathbb{E}[|\tilde{Y}-\tilde{X}|] \le \{\mathbb{E}[(\tilde{Y}-\tilde{X})^2]\}^{1/2} \le 2\,. \tag{5.36}$$

The root mean square error of approximating $\tilde{Y}$ by $\tilde{X}$ cannot be larger than 2. Upon using the Chebyshev inequality, this entails:

$$\mathbb{P}\left[|\tilde{Y}-\tilde{X}| \ge \lambda\right] \le \frac{\mathbb{E}[(\tilde{Y}-\tilde{X})^2]}{\lambda^2} \le \frac{4}{\lambda^2}\,. \tag{5.37}$$

Since

$$X = \mu_X + \sigma_X \tilde{X}\,, \quad Y = \mu_Y + \sigma_Y \tilde{Y}\,, \tag{5.38}$$

we get

$$\begin{aligned}
\mathbb{E}[(Y-X)^2] &= \mathbb{E}\left\{[(\mu_Y + \sigma_Y\tilde{Y}) - (\mu_X + \sigma_X\tilde{X})]^2\right\} \\
&= \mathbb{E}\left\{[(\sigma_Y\tilde{Y} - \sigma_X\tilde{X}) + (\mu_Y - \mu_X)]^2\right\} \\
&= \mathbb{E}\left\{[(\sigma_Y\tilde{Y} - \sigma_X\tilde{X}) + (\mu_Y - \mu_X)]^2\right\}
\end{aligned}$$

$$= \quad \mathbb{E}[(\sigma_Y \tilde{Y} - \sigma_X \tilde{X})^2] + (\mu_Y - \mu_X)^2 \tag{5.39}$$

hence

$$\begin{aligned}\mathbb{E}[(Y-X)^2] &= \sigma_Y^2 \mathbb{E}\left[\left(\tilde{Y} - \frac{\sigma_X}{\sigma_Y}\tilde{X}\right)^2\right] + (\mu_Y - \mu_X)^2 \\ &= \sigma_Y^2 \left[1 + \left(\frac{\sigma_X}{\sigma_Y}\right)^2 - 2\left(\frac{\sigma_X}{\sigma_Y}\right)\rho_{XY}\right] + (\mu_Y - \mu_X)^2, \quad \text{if } \sigma_Y \neq 0,\end{aligned} \tag{5.40}$$

and

$$\mathbb{E}[(Y-X)^2] = \sigma_X^2 + (\mu_Y - \mu_X)^2, \quad \text{if } \sigma_Y = 0. \tag{5.41}$$

If the variances of X and Y are the same, *i.e.*

$$\sigma_Y^2 = \sigma_X^2, \tag{5.42}$$

we have:

$$\begin{aligned}\mathbb{E}[(Y-X)^2] &= 2\sigma_Y^2(1-\rho_{XY}) + (\mu_Y - \mu_X)^2 \\ &= 2\sigma_X^2(1-\rho_{XY}) + (\mu_Y - \mu_X)^2.\end{aligned} \tag{5.43}$$

If the means and variances of X and Y are the same, *i.e.*

$$\mu_Y = \mu_X \text{ and } \sigma_Y^2 = \sigma_X^2, \tag{5.44}$$

we have:

$$\mathbb{E}[(Y-X)^2] = 2\sigma_Y^2\,(1-\rho_{XY}) = 2\sigma_X^2\,(1-\rho_{XY}) \tag{5.45}$$

and

$$0 \leq \mathbb{E}[(Y-X)^2] \leq 4\sigma_X^2 \tag{5.46}$$

so that

$$\mathbb{E}[(Y-X)^2] = 0 \text{ and } \mathbb{P}[Y = X] = 1, \text{ if } \rho_{XY} = 1, \tag{5.47}$$

and, using Chebyshev's inequality,

$$\mathbb{P}[|Y-X| > c] \leq \frac{\mathbb{E}[(Y-X)^2]}{c^2} = \frac{2\sigma_X^2\,(1-\rho_{XY})}{c^2} \text{ for any } c > 0, \tag{5.48}$$

$$\mathbb{P}\,[|Y-X| > c\sigma_X] \leq \frac{\mathbb{E}[(Y-X)^2]}{\sigma_X^2\, c^2} = \frac{2(1-\rho_{XY})}{c^2} \text{ for any } c > 0. \tag{5.49}$$

If $\mu_Y = \mu_X$ and $\sigma_Y^2 = \sigma_X^2 > 0$, we also have:

$$\mathbb{E}[(Y-X)^2] = 0 \Leftrightarrow \rho_{XY} = 1\,, \tag{5.50}$$

$$\mathbb{E}[(Y-X)^2] = 2\sigma_X^2 \Leftrightarrow \rho_{XY} = 0\,, \tag{5.51}$$

$$\mathbb{E}[(Y-X)^2] = 4\sigma_X^2 \Leftrightarrow \rho_{XY} = -1\,. \tag{5.52}$$

Since

$$\sigma_Y(\tilde{Y}-\tilde{X}) = Y - \mu_Y - \frac{\sigma_Y}{\sigma_X}(X-\mu_X) = Y - \left(\mu_Y + \frac{\sigma_Y}{\sigma_X}\mu_X\right) - \frac{\sigma_Y}{\sigma_X}X\,, \tag{5.53}$$

the linear function

$$L_0(X) = \left(\mu_Y + \frac{\sigma_Y}{\sigma_X}\mu_X\right) + \frac{\sigma_Y}{\sigma_X}X \tag{5.54}$$

can be viewed as a "forecast" of Y based on X such that

$$\mathbb{E}[(Y-L_0(X))^2] = \sigma_Y^2\,\mathbb{E}[(\tilde{Y}-\tilde{X})^2] = 2\sigma_Y^2(1-\rho_{XY})\,. \tag{5.55}$$

It is then of interest to note that

$$\mathbb{E}[(Y-L_0(X))^2] \leq \mathbb{E}[(Y-\mu_Y)^2] = \sigma_Y^2 \Leftrightarrow \rho_{XY} \geq 0.5\,, \tag{5.56}$$

with

$$\mathbb{E}[(Y-L_0(X))^2] < \mathbb{E}[(Y-\mu_Y)^2] = \sigma_Y^2 \Leftrightarrow \rho_{XY} > 0.5 \tag{5.57}$$

when $\sigma_Y^2 > 0$. Thus $L_0(X)$ provides a "better forecast" of Y than the mean of Y, when $\rho_{XY} > 0.5$. If $\rho_{XY} < 0.5$ and $\sigma_Y^2 > 0$, the opposite holds: $\mathbb{E}[(Y-L_0(X))^2] > \sigma_Y^2$.

5.5. Polarization identities

Since

$$\mathbb{E}[(Y-X)^2] = \mathbb{E}(X^2+Y^2-2XY) = \mathbb{E}(X^2)+\mathbb{E}(Y^2)-2\mathbb{E}(XY)\,, \tag{5.58}$$

$$\mathbb{E}[(Y+X)^2] = \mathbb{E}(X^2+Y^2+2XY) = \mathbb{E}(X^2)+\mathbb{E}(Y^2)+2\mathbb{E}(XY)\,, \tag{5.59}$$

we get on summing the above two equations:

$$\mathbb{E}(XY) = \frac{1}{4}\{\mathbb{E}[(Y+X)^2] - \mathbb{E}[(Y-X)^2]\}\,. \tag{5.60}$$

Similarly, since

$$\mathbb{V}(X+Y) = \mathbb{V}(X) + \mathbb{V}(Y) + 2\mathsf{C}(X,Y)\,, \tag{5.61}$$

$$\mathbb{V}(X-Y)=\mathbb{V}(X)+\mathbb{V}(Y)-2\mathsf{C}(X,Y)\,, \tag{5.62}$$

we have:

$$\mathsf{C}(X,Y)=\frac{1}{4}[\mathbb{V}(X+Y)-\mathbb{V}(X-Y)]\,. \tag{5.63}$$

(5.63) is sometimes called the "polarization identity". Further,

$$\rho\,(X,Y)=\frac{1}{4}\frac{\mathbb{V}(X+Y)-\mathbb{V}(X-Y)}{\sigma_X\,\sigma_Y}=\frac{1}{4}\left[\frac{\sigma^2_{X+Y}}{\sigma_X\,\sigma_Y}-\frac{\sigma^2_{X-Y}}{\sigma_X\,\sigma_Y}\right] \tag{5.64}$$

and, if $\mathbb{V}(X)=\mathbb{V}(Y)=1$,

$$\rho\,(X,Y)=\frac{\mathbb{V}(X+Y)-\mathbb{V}(X-Y)}{4}=\frac{\sigma^2_{X+Y}-\sigma^2_{X-Y}}{4}\,. \tag{5.65}$$

On $X+Y$ and $X-Y$, it also interesting to observe that

$$\mathsf{C}(X+Y,X-Y)=[\mathbb{V}(X)-\mathbb{V}(Y)]+[\mathsf{C}(Y,X)-\mathsf{C}(X,Y)]=\mathbb{V}(X)-\mathbb{V}(Y) \tag{5.66}$$

so that

$$\mathsf{C}\big((X+Y)/2,X-Y\big)=\mathsf{C}(X+Y,X-Y)=0\,,\quad \text{if } \mathbb{V}(X)=\mathbb{V}(Y)\,. \tag{5.67}$$

This holds irrespective of the covariance between between X and Y. In particular, if the vector (X,Y) is multinormal $X+Y$ and $X-Y$ are independent when $\mathbb{V}(X)=\mathbb{V}(Y)$.

On applying (5.64) to the normalized variables $\tilde{Y}$ and $\tilde{X}$, we get a polarization formula in terms of normalized variables:

$$\rho\,(X,Y)=\frac{\mathbb{V}(\tilde{Y}+\tilde{X})-\mathbb{V}(\tilde{Y}-\tilde{X})}{4}=\frac{\mathbb{E}[(\tilde{Y}+\tilde{X})^2]-\mathbb{E}[(\tilde{Y}-\tilde{X})^2]}{4}\,. \tag{5.68}$$

This also follows on applying (5.64) to $\tilde{Y}$ and $\tilde{X}$.

6. 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.