Pairwise Difference Representations of Moments: Gini and Generalized Lagrange identities

Jean-Marie Dufour [*] Abderrahim Taamouti [†] Meilin Tong [‡]

McGill University University of Liverpool McGill University

December 2025

[*] 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

[†] Professor and Chair in Applied Econometrics, University of Liverpool Management School, Chatham St, Liverpool, L69 7ZH, UK. TEL: 44 (0)151 795 1773; e-mail: Abderrahim.Taamouti@liverpool.ac.uk

[‡] Department of Economics, McGill University, Leacock Building, Room 414, 855 Sherbrooke Street West, Montréal, Québec H3A 2T7, Canada. e-mail: meilin.tong@mail.mcgill.ca.

ABSTRACT

We provide pairwise-difference (Gini-type) representations of higher-order central moments for both general random variables and empirical moments. Such representations do not require a measure of location. For third and fourth moments, this yields pairwise-difference representations of skewness and kurtosis coefficients. We show that all central moments possess such representations, so no reference to the mean is needed for moments of any order. This is done by considering i.i.d. *replications* of the random variables considered, by observing that central moments can be interpreted as covariances between a random variable and powers of the same variable, and by giving recursions which link the pairwise-difference representation of any moment to lower order ones. Numerical summation identities are deduced. Through a similar approach, we give analogues of the Lagrange and Binet-Cauchy identities for general random variables, along with a simple derivation of the classic Cauchy-Schwarz inequality for covariances. Finally, an application to unbiased estimation of centered moments is discussed.

Key words: Gini; Covariance; Skewness; Kurtosis; Pairwise differences; Lagrange identity; Moments; Unbiased estimation.

1. Introduction

The sample variance is certainly the most widely used measure of dispersion among observations in statistics. It is typically defined as the average of the squared deviations of the observations $x_1, \ldots, x_n$ from their sample mean:

$$s_X^2(n) := \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2 \tag{1.1}$$

where $\bar{x} = \sum_{i=1}^{n} x_i / n$. At first sight, the variance depends crucially on "centering" the observations with respect to their mean. A century ago, however, Gini (1912) noticed that $s_X^2(n)$ can be rewritten as the average of pairwise differences between all the observations:

$$s_X^2(n) = \frac{1}{2} \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j=1}^{n} (x_i - x_j)^2 = \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j>i} (x_i - x_j)^2 ; \tag{1.2}$$

see Heffernan (1988). The Gini representation underscores that $s_X^2(n)$ does not depend on using the sample mean as location parameter, and thus may be viewed as a measure of "intrinsic variability" between observations. In particular, it provides a way of measuring dispersion when there is no natural location parameter, as with categorical data; see Light and Margolin (1971, 1974). Further, by focusing on pairwise differences $x_i - x_j$, it may be easier to determine which observations have the greatest influence on $s_X^2(n)$, because deviations from the mean $x_i - \bar{x}$ depend on all the observations (through $\bar{x}$). A comprehensive review of statistical methods based on the Gini approach is presented by Yitzhaki and Schechtman (2013).

If we have bivariate observations $z_i := (x_i, y_i)'$, $i = 1, \ldots, n$, the sample covariance

$$s_{XY}(n) := \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) \tag{1.3}$$

is in turn the most widely used "measure of association" between x and y. For the covariance, a pairwise representation is provided by the formula

$$s_{XY}(n) = \frac{1}{2}\frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j=1}^{n}(x_i - x_j)(y_i - y_j) = \frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j>i}(x_i - x_j)(y_i - y_j)\,; \tag{1.4}$$

see Hayes (2011). Clearly, (1.2) is a special case of (1.4) obtained by taking $x_i = y_i$, $i = 1, \ldots, n$. Formula (1.4) shows that the sample covariance measures the tendency of x and y to move in the same direction, without reference to sample means. Similar expressions for the variances and covariances of general random variables have also been used in the literature on U-statistics; see Lee (1990, Chapter 1). Since centering is not needed, formula (1.4) provides a basis for developing alternative measures of dependence, such as Kendall's and Spearman's measures of dependence [see Lehmann (1966)]. Again, pairwise differences $x_i - x_j$ and $y_i - y_j$ may be easier to interpret than $x_i - \bar{x}$ and $y_i - \bar{y}$, because $\bar{x}$ and $\bar{y}$ depend on all the observations.

Another similar result we consider here is the Lagrange identity:

$$\begin{aligned}\Big(\sum_{i=1}^{n}x_i^2\Big)\Big(\sum_{i=1}^{n}y_i^2\Big) - \Big(\sum_{i=1}^{n}x_i y_i\Big)^2 &= \sum_{1\le i<j\le n}(x_i y_j - x_j y_i)^2 \\ &= \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}(x_i y_j - x_j y_i)^2 = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}(\det[z_i, z_j])^2\,;\end{aligned} \tag{1.5}$$

see Wright (1992) and Steele (2004, Chapter 3). An interesting feature of this identity is that it involves pairwise comparisons between cross-products of the components of $x :=$

$(x_1, \ldots, x_n)'$ and $y := (y_1, \ldots, y_n)'$. Since the right-hand side is non-negative, the Lagrange identity provides a simple way of proving the discrete Cauchy-Schwarz inequality:

$$\left(\sum_{i=1}^{n} x_i y_i\right)^2 \leq \left(\sum_{i=1}^{n} x_i^2\right)\left(\sum_{i=1}^{n} y_i^2\right). \tag{1.6}$$

Interestingly, (1.5) directly shows when the bound is an equality $(x_i y_j = x_j y_i$ for all i and $j)$ and what determines its tightness. The Lagrange identity is itself a special case of the Binet-Cauchy identity

$$\left(\sum_{i=1}^{n} a_i c_i\right)\left(\sum_{j=1}^{n} b_j d_j\right) - \left(\sum_{i=1}^{n} a_i d_i\right)\left(\sum_{j=1}^{n} b_j c_j\right) = \sum_{1 \leq i < j \leq n} (a_i b_j - a_j b_i)(c_i d_j - c_j d_i) \tag{1.7}$$

where a_i, b_i, c_i, d_i, $i = 1, \ldots, n$, are arbitrary real constants [see Steele (2004, p. 49)]. Formula (1.7) relates several inner products to pairwise comparisons between the components of the vectors considered.

Since the representations in (1.2) and (1.4) only depend on *differences* between observations, we will call such representations (and similar ones) *D-representations* (or *Gini-type representations*). In this paper, we provide D-representations of higher-order central moments for both general random variables and empirical moments. For third and fourth moments, this yields new representations of skewness and kurtosis coefficients, where reference to a location parameter is eliminated. We show that all central moments possess D-representations, so again no reference to the mean is needed for moments of any order. This is done by considering i.i.d. *replications* (or realizations) of the random variables considered, by observing that central moments can be interpreted as covariances between a random variable and powers of the same variable (a *nonlinear* relation), and by giving *recursions* which link the D-representation of any moment to lower order ones. Numerical summation identities similar to (1.2) - (1.4) are also deduced. Finally, through a similar approach, we give the analogues of the Lagrange and Binet-Cauchy identities for general random variables, along with a simple derivation of the classic Cauchy-Schwarz inequality for covariances.

The paper is organized as follows. In Section 2, we discuss *D*-representations for covariances and regression coefficients, and we show how corresponding numerical identities can be derived from these. In Section 3, we provide *D*-representations for third and fourth moments, along with corresponding expressions for skewness and kurtosis coefficients. In Section 4, we show that all moments possess such representations, and we give recursive formulae for computing a *D*-representation for any moment. Generalized Lagrange and Binet-Cauchy identities are given in Section 5. In Section 6, we discuss some applications of the proposed moment representations. We conclude in Section 7. Proofs are given in an online Appendix.

2. *D*-representation of Covariances

In order to generalize the expressions (1.2) - (1.4) to higher-order moments, it will be useful to discuss the pairwise difference representation of the covariance between general random variables. Let X and Y be random variables with finite second moments, and means $\mu_X = \mathbb{E}\{X\}$, $\mu_Y = \mathbb{E}\{Y\}$. We denote by $\mathsf{C}[X,Y] := \mathbb{E}\{(X-\mu_X)(Y-\mu_Y)\}$ the covariance between X and Y. Consider two independent and identically distributed replications of the random vector $(X,Y)'$: $(X_1,Y_1)'$ and $(X_2,Y_2)'$. On observing

$$\mathbb{E}\{(X_1-X_2)(Y_1-Y_2)\} \quad = \quad \mathbb{E}\{X_1Y_1\}+\mathbb{E}\{X_2Y_2\}-\mathbb{E}\{X_1Y_2\}-\mathbb{E}\{X_2Y_1\}$$

$$
\begin{aligned}
&= \mathbb{E}\{X_1Y_1\} + \mathbb{E}\{X_2Y_2\} - \mathbb{E}\{X_1\}\mathbb{E}\{Y_2\} - \mathbb{E}\{X_2\}\mathbb{E}\{Y_1\} \\
&= 2(\mathbb{E}\{XY\} - \mathbb{E}\{X\}\mathbb{E}\{Y\}) = 2\mathsf{C}[X, Y], \qquad (2.1)
\end{aligned}
$$

it follows that

$$\mathsf{C}[X, Y] = \frac{1}{2}\mathbb{E}\{(X_1 - X_2)(Y_1 - Y_2)\}. \qquad (2.2)$$

It is also easy to see that

$$
\begin{aligned}
\mathsf{C}[X, Y] &= \mathbb{E}\{X_1(Y_1 - Y_2)\} = -\mathbb{E}\{X_2(Y_1 - Y_2)\} \\
&= \mathbb{E}\{(X_1 - X_2)Y_1\} = -\mathbb{E}\{(X_1 - X_2)Y_2\}. \qquad (2.3)
\end{aligned}
$$

In other words, only one of the variables in $(X, Y)'$ need be replicated to avoid centering with respect to the means (μ_X and μ_Y). (2.1) - (2.3) also hold under the weaker assumption (without identical distributions) that $(X_1, Y_1)'$ and $(X_2, Y_2)'$ have the same first and second moments as (X, Y) with

$$\mathbb{E}\{X_1Y_2\} = \mathbb{E}\{X_2Y_1\} = \mathbb{E}\{X\}\mathbb{E}\{Y\}. \qquad (2.4)$$

For $X_1 = Y_1$ and $X_2 = Y_2$, we obtain the following formulae for the variance of X :

$$
\begin{aligned}
\sigma_X^2 &= \mathrm{Var}(X) = \frac{1}{2}\mathbb{E}\{(X_1 - X_2)^2\} \\
&= \mathbb{E}\{X_1(X_1 - X_2)\} = -\mathbb{E}\{X_2(X_1 - X_2)\}. \qquad (2.5)
\end{aligned}
$$

If X_3 is a third replication of X [so that X_1, X_2, X_3 are i.i.d.], it is also easy to check that

$$\sigma_X^2 = \mathbb{E}\{(X_1 - X_3)(X_1 - X_2)\}. \qquad (2.6)$$

The uncentered second moment of X can be written as

$$\begin{aligned}\mathbb{E}\{X^2\} &= \frac{1}{2}\mathbb{E}\{(X_1-X_2)^2\}+\mu_X^2 \\ &= \mathbb{E}\{X_1(X_1-X_2)\}+\mu_X^2 = -\mathbb{E}\{X_2(X_1-X_2)\}+\mu_X^2 \end{aligned} \tag{2.7}$$

and the linear regression coefficient of Y on X is

$$\begin{aligned}\beta_{Y\cdot X} &= \frac{\sigma_{XY}}{\sigma_X^2} = \frac{\mathbb{E}\{(X_1-X_2)(Y_1-Y_2)\}}{\mathbb{E}\{(X_1-X_2)^2\}} \\ &= \frac{\mathbb{E}\{(X_1-X_2)Y_1\}}{\mathbb{E}\{(X_1-X_2)X_1\}} = \frac{\mathbb{E}\{X_1(Y_1-Y_2)\}}{\mathbb{E}\{X_1(X_1-X_2)\}}. \end{aligned} \tag{2.8}$$

In view of the above discussion on the covariance and the variance, we will define more formally what we mean by the D-representation of a parameter.

Definition 2.1 *A parameter γ has a D-representation for the variables $X_1,\ldots,X_n$ if there is a function $g(x_1,\ldots,x_n)=h(x_2-x_1,\ldots,x_n-x_{n-1})$ such that*

$$\gamma=\mathbb{E}\{h(X_2-X_1,\ldots,X_n-X_{n-1})\}. \tag{2.9}$$

In other words, γ has a D-representation for the variables $X_1,\ldots,X_n$ if it can be written as the expected value of a function $g(X_1,\ldots,X_n)$ which depends on $X_1,\ldots,X_n$ only through the differences between the variables $X_1,\ldots,X_n$. Equivalently, this means that

$$\gamma=\mathbb{E}\{g(X_1,\ldots,X_n)\} \tag{2.10}$$

where $g(x_1, \ldots, x_n)$ is invariant to uniform location changes, *i.e.*

$$g(x_1 + c, \ldots, x_n + c) = g(x_1, \ldots, x_n) \quad \text{for any } c. \tag{2.11}$$

A potentially useful feature of the representation (2.2) follows from the fact that the variables $(X_1 - X_2)$ and $(Y_1 - Y_2)$ have marginal distributions symmetric around zero. Consequently, the odd moments of $(X_1 - X_2)$ and $(Y_1 - Y_2)$ are all zero (when they exist). More generally, the distribution of the vector $(X_1 - X_2, Y_1 - Y_2)'$ is jointly symmetric around zero in the sense that

$$(X_1 - X_2, Y_1 - Y_2)' \sim -(X_1 - X_2, Y_1 - Y_2)'. \tag{2.12}$$

This type of symmetry can be useful for proving unbiasedness results in various setups; see Dufour (1984, 1985). This property is shared by all D-representations when the paired differences are based on i.i.d. observations.

It is of interest to note that the numerical identity (1.4) can be derived from the general moment formula (2.2). Let (I, J) be a pair of random integers such that

$$\mathbb{P}[(I, J) = (i, j)\} = (\frac{1}{n})^2, \; i, \, j = 1, ..., n, \tag{2.13}$$

and consider the random vector

$$(X^*, Y^*) = (x_I, y_J)\,. \tag{2.14}$$

The values $x_1, \dots, x_n$ and $y_1, \dots, y_n$ need not be distinct. It is then easy to see that:

$$\mathbb{E}\{X^*\} = \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} x_i = \bar{x}, \quad \mathbb{E}\{Y^*\} = \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} y_j = \bar{y}, \tag{2.15}$$

$$\mathbb{E}\{X^* Y^*\} = \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} x_i y_j \,, \tag{2.16}$$

and by the definition of the covariance,

$$\sigma_{X^* Y^*} = \mathbb{E}\{X^* Y^*\} - \mathbb{E}\{X^*\}\mathbb{E}\{Y^*\} = \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} x_i y_j - \bar{x}\bar{y}\,. \tag{2.17}$$

Consider now two i.i.d. replications of (X^*, Y^*): (X_1^*, Y_1^*) and (X_2^*, Y_2^*). By applying (2.2) and (2.16) to (X^*, Y^*), the covariance between X^* and Y^* is given by:

$$\sigma_{X^* Y^*} = \frac{1}{2}\mathbb{E}\{(X_1^* - X_2^*)(Y_1^* - Y_2^*)\} = \frac{1}{2}\frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} (x_i - x_j)(y_i - y_j)\,. \tag{2.18}$$

Since the above sum contains a zero whenever $i = j$, it is natural to divide by $(n^2 - n)$ rather than n^2. This leads to:

$$s_{X^* Y^*} := (\frac{n}{n-1})\sigma_{X^* Y^*} = \frac{1}{2}\frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j=1}^{n} (x_i - x_j)(y_i - y_j) = s_{XY}(n)\,. \tag{2.19}$$

In the special case where $X^* = Y^*$, we have

$$\sigma_{X^*}^2 = \frac{1}{2}\mathbb{E}\{(X_1^* - X_2^*)^2\} = \frac{1}{2}\frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} (x_i - x_j)^2 \tag{2.20}$$

hence, on setting $s^2_{X^*} := (\frac{n}{n-1})\sigma^2_{X^*}$, the Gini formula (1.2) for the variance is:

$$s^2_{X^*} = \frac{1}{2}\frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j=1}^{n}(x_i - x_j)^2 = s^2_X(n)\,. \tag{2.21}$$

3. Skewness and Kurtosis

In this section, we express skewness and kurtosis coefficients in terms of pairwise differences. Such representations can be written as functions of differences of the form $X_i - X_j$ or differences between powers of variables $[X_i^2 - X_j^2]$. We first underscore that the third and fourth central moments of a random variable X can be interpreted as covariances, which in turn can be represented in terms of pairwise differences between replicated random variables. In each case, we give several expressions which may have independent interest.

Proposition 3.1 COVARIANCE REPRESENTATION OF THIRD AND FOURTH CENTRAL MOMENTS. *Let X be a random variable with finite mean μ_X and variance σ^2_X. Let X_1, X_2 two* i.i.d. *replications of X. If X has finite third moment, then*

$$\begin{aligned}
\mathbb{E}\{(X-\mu_X)^3\} &= \mathsf{C}[X, (X-\mu_X)^2] = \mathsf{C}[X, X^2] - 2\mu_X\sigma^2_X \\
&= \mathbb{E}\{(X_1 - X_2)(X_1-\mu_X)^2\} = \mathbb{E}\{(X_1-X_2)X_1^2\} - 2\mu_X\sigma^2_X \\
&= \frac{1}{2}\mathbb{E}\{(X_1-X_2)[(X_1-\mu_X)^2 - (X_2-\mu_X)^2]\} \\
&= \frac{1}{2}\mathbb{E}\{(X_1-X_2)(X_1^2 - X_2^2)\} - \mu_X\mathbb{E}\{(X_1-X_2)^2\}\,.
\end{aligned} \tag{3.1}$$

If X has finite fourth moment, then

$$\mathbb{E}\{(X-\mu_X)^4\} = \mathsf{C}[X, (X-\mu_X)^3] = \mathsf{C}[X, X^3] - 3\mu_X\mathsf{C}[X, X^2] + 3\mu_X^2\sigma^2_X$$

$$
\begin{aligned}
&= \mathbb{E}\{(X_1 - X_2)(X_1 - \mu_X)^3\} = \mathbb{E}\{(X_1 - X_2)(X_1^2 - 2\mu_X X_1)\} \\
&= \frac{1}{2}\mathbb{E}\{(X_1 - X_2)[(X_1 - \mu_X)^3 - (X_2 - \mu_X)^3]\} \\
&= \frac{1}{2}\mathbb{E}\{(X_1 - X_2)(X_1^3 - X_2^3)\} - 3\mu_X \mathbb{E}\{(X - \mu_X)^3\} - 3\mu_X^2 \sigma_X^2 \,. \quad (3.2)
\end{aligned}
$$

To interpret Proposition 3.1, consider first the case where the mean is zero ($\mu_X = 0$), so the third moment does not depend on the mean or the variance. When the distribution is skewed to the right [$\mathbb{E}\{(X - \mu_X)^3\} > 0$], equation (3.1) shows that an increase in the value of X ($X_1 - X_2 > 0$) is associated with an increase in the absolute value of X [$X_1^2 - X_2^2 > 0$], while a shift to the left would go with a decrease in the absolute value of X. More generally, $\mathbb{E}\{(X - \mu_X)^3\}$ can be written as

$$
\mathbb{E}\{(X - \mu_X)^3\} = \frac{1}{2}\mathbb{E}\{[(X_1 - \mu_X) - (X_2 - \mu_X)][(X_1 - \mu_X)^2 - (X_2 - \mu_X)^2]\} \qquad (3.3)
$$

so that $\mathbb{E}\{(X - \mu_X)^3\}$ measures the association between a change in the deviation from the mean and the corresponding change in the absolute values of the deviations. A positive (negative) association entails positive (negative) skewness, which is clearly exhibited by pairwise differences.

The formulae given by Proposition 3.1 depend on unknown parameters (μ_X and σ_X^2). It is possible to avoid this by using a larger number of replications. For the third moment, only one additional replication of X is needed. If X_1, X_2, X_3 are three i.i.d. replications of X, we see [from (3.1)] that

$$
\begin{aligned}
\mathbb{E}\{(X - \mu_X)^3\} &= \frac{1}{2}\mathbb{E}\{(X_1 - X_2)(X_1^2 - X_2^2)\} - \mathbb{E}\{X_3\}\mathbb{E}\{(X_1 - X_2)^2\} \\
&= \frac{1}{2}\mathbb{E}\{(X_1 - X_2)[(X_1^2 - X_2^2) - 2X_3(X_1 - X_2)]\} \,. \qquad (3.4)
\end{aligned}
$$

For the fourth moment, this can be done with four replications. If X_1, X_2, X_3, X_4 are four i.i.d. replications of X, we have:

$$\begin{aligned}\mathbb{E}\{(X-\mu_X)^4\} =& \frac{1}{2}\mathbb{E}\{(X_1-X_2)(X_1^3-X_2^3)\} - \frac{3}{2}\mathbb{E}\{X_3\}\mathbb{E}\{(X_1-X_2)(X_1^2-X_2^2)\} \\ &+ \frac{3}{2}\mathbb{E}\{X_3\}\mathbb{E}\{X_4\}\mathbb{E}\{(X_1-X_2)^2\} \\ =& \frac{1}{2}\mathbb{E}\{(X_1-X_2)(X_1^3-X_2^3) - 3X_3(X_1-X_2)(X_1^2-X_2^2) + 3X_3X_4(X_1-X_2)^2\}. \quad (3.5)\end{aligned}$$

(3.4) - (3.5) still do not provide D-representations, because X_3 and X_4 are in levels. This is done in the following propositions by using additional replicates of X. A number of alternative expressions are provided in the following proposition.

Proposition 3.2 D-REPRESENTATION OF THIRD CENTRAL MOMENT. *Let X be a random variable with mean μ_X, variance σ_X^2, and finite third moment. If X_1, X_2, X_3 are three i.i.d. replications of X, then*

$$\begin{aligned}\mathbb{E}\{(X-\mu_X)^3\} &= \mathbb{E}\{(X_1-X_3)(X_1-X_2)^2\} = \frac{1}{2}\mathbb{E}\{(X_1-X_2)[(X_1-X_3)^2-(X_2-X_3)^2]\} \\ &= \frac{1}{2}\mathbb{E}\{(X_1-X_2)^2[(X_1-X_3)+(X_2-X_3)]\} = \frac{1}{6}\mathbb{E}\{D_1^{(3)}D_2^{(3)}D_3^{(3)}\} \\ &= \frac{9}{2}\mathbb{E}\{(X_1-\bar{X}^{(3)})(X_2-\bar{X}^{(3)})(X_3-\bar{X}^{(3)})\} \qquad (3.6)\end{aligned}$$

where $\bar{X}^{(3)} := \sum_{j=1}^{3} X_j/3$ *and*

$$D_i^{(3)} := \sum_{j\neq i}(X_i - X_j) = 3(X_i - \bar{X}^{(3)}), \quad i = 1, 2, 3. \qquad (3.7)$$

As the variables $D_i^{(3)}$ are sums of pairwise differences, (3.6) provides simple D-representations of the third central moment of X. Note also that $\mathbb{E}\{(X-\mu_X)^3\}$ can be rewritten as a function of deviations with respect to a "reference" value X_3:

$$\begin{aligned}\mathbb{E}\{(X-\mu_X)^3\} &= \frac{1}{2}\mathbb{E}\{[(X_1-X_3)-(X_2-X_3)\}[(X_1-X_3)^2-(X_2-X_3)^2]\} \\ &= \frac{1}{2}\mathbb{E}\{[(X_1-X_3)-(X_2-X_3)]^2[(X_1-X_3)+(X_2-X_3)]\}\,. \end{aligned} \tag{3.8}$$

We now consider the fourth central moment.

Proposition 3.3 *D*-REPRESENTATION OF FOURTH CENTRAL MOMENT. *Let X be a random variable with mean μ_X, variance σ_X^2, and finite fourth moment. If X_1, X_2, X_3, X_4 four i.i.d. replications of X, then*

$$\begin{aligned}\mathbb{E}\{(X-\mu_X)^4\} &= \frac{1}{2}\mathbb{E}\{(X_1-X_2)^4\}-3\sigma_X^4 = \frac{1}{2}\mathbb{E}\{(X_1-X_2)^4\}-\frac{3}{4}(\mathbb{E}\{(X_1-X_2)^2\})^2 \\ &= \mathbb{E}\big\{\frac{1}{2}(X_1-X_2)^4-\frac{3}{4}(X_1-X_2)^2(X_3-X_4)^2\big\}\,. \end{aligned} \tag{3.9}$$

The last two identities in (3.9) provide simple D-representations of the fourth central moment of X. We now apply the above results to skewness, kurtosis and excess kurtosis coefficients:

$$\mathrm{Sk}(X) := \frac{\mathbb{E}\{(X-\mu_X)^3\}}{\sigma_X^3}\,,\ \ \mathrm{Kur}(X) := \frac{\mathbb{E}\{(X-\mu_X)^4\}}{\sigma_X^4}\,,\ \ \mathrm{EKur}(X) = \mathrm{Kur}(X)-3\,. \tag{3.10}$$

$\mathrm{Kur}(X)$ is also called the "Pearson kurtosis" coefficient, while the excess kurtosis $\mathrm{EKur}(X)$ is the "Fisher kurtosis".

Proposition 3.4 *D*-REPRESENTATION OF SKEWNESS AND KURTOSIS. *Let X be a random variable with mean μ_X and variance σ_X^2, and X_1, X_2, X_3, X_4 four i.i.d. replications of*

X. If X has finite third moment, then

$$\begin{aligned}
\mathrm{Sk}(X) &= \frac{\mathbb{E}\{(X_1-X_2)(X_1^2-X_2^2)\}}{2\sigma_X^3} - 2\frac{\mu_X}{\sigma_X} = \frac{\mathbb{E}\{(X_1-X_3)(X_1-X_2)^2\}}{\sigma_X^3} \\
&= \frac{\sqrt{8}\,\mathbb{E}\{(X_1-X_3)(X_1-X_2)^2\}}{(\mathbb{E}\{(X_1-X_2)^2\})^{3/2}} = \frac{\sqrt{2}}{3}\frac{\mathbb{E}\{D_1^{(3)}D_2^{(3)}D_3^{(3)}\}}{(\mathbb{E}\{(X_1-X_2)^2\})^{3/2}} \qquad (3.11)
\end{aligned}$$

where $D_i^{(3)}$ is defined by (3.7). If X has finite fourth moment, then

$$\mathrm{Kur}(X) = \frac{2\mathbb{E}\{(X_1-X_2)^4\}}{(\mathbb{E}\{(X_1-X_2)^2\})^2} - 3\,. \qquad (3.12)$$

If X_1 and X_2 are i.i.d. normal, it is easy to see that

$$\frac{\mathbb{E}\{(X_1-X_2)^4\}}{(\mathbb{E}\{(X_1-X_2)^2\})^2} = \frac{12\sigma_X^4}{4\sigma_X^4} = 3 \qquad (3.13)$$

so that $\mathrm{EKur}(X) = 0$.

4. Higher-order Moments

In this section, we study how higher-order central moments can be represented in terms of pairwise differences between i.i.d. realizations of a random variable. Consider a random variable X with mean μ_X and finite moments up to order $n \geq 1$. We denote the k-th central moment of X by

$$\mu_k := \mathbb{E}\{(X-\mu_X)^k\}, \quad k = 1,\ldots,n\,. \qquad (4.1)$$

Suppose that $X_0, X_1, \ldots, X_n$ are i.i.d. replications of X, and define

$$P_k := \prod_{i=1}^{k}(X_0 - X_i)\,, \quad \tilde{X}_i := X_i - \mu_X\,, \quad i = 1, \ldots, n\,. \tag{4.2}$$

Since $\mathbb{E}\{\tilde{X}_i\} = \mu_1 = 0$, for $i = 0, \ldots, n$, the independence of $\tilde{X}_0, \ldots, \tilde{X}_n$ entails:

$$\mathbb{E}\{P_k\} = \mathbb{E}\Big\{\prod_{i=1}^{k}(\tilde{X}_0 - \tilde{X}_i)\Big\} = \mathbb{E}\{\tilde{X}_0^k\} = \mu_k\,, \quad k = 1, \ldots, n\,. \tag{4.3}$$

In other words, P_k can be viewed as an unbiased estimator of μ_k, for $k = 1, \ldots, n$. This immediately shows that all the central moments of X have a D-representation for moments up to order n, provided X has moments up to order n.

The function P_n requires $n+1$ replications of X to represent the n-th central moment of X. However, for $n = 2, 3, 4$, we have given representations which only require n replications (see Section 3). We will now show this is indeed feasible for $n \geq 5$. We first state some useful recursions.

Proposition 4.1 RECURSIONS FOR HIGHER-ORDER MOMENTS. *Let X be a random variable with mean μ_X, and finite central moments μ_j for $j = 1, \ldots, n+1$, where $n \geq 1$. Then*

$$\mu_{n+1} = \mathsf{C}[X, (X - \mu_X)^n] = \mathbb{E}\{X(X - \mu_X)^n\} - \mu_X \mu_n\,. \tag{4.4}$$

If X_1, X_2, X_3 are i.i.d. replications of X, then

$$\begin{aligned} \mu_{n+1} &= \mathbb{E}\{(X_1 - X_2)(X_1 - \mu_X)^n\} \\ &= \mathbb{E}\{(X_1 - X_3)(X_1 - X_2)^n\} - \sum_{j=2}^{n-1}(-1)^j C_n^j \mu_j \mu_{n+1-j} \end{aligned} \tag{4.5}$$

where $C_n^j = n!/[j!(n-j)!]$ and, for $n+1$ even,

$$\mu_{n+1} = \frac{1}{2}\big[\mathbb{E}\{(X_1 - X_2)^{n+1}\} - \sum_{j=2}^{n-1}(-1)^j C_{n+1}^j\, \mu_j \mu_{n+1-j}\big]\,. \tag{4.6}$$

When the summations in (4.5) - (4.6) are empty (for $n = 1$ and $n = 2$), the corresponding formulae yield [as expected from (2.6) and (3.6)]:

$$\mu_2 = \mathbb{E}\{(X_1 - X_3)(X_1 - X_2)\} = \frac{1}{2}\mathbb{E}\{(X_1 - X_2)^2\}\,, \quad \mu_3 = \mathbb{E}\{(X_1 - X_3)(X_1 - X_2)^2\}\,. \tag{4.7}$$

Now, let $X_1, \ldots, X_n$ be i.i.d. replications of X with $n \geq 5$, and consider the following sequence of functions:

$$\bar{\mu}_1 = 0\,, \quad \bar{\mu}_2(x_1, x_2) = \frac{1}{2}(x_1 - x_2)^2\,, \tag{4.8}$$

$$\bar{\mu}_3(x_1, x_2, x_3) = (x_1 - x_3)(x_1 - x_2)^2\,, \tag{4.9}$$

$$\bar{\mu}_4(x_1, x_2, x_3, x_4) = \frac{1}{2}(x_1 - x_2)^4 - \frac{3}{4}(x_1 - x_2)^2(x_3 - x_4)^2, \tag{4.10}$$

and for $k \geq 4$,

$$\bar{\mu}_{k+1}(x_1, \ldots, x_{k+1}) = (x_1 - x_3)(x_1 - x_2)^k - \sum_{j=2}^{k-1}(-1)^j C_k^j \bar{\mu}_j(x_1, \ldots, x_j)\bar{\mu}_{k+1-j}(x_{j+1}, \ldots, x_{k+1})\,. \tag{4.11}$$

If $k+1$ is even (with $k \geq 1$), we can also consider the functions

$$\begin{aligned}\tilde{\mu}_{k+1}(x_1, \ldots, x_{k+1}) \quad = \quad & \frac{1}{2}\big[(x_1 - x_2)^{k+1} \\ & - \sum_{j=2}^{k-1}(-1)^j C_{k+1}^j \bar{\mu}_j(x_1, \ldots, x_j)\bar{\mu}_{k+1-j}(x_{j+1}, \ldots, x_{k+1})\big]\,. \end{aligned} \tag{4.12}$$

From the above definitions, it is clear that the functions $\bar{\mu}_k(x_1, \ldots, x_k)$ and $\tilde{\mu}_k(x_1, \ldots, x_k)$ depend on their arguments only through differences $x_i - x_j$ $(i, j = 1, \ldots, k)$. Note also that $\bar{\mu}_j$ can be replaced by $\tilde{\mu}_j$ in the summations of (4.11) - (4.12) whenever j is even.

Proposition 4.2 RECURSIVE *D*-REPRESENTATIONS FOR HIGHER-ORDER MOMENTS. *Let X be a random variable with mean μ_X, and finite moments up to order $n+1$, where $n \geq 1$, and let $\bar{\mu}_k(x_1, \ldots, x_k)$ and $\tilde{\mu}_k(x_1, \ldots, x_k)$ be defined by* (4.8) - (4.12), *for $k \geq 1$. If $X_1, \ldots, X_{n+1}$ are i.i.d. replications of X, then*

$$\mathbb{E}\{\bar{\mu}_{k+1}(X_1, \ldots, X_{k+1})\} = \mu_{k+1}, \quad \textit{for } k = 1, \ldots, n, \tag{4.13}$$

and, if $k+1$ is even,

$$\mathbb{E}\{\tilde{\mu}_{k+1}(X_1, \ldots, X_{k+1})\} = \mu_{k+1}, \quad \textit{for } 1 \leq k \leq n. \tag{4.14}$$

The above proposition shows that the k-th central moment of X has a D-representation which only requires k replications of X (provided μ_k is finite).

5. Generalized Lagrange Identities

In this section, we give some generalizations of the Lagrange identity. We first give a moment-based Lagrange-type identity using independent random vectors.

Proposition 5.1 GENERALIZED LAGRANGE IDENTITY. *Let $(X_1, Y_1)'$ and $(X_2, Y_2)'$ be* *two independent random vectors with finite second moments. Then*

$$\frac{1}{2}\big(\mathbb{E}\{X_1^2\}\mathbb{E}\{Y_2^2\}+\mathbb{E}\{X_2^2\}\mathbb{E}\{Y_1^2\}\big)-\mathbb{E}\{X_1Y_1\}\mathbb{E}\{X_2Y_2\}=\frac{1}{2}\mathbb{E}\{(X_1Y_2-X_2Y_1)^2\}. \quad (5.1)$$

In Proposition 5.1, $(X_1,Y_1)'$ and $(X_2,Y_2)'$ need not be identically distributed. Since the right-hand side of (5.1) is nonnegative, it follows that

$$\mathbb{E}\{X_1Y_1\}\mathbb{E}\{X_2Y_2\}\leq\frac{1}{2}(\mathbb{E}\{X_1^2\}\mathbb{E}\{Y_2^2\}+\mathbb{E}\{X_2^2\}\mathbb{E}\{Y_1^2\}) \quad (5.2)$$

and, on replacing each variable X_1 , X_2 , Y_1 , Y_2 by its deviation from the mean ($X_1-\mathbb{E}\{X_1\}$, etc.),

$$\mathsf{C}[X_1,\,Y_1]\mathsf{C}[X_2,\,Y_2]\leq\frac{1}{2}(\sigma_{X_1}^2\sigma_{Y_2}^2+\sigma_{X_2}^2\sigma_{Y_1}^2)\,. \quad (5.3)$$

The latter inequality holds even if the random variables X_1 , X_2 , Y_1 , Y_2 have different means and variances. When $(X_1,Y_1)'$ and $(X_2,Y_2)'$ have the same covariance matrix (though possibly different means), (5.2) yields the usual Cauchy-Schwarz inequality for covariances:

$$\mathsf{C}[X_1,\,Y_1]^2\leq\sigma_{X_1}^2\sigma_{Y_1}^2\,. \quad (5.4)$$

Interestingly, (5.1) provides a remarkably simple way of proving the Cauchy-Schwarz inequality for covariances.

The distance $\mathbb{E}\{(X_1Y_2-X_2Y_1)^2\}$ can be viewed as a measure of non-proportionality between Y and X : if $Y_1=b_1X_1$ and $Y_2=b_2X_2$, then

$$\mathbb{E}\{(X_1Y_2-X_2Y_1)^2\}=\mathbb{E}\{(b_2X_1X_2-b_1X_2X_1)^2\}=(b_2-b_1)^2\mathbb{E}\{(X_1X_2)^2\} \quad (5.5)$$

with $\mathbb{E}\{(X_1Y_2 - X_2Y_1)^2\} = 0$ when $b_1 = b_2$. When $\mathbb{E}\{(X_1X_2)^2\} \neq 0$, the condition $b_1 = b_2$ is necessary and sufficient for $\mathbb{E}\{(X_1Y_2 - X_2Y_1)^2\} = 0$. The identity (5.5) holds even if X_1 and X_2 do not have the same distribution.

When additional homogeneity restrictions are imposed, we get simpler identities which are spelled out in the following corollary.

Corollary 5.2 COVARIANCE GENERALIZED LAGRANGE IDENTITY. *Let $(X_1, Y_1)'$ and $(X_2, Y_2)'$ be two independent random vectors with finite second moments, $\tilde{X}_i := X_i - \mathbb{E}\{X_i\}$ and $\tilde{Y}_i := Y_i - \mathbb{E}\{Y_i\}$, $i = 1, 2$. Then*

$$\frac{1}{2}(\sigma^2_{X_1}\sigma^2_{Y_2} + \sigma^2_{X_2}\sigma^2_{Y_1}) - \mathsf{C}[X_1, Y_1]\mathsf{C}[X_2, Y_2] = \frac{1}{2}\mathbb{E}\{(\tilde{X}_1\tilde{Y}_2 - \tilde{X}_2\tilde{Y}_1)^2\}\,. \tag{5.6}$$

If $\sigma^2_{X_1} = \sigma^2_{X_2}$ and $\sigma^2_{Y_1} = \sigma^2_{Y_2}$, then

$$\sigma^2_{X_1}\sigma^2_{Y_1} - \mathsf{C}[X_1, Y_1]\mathsf{C}[X_2, Y_2] = \frac{1}{2}\mathbb{E}\{(\tilde{X}_1\tilde{Y}_2 - \tilde{X}_2\tilde{Y}_1)^2\}\,. \tag{5.7}$$

If the two random vectors $(X_1, Y_1)'$ and $(X_2, Y_2)'$ have the same covariance matrix, then

$$\sigma^2_{X_1}\sigma^2_{Y_1} - \mathsf{C}[X_1, Y_1]^2 = \frac{1}{2}\mathbb{E}\{(\tilde{X}_1\tilde{Y}_2 - \tilde{X}_2\tilde{Y}_1)^2\}\,. \tag{5.8}$$

The identity (5.8) holds *a fortiori* when $(X_1, Y_1)'$ and $(X_2, Y_2)'$ are i.i.d. (with finite second moments). When $(X_1, Y_1)'$ and $(X_2, Y_2)'$ have the same covariance matrix, the correlation $\rho(X_1, Y_1)$ between X_1 and Y_1 satisfies:

$$\rho(X_1, Y_1)^2 = \rho(X_2, Y_2)^2 = 1 - \frac{\mathbb{E}\{(\tilde{X}_1\tilde{Y}_2 - \tilde{X}_2\tilde{Y}_1)^2\}}{2\sigma^2_{X_1}\sigma^2_{Y_1}}\,. \tag{5.9}$$

Thus, if $(X_1, Y_1)'$ and $(X_2, Y_2)'$ are i.i.d. replications of the random vector $(X, Y)'$, $\rho(X_1, Y_1)^2$ measures how close two independent replications of $(X, Y)'$ are according to the distance $\mathbb{E}\{(\tilde{X}_1\tilde{Y}_2 - \tilde{X}_2\tilde{Y}_1)^2\}$.

The following proposition extends in a similar way the Binet-Cauchy identity (1.7).

Proposition 5.3 GENERALIZED BINET-CAUCHY IDENTITY.
Let $Z_i := (A_i, B_i, C_i, D_i)'$ and $\tilde{Z}_i := Z_i - \mathbb{E}\{Z_i\} = (\tilde{A}_i, \tilde{B}_i, \tilde{C}_i, \tilde{D}_i)'$, $i = 1, 2$, be independent random vectors with finite fourth moments. Then

$$\begin{aligned} \mathbb{E}\{(A_1B_2 - A_2B_1)(C_1D_2 - C_2D_1)\} &= \mathbb{E}\{A_1C_1\}\mathbb{E}\{B_2D_2\} + \mathbb{E}\{A_2C_2\}\mathbb{E}\{B_1D_1\} \\ &- [\mathbb{E}\{A_1D_1\}\mathbb{E}\{B_2C_2\} + \mathbb{E}\{A_2D_2\}\mathbb{E}\{B_1C_1\}] \,. \quad (5.10) \end{aligned}$$

If $(A_1, B_1, C_1, D_1)'$ and $(A_2, B_2, C_2, D_2)'$ have the same covariance matrix, then

$$\mathsf{C}[A_1, C_1]\mathsf{C}[B_1, D_1] - \mathsf{C}[A_1, D_1]\mathsf{C}[B_1, C_1] = \frac{1}{2}\mathbb{E}\{(\tilde{A}_1\tilde{B}_2 - \tilde{A}_2\tilde{B}_1)(\tilde{C}_1\tilde{D}_2 - \tilde{C}_2\tilde{D}_1)\} \,. \quad (5.11)$$

The above proposition yields restrictions on the covariances of a four-dimensional random vector. From (5.11), we see that $A_1B_2 - A_2B_1 = 0$ (a.s.) or $C_1D_2 - C_2D_1 = 0$ (a.s.) entails

$$\mathsf{C}[A_1, C_1]\mathsf{C}[B_1, D_1] = \mathsf{C}[A_1, D_1]\mathsf{C}[B_1, C_1] \,. \quad (5.12)$$

On taking $A_i = C_i$ and $B_i = D_i$ for $i = 1, 2$, we get the Lagrange identity:

$$\sigma_{A_1}^2\sigma_{B_1}^2 - (\mathsf{C}[A_1, B_1])^2 = \frac{1}{2}\mathbb{E}\{(\tilde{A}_1\tilde{B}_2 - \tilde{A}_2\tilde{B}_1)^2\} \,. \quad (5.13)$$

6. Applications

The D-representations given above show that the central moments of a real random variable X can all be represented as expected values of polynomial functions in differences between a small number of replications of X. For $k \geq 2$, the k-th central moment of X only requires k replications $X_1, \ldots, X_k$. More precisely, we can write:

$$\mu_k := \mathbb{E}\{(X - \mu_X)^k\} = \mathbb{E}\{h_k(X_1, \ldots, X_k)\}, \quad \text{for } k \geq 2, \tag{6.1}$$

where

$$h_2(X_1, X_2) = \frac{1}{2}(X_1 - X_2)^2, \tag{6.2}$$

$$h_3(X_1, X_2, X_3) = \frac{1}{2}(X_1 - X_2)^2[(X_1 - X_3) + (X_2 - X_3)], \tag{6.3}$$

$$h_4(X_1, X_2, X_3, X_4) = \frac{1}{2}(X_1 - X_2)^4 - \frac{3}{4}(X_1 - X_2)^2(X_3 - X_4)^2, \tag{6.4}$$

$$\begin{aligned} h_{k+1}(X_1, \ldots, X_{k+1}) &= (X_1 - X_3)(X_1 - X_2)^k \\ &\quad - \sum_{j=2}^{k-1} (-1)^j C_k^j h_j(X_1, \ldots, X_j) h_{k+1-j}(X_{j+1}, \ldots, X_{k+1}), \quad k \geq 4. \end{aligned} \tag{6.5}$$

In other words, $h_k(X_1, \ldots, X_k)$ is an unbiased estimator of μ_k based on k observations.

If n observations $X_1, \ldots, X_n$ are available, any (fixed or randomly selected) subset of k observations from $\{X_1, \ldots, X_n\}$ yields an unbiased estimator of μ_k, and similarly for the arithmetic (or a weighted) average of such subsets. In particular, the arithmetic average of $h_k(X_{i_1}, \ldots, X_{i_k})$ over N randomly selected $\{i_1, \ldots, i_k\}$ of $\{1, \ldots, n\}$ is an unbiased estimator of μ_k. *A fortiori*, this will also hold if the average is taken over all subsets of k observations from $\{X_1, \ldots, X_n\}$, but the large number of such subsets can easily make this approach computationally prohibitive. We will call such estimators based on observation differences D-estimators.

It is of interest to compare D-estimators of μ_k with the "natural" estimators based on deviations from the sample mean:

$$\hat{m}_k(n) = \frac{1}{n-1}\sum_{i=i}^{n}(X_i - \bar{X}_n)^k\,, \quad \bar{X}_n = \frac{1}{n}\sum_{i=i}^{n} X_i\,. \tag{6.6}$$

For $k = 2$, $\hat{m}_k(n)$ is unbiased, but it is not generally unbiased for $k \geq 3$. To see how big the bias difference can be, we take $n = k$ and compare by simulation $\hat{\mu}_k(k)$ with the corresponding D-estimator with $h_k(X_1, \ldots, X_k)$; see Table 1. We use the Exponential distribution as an example because its central moments are nonzero at all orders, which helps us assess estimator bias without interference from sign-canceling or offsetting effects that occur when the true moment is zero. We see from these results that $\hat{m}_k(k)$ is heavily biased, while the corresponding D-estimators exhibit no bias as predicted by the results presented above.

Table 1. Simulation Comparison of Bias for Exponential(2) when n=k

Central Moment **Number of Observations**	**Ord. 2** **2**	**Ord. 3** **3**	**Ord. 4** **4**	**Ord. 5** **5**	**Ord. 6** **6**	**Ord. 7** **7**	**Ord. 8** **8**
True Value	*0.25*	*0.25*	*0.56*	*1.38*	*4.00*	*14.48*	*57.94*
Natural Estimator $\hat{m}_k(k)$	**0.000**	-0.167	-0.258	-0.768	-2.171	-8.321	-33.739
D-estimators $\hat{\mu}_k(k)$	**0.000**	**0.000**	**0.000**	**0.003**	**0.094**	**0.131**	**0.053**

* *Number of Simulation = 20000000*

Of course, if $n > k$, more efficient unbiased estimators can be obtained by averaging several minimal estimators. This can be easily implemented through computer programming by selecting the index tuples $(i_1, \ldots, i_{k+1})$ independently and identically distributed (i.i.d.) from a uniform distribution over the index set I. Due to its stochastic nature, we refer to this estimator as a Monte Carlo approximation of the D-estimator, $\hat{\mu}_k(n)^{MC}$. Table 2 presents simulation results comparing the bias of the "natural" estimators $\hat{m}_k(n)$ with the proposed D-estimator Monte Carlo approximation for central moments of orders $k = 2, \ldots, 8$ under an Exponential(2) distribution. For each replication, we generate 30,000 Monte Carlo samples.

Table 2. Simulation Comparison of Bias for Exponential(2) when n>k (# of Simulation = 1000)

	Central Moment	Ord.2	Ord.3	Ord.4	Ord.5	Ord.6	Ord.7	Ord.8
	True Value	*0.25*	*0.25*	*0.56*	*1.38*	*4.00*	*14.48*	*57.94*
n=50	**Natural Estimator** $\hat{m}_k(n)$	-0.004	-0.018	-0.059	-0.240	-0.921	-5.203	-27.483
	D-Estimator MC $\hat{\mu}_k(n)^{MC}$	**-0.004**	**-0.009**	**-0.035**	**-0.162**	**-0.592**	**-4.048**	**-22.484**
n=100	**Natural Estimator** $\hat{m}_k(n)$	**-0.001**	-0.008	-0.020	-0.084	-0.295	-2.621	-16.574
	D-Estimator MC $\hat{\mu}_k(n)^{MC}$	-0.001	**-0.003**	**-0.005**	**-0.035**	**-0.133**	**-1.991**	**-13.020**

** Number of Simulation = 1000*

The simulation results demonstrate that the proposed D-estimator Monte Carlo approximation substantially reduces bias across all orders of the central moment compared to the natural estimator. This improvement is particularly notable for higher-order moments where the naive estimator exhibits larger negative bias. Moreover, increasing the sample size from 50 to 100 decreases the bias for both estimators, but the D-estimator approach consistently outperforms the natural estimator at each sample size. These findings highlight the effectiveness of the Monte Carlo approximation in providing more accurate and unbiased estimates of central moments, especially for higher-order moments that are typically more challenging to estimate.

From (3.6), one can see that several choices of minimal estimator h_3 are possible (and similarly for higher-order moments), which may have different efficiency properties. The above results can also be used to build unbiased estimators of cumulants as well as to analyze which observations deviate from the rest of the sample. The latter application can be based on analyzing the population of subgroups of observations $(X_{i_1}, \ldots, X_{i_k})$ along with the corresponding unbiased estimates $h_k(X_{i_1}, \ldots, X_{i_k})$. Discussing in detail variants of D-estimators, as well as other applications, goes beyond the scope of this paper.

7. Conclusion

In this paper, following an approach (apparently) initiated by Gini (1912) for the variance, we have studied the general problem of representing the central moments of a random variable in terms of pairwise differences between i.i.d. replications, without reference to a location parameter. We first gave pairwise-difference representations (defined formally as *D-representations*) for the third and fourth central moments of a general random variable. These yield intuitive interpretations of the familiar skewness and kurtosis coefficients. For general moments, we have observed that central moments can be interpreted as covariances between a random variable and powers of the same variable (a *nonlinear* relation). These then provide *recursions* which link the D-representation of any moment to lower order ones, hence D-representations for all central moments (as long as they are finite). Through a similar approach, analogues of Lagrange and Binet-Cauchy identities were established for general random variables. These provide a simple derivation of the classic Cauchy-Schwarz inequality for covariances. Finally, it is of interest to note that the formulas given in this paper can be interpreted as applications of the "coupling" method [see Thorisson (2000)], which opens the way to additional applications.

Acknowledgements

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).

References

Dufour, J.-M. (1984), 'Unbiasedness of predictions from estimated autoregressions when the true order is unknown', *Econometrica* **52**(1), 209–215.

Dufour, J.-M. (1985), 'Unbiasedness of predictions from estimated vector autoregressions', *Econometric Theory* **1**(1), 387–402.

Gini, C. W. (1912), 'Variabilita e mutabilita', *Studi Economico-Giuridici Della R. Universita di Cagliari* **3**(2), 1–158.

Hayes, K. (2011), 'A geometrical interpretation of an alternative formula for the sample covariance', *The American Statistician* **65**(2), 110–112.

Heffernan, P. M. (1988), 'New measures of spread and a simpler formula for the normal distribution', *The American Statistician* **42**(2), 100–102.

Lee, A. J. (1990), *U-Statistics: Theory and Practice*, Marcel Dekker, New York.

Lehmann, E. L. (1966), 'Some concepts of dependence', *The Annals of Mathematical Statistics* **37**, 1137–1153.

Light, R. J. and Margolin, B. H. (1971), 'An analysis of variance for categorical data', *Journal of the American Statistical Association* **66**(335), 534–544.

Light, R. J. and Margolin, B. H. (1974), 'An analysis of variance for categorical data, II: Small sample comparisons with chi square and other competitors', *Journal of the American Statistical Association* **69**(347), 755–764.

Steele, J. M. (2004), *The Cauchy-Schwarz Master Class: An Introduction to the Art of Mathematical Inequalities*, Cambridge University Press, Cambridge, U.K.

Thorisson, H. (2000), *Coupling, Stationarity, and Regeneration*, Springer-Verlag, New York.

Wright, T. (1992), 'Lagrange's identity reveals correlation coefficient and straight-line connection', *The American Statistician* **46**(2), 106–107.

Yitzhaki, S. and Schechtman, E. (2013), *The Gini Methodology: A Primer on a Statistical Methodology*, Springer, New York.

A. Appendix: Proofs

PROOF OF PROPOSITION 3.1 The third central moment can be written as follows:

$$\begin{aligned}\mathbb{E}\{(X-\mu_X)^3\} &= \mathbb{E}\{(X-\mu_X)(X-\mu_X)^2\} = \mathsf{C}[X,\,(X-\mu_X)^2] \\ &= \mathsf{C}[X,\,(X^2-2\mu_X X+\mu_X^2)] = \mathsf{C}[X,\,(X^2-2\mu_X X)]\,. \qquad \text{(A.1)}\end{aligned}$$

Using (2.2) and (2.5), we can write:

$$\begin{aligned}\mathsf{C}[X,\,(X-\mu_X)^2] &= \frac{1}{2}\mathbb{E}\{(X_1-X_2)[(X_1-\mu_X)^2-(X_2-\mu_X)^2]\} \\ &= \frac{1}{2}\mathbb{E}\{(X_1-X_2)[(X_1^2-2\mu_X X_1)-(X_2^2-2\mu_X X_2)]\} \\ &= \frac{1}{2}\mathbb{E}\{(X_1-X_2)[(X_1^2-X_2^2)-2\mu_X(X_1-X_2)]\} \\ &= \frac{1}{2}\mathbb{E}\{(X_1-X_2)(X_1^2-X_2^2)\}-\mu_X\mathbb{E}\{(X_1-X_2)^2\} \\ &= \frac{1}{2}\mathbb{E}\{(X_1-X_2)(X_1^2-X_2^2)\}-2\mu_X\sigma_X^2\,. \qquad \text{(A.2)}\end{aligned}$$

(3.1) follows on combining (A.1) and (A.2). To show (3.2), we proceed similarly. We have:

$$\begin{aligned}\mathbb{E}\{(X-\mu_X)^4\} &= \mathbb{E}\{(X-\mu_X)(X-\mu_X)^3\} = \mathsf{C}[X,\,(X-\mu_X)^3] \\ &= \mathsf{C}[X,\,(X^3-3X^2\mu_X+3X\mu_X^2-\mu_X^3)] \\ &= \mathsf{C}[X,\,(X^3-3X^2\mu_X+3X\mu_X^2)] \qquad \text{(A.3)}\end{aligned}$$

hence, by (2.2),

$$\mathsf{C}[X,\,(X-\mu_X)^3] = \frac{1}{2}\mathbb{E}\{(X_1-X_2)[(X_1-\mu_X)^3-(X_2-\mu_X)^3]\}$$

$$
\begin{aligned}
&= \frac{1}{2}\mathbb{E}\{(X_1-X_2)[(X_1^3-3X_1^2\mu_X+3X_1\mu_X^2-\mu_X^3)-(X_2^3-3X_2^2\mu_X+3X_2\mu_X^2-\mu_X^3)]\} \\
&= \frac{1}{2}\mathbb{E}\{(X_1-X_2)[(X_1^3-X_2^3)-3\mu_X(X_1^2-X_2^2)+3\mu_X^2(X_1-X_2)]\} \\
&= \frac{1}{2}\mathbb{E}\{(X_1-X_2)(X_1^3-X_2^3)\}-\frac{3}{2}\mu_X\mathbb{E}\{(X_1-X_2)(X_1^2-X_2^2)\}+\frac{3}{2}\mu_X^2\mathbb{E}\{(X_1-X_2)^2\} \\
&= \frac{1}{2}\mathbb{E}\{(X_1-X_2)(X_1^3-X_2^3)\}-3\mu_X\left[\mathbb{E}\{(X-\mu_X)^3\}+\mu_X\mathbb{E}\{(X_1-X_2)^2\}\right] \\
&\quad +\frac{3}{2}\mu_X^2\mathbb{E}\{(X_1-X_2)^2\} \\
&= \frac{1}{2}\mathbb{E}\{(X_1-X_2)(X_1^3-X_2^3)\}-3\mu_X\mathbb{E}\{(X-\mu_X)^3\}-\frac{3}{2}\mu_X^2\mathbb{E}\{(X_1-X_2)^2\} \\
&= \frac{1}{2}\mathbb{E}\{(X_1-X_2)(X_1^3-X_2^3)\}-3\mu_X\mathbb{E}\{(X-\mu_X)^3\}-3\mu_X^2\sigma_X^2\,. \qquad \text{(A.4)}
\end{aligned}
$$

(3.2) follows on combining (A.3) and (A.4). □

PROOF OF PROPOSITION 3.2 Since X_1, X_2, X_3 are i.i.d. replications of X, we have:

$$
\begin{aligned}
\mathbb{E}\{(X_1-X_2)[(X_1-X_3)^2-(X_2-X_3)^2]\} &= \mathbb{E}\{(X_1-X_2)[(X_1^2-X_2^2)-2X_3(X_1-X_2)]\} \\
&= \mathbb{E}\{(X_1-X_2)(X_1^2-X_2^2)\}-2\mathbb{E}\{X_3\}\mathbb{E}\{(X_1-X_2)^2\} \\
&= \mathbb{E}\{(X_1-X_2)(X_1^2-X_2^2)\}-2\mu_X\mathbb{E}\{(X_1-X_2)^2\} \qquad \text{(A.5)}
\end{aligned}
$$

and, using (3.1),

$$
\begin{aligned}
\mathbb{E}\{(X-\mu_X)^3\} &= \frac{1}{2}\mathbb{E}\{(X_1-X_2)(X_1^2-X_2^2)\}-\mu_X\mathbb{E}\{(X_1-X_2)^2\} \\
&= \frac{1}{2}\mathbb{E}\{(X_1-X_2)[(X_1-X_3)^2-(X_2-X_3)^2]\} \\
&= \frac{1}{2}\{\mathbb{E}\{(X_1-X_2)(X_1^2-X_2^2)-2X_3(X_1-X_2)^2]\} \\
&= \frac{1}{2}\mathbb{E}\{(X_1-X_2)^2[(X_1+X_2)-2X_3]\}
\end{aligned}
$$

$$= \frac{1}{2}\mathbb{E}\{(X_1 - X_2)^2[(X_1 - X_3) + (X_2 - X_3)]\}\,. \tag{A.6}$$

Now, set $\tilde{X}_i = X_i - \mu_X$, $i = 1, 2, 3$, so that $\mathbb{E}\{\tilde{X}_i\} = 0$ for all i. Using the mutual independence of $\tilde{X}_1$, $\tilde{X}_2$, $\tilde{X}_3$, we can then write:

$$\begin{aligned} \mathbb{E}\{(X_1 - X_3)(X_1 - X_2)^2\} &= \mathbb{E}\{(\tilde{X}_1 - \tilde{X}_3)(\tilde{X}_1 - \tilde{X}_2)^2\} = \mathbb{E}\{(\tilde{X}_1 - \tilde{X}_3)(\tilde{X}_1^2 - 2\tilde{X}_1\tilde{X}_2 + \tilde{X}_2^2)\} \\ &= \mathbb{E}\{\tilde{X}_1^3\} = \mathbb{E}\{(X - \mu_X)^3\}\,. \end{aligned} \tag{A.7}$$

This establishes the first three identities of (3.6). Further,

$$D_i^{(3)} = \sum_{j=1}^{3} D_{ij} = \sum_{j \neq i} (\tilde{X}_i - \tilde{X}_j) = 2\tilde{X}_i - \sum_{j \neq i} \tilde{X}_j\,, \quad i = 1, 2, 3, \tag{A.8}$$

where $\mathbb{E}\{\tilde{X}_i\tilde{X}_j X_k\} = 0$ unless $i = j = k$, hence

$$\begin{aligned} \mathbb{E}\{D_1^{(3)} D_2^{(3)} D_3^{(3)}\} &= \mathbb{E}\{(2\tilde{X}_1 - \tilde{X}_2 - \tilde{X}_3)(2\tilde{X}_2 - \tilde{X}_1 - \tilde{X}_3)(2\tilde{X}_3 - \tilde{X}_1 - \tilde{X}_2)\} \\ &= 2\mathbb{E}\{\tilde{X}_1^3\} + 2\mathbb{E}\{\tilde{X}_2^3\} + 2\mathbb{E}\{\tilde{X}_3^3\} = 6\mathbb{E}\{(X - \mu_X)^3\}\,. \end{aligned} \tag{A.9}$$

The last identity of (3.6) follows on observing that

$$X_i - \bar{X}^{(3)} = \frac{1}{3}(3X_i - \sum_{j=1}^{3} X_j) = \frac{1}{3}\sum_{j \neq i}(X_i - X_j) = \frac{1}{3}D_i^{(3)}. \tag{A.10}$$

□

PROOF OF PROPOSITION 3.3 Set $\tilde{X}_i = X_i - \mu_X$, $i = 1, \ldots, 4$, and consider the binomial expansion:

$$
\begin{aligned}
(X_1 - X_2)^4 &= (\tilde{X}_1 - \tilde{X}_2)^4 = \sum_{j=0}^{4} (-1)^j C_4^j \tilde{X}_1^{4-j} \tilde{X}_2^j \\
&= (\tilde{X}_1^4 + \tilde{X}_2^4) - 4(\tilde{X}_1^3 \tilde{X}_2 + \tilde{X}_1 \tilde{X}_2^3) + 6\tilde{X}_1^2 \tilde{X}_2^2 . \qquad \text{(A.11)}
\end{aligned}
$$

Taking the expected value on both sides of the above equation, we get:

$$
\begin{aligned}
\mathbb{E}\{(X_1 - X_2)^4\} &= 2\mathbb{E}\{(X - \mu_X)^4\} - 4(\mathbb{E}\{\tilde{X}_1^3\}\mathbb{E}\{\tilde{X}_2\} + \mathbb{E}\{\tilde{X}_1\}\mathbb{E}\{\tilde{X}_2^3\}) + 6\mathbb{E}\{\tilde{X}_1^2\}\mathbb{E}\{\tilde{X}_2^2\} \\
&= 2\mathbb{E}\{(X - \mu_X)^4\} + 6\sigma_X^4 = 2\mathbb{E}([(X - \mu_X)^4\} + \frac{3}{2}(\mathbb{E}\{(X_1 - X_2)^2\})^2 \\
&= 2\mathbb{E}\{(X - \mu_X)^4\} + \frac{3}{2}\mathbb{E}\{(X_1 - X_2)^2\}\mathbb{E}\{(X_3 - X_4)^2\} \\
&= 2\mathbb{E}\{(X - \mu_X)^4\} + \frac{3}{2}\mathbb{E}\{(X_1 - X_2)^2(X_3 - X_4)^2\} \qquad \text{(A.12)}
\end{aligned}
$$

hence

$$
\begin{aligned}
\mathbb{E}\{(X - \mu_X)^4\} &= \frac{1}{2}\mathbb{E}\{(X_1 - X_2)^4\} - 3\sigma_X^4 \\
&= \frac{1}{2}\mathbb{E}\{(X_1 - X_2)^4\} - \frac{3}{4}(\mathbb{E}\{(X_1 - X_2)^2\})^2 \\
&= \frac{1}{2}\mathbb{E}\{(X_1 - X_2)^4\} - \frac{3}{4}\mathbb{E}\{(X_1 - X_2)^2(X_3 - X_4)^2\} \\
&= \frac{1}{2}\mathbb{E}\{(X_1 - X_2)^4 - \frac{3}{2}(X_1 - X_2)^2(X_3 - X_4)^2\} . \qquad \text{(A.13)}
\end{aligned}
$$

This completes the proof of (3.9). □

PROOF OF PROPOSITION 3.4 If X has finite third moment, we get from Propositions

3.1 and 3.2:

$$
\begin{aligned}
\mathrm{Sk}(X) &= \frac{\mathbb{E}\{(X-\mu_X)^3\}}{\sigma_X^3} = \frac{(\mathbb{E}\{(X_1-X_2)(X_1^2-X_2^2)\}/2)-2\mu_X\sigma_X^2}{\sigma_X^3} \\
&= \frac{\mathbb{E}\{(X_1-X_3)(X_1-X_2)^2\}}{\sigma_X^3} = \frac{\mathbb{E}\{(X_1-X_3)(X_1-X_2)^2\}}{(\sigma_X^2)^{3/2}} \\
&= \frac{\mathbb{E}\{(X_1-X_3)(X_1-X_2)^2\}}{(\mathbb{E}\{(X_1-X_2)^2\}/2)^{3/2}} = \frac{\sqrt{8}\mathbb{E}\{(X_1-X_3)(X_1-X_2)^2\}}{(\mathbb{E}\{(X_1-X_2)^2\})^{3/2}} \\
&= \frac{\sqrt{8}}{6}\frac{\mathbb{E}\{D_1^{(3)}D_2^{(3)}D_3^{(3)}\}}{(\mathbb{E}\{(X_1-X_2)^2\})^{3/2}} = \frac{\sqrt{2}}{3}\frac{\mathbb{E}\{D_1^{(3)}D_2^{(3)}D_3^{(3)}\}}{(\mathbb{E}\{(X_1-X_2)^2\})^{3/2}}\,.
\end{aligned}
\tag{A.14}
$$

If X has finite fourth moment, we get from Propositions 3.1 and 3.3:

$$
\begin{aligned}
\mathrm{Kur}(X) &= \frac{\mathbb{E}\{(X-\mu_X)^4\}}{\sigma_X^4} \\
&= \frac{\mathbb{E}\{(X_1-X_2)^4\}-6\sigma_X^4}{2\sigma_X^4} = \frac{\mathbb{E}\{(X_1-X_2)^4\}}{2\sigma_X^4}-3 \\
&= \frac{\mathbb{E}\{(X_1-X_2)^4\}}{2(\mathbb{E}\{(X_1-X_2)^2\}/2)^2}-3 = \frac{2\mathbb{E}\{(X_1-X_2)^4\}}{(\mathbb{E}\{(X_1-X_2)^2\})^2}-3\,.
\end{aligned}
\tag{A.15}
$$

□

PROOF OF PROPOSITION 4.1 By definition, for $n \geq 1$,

$$
\begin{aligned}
\mu_{n+1} &= \mathbb{E}\{(X-\mu_X)(X-\mu_X)^n\} = \mathsf{C}[X,\,(X-\mu_X)^n] \\
&= \mathbb{E}\{X(X-\mu_X)^n\} - \mu_X\mu_n
\end{aligned}
\tag{A.16}
$$

which yields (4.4). Let X_1, X_2, X_3 be i.i.d. replications of X. Then, the above identity entails:

$$\begin{aligned}
\mu_{n+1} &= \mathbb{E}\{X_1(X_1-\mu_X)^n\} - \mathbb{E}\{X_2\}\mathbb{E}\{(X_1-\mu_X)^n\} \\
&= \mathbb{E}\{X_1(X_1-\mu_X)^n\} - \mathbb{E}\{X_2(X_1-\mu_X)^n\} \\
&= \mathbb{E}\{(X_1-X_2)(X_1-\mu_X)^n\}\,. \qquad \text{(A.17)}
\end{aligned}$$

Setting $\tilde{X}_j := X_j - \mu_X\,,\ j = 1, 2\,, 3$, and using the binomial theorem, we can write:

$$\begin{aligned}
(X_1-X_2)^n &= (\tilde{X}_1-\tilde{X}_2)^n = \sum_{j=0}^{n} C_n^j \tilde{X}_1^{n-j}(-\tilde{X}_2)^j \\
&= \sum_{j=0}^{n}(-1)^j C_n^j \tilde{X}_1^{n-j}\tilde{X}_2^j = \tilde{X}_1^n + (-1)^n\tilde{X}_2^n + \sum_{j=1}^{n-1}(-1)^j C_n^j \tilde{X}_1^{n-j}\tilde{X}_2^n \qquad \text{(A.18)}
\end{aligned}$$

and, multiplying the above equation by (X_1-X_3),

$$\begin{aligned}
&(X_1-X_3)(X_1-X_2)^n = (\tilde{X}_1-\tilde{X}_3)(\tilde{X}_1-\tilde{X}_2)^n \\
&= (\tilde{X}_1-\tilde{X}_3)[\tilde{X}_1^n + (-1)^n\tilde{X}_2^n] + (\tilde{X}_1-\tilde{X}_3)\sum_{j=1}^{n-1}(-1)^j C_n^j \tilde{X}_1^{n-j}\tilde{X}_2^j\,. \qquad \text{(A.19)}
\end{aligned}$$

Taking the expected value on both sides of (A.19) and using the fact that $\mathbb{E}\{\tilde{X}_3\} = 0$ and $\mathbb{E}\{\tilde{X}_1^k\} = \mathbb{E}\{\tilde{X}_2^k\} = \mu_k$, we get:

$$\begin{aligned}
\mathbb{E}\{(X_1-X_3)(X_1-X_2)^n\} &= \mathbb{E}\{\tilde{X}_1^{n+1}\} + \sum_{j=1}^{n-1}(-1)^j C_n^j \mathbb{E}\{\tilde{X}_1^{n+1-j}\}\mathbb{E}\{\tilde{X}_2^j\} \\
&= \mu_{n+1} + \sum_{j=1}^{n-1}(-1)^j C_n^j \mu_{n+1-j}\mu_j \qquad \text{(A.20)}
\end{aligned}$$

hence, noting that $\mu_1 = 0$,

$$\mu_{n+1} = \mathbb{E}\{(X_1 - X_3)(X_1 - X_2)^n\} - \sum_{j=2}^{n-1} (-1)^j C_n^j \mu_j \mu_{n+1-j}\,. \tag{A.21}$$

This establishes (4.5). When n is even, we can take the expected value on both sides of (A.18):

$$\begin{aligned}\mathbb{E}\{(X_1 - X_2)^n\} &= \mathbb{E}\{\tilde{X}_1^n + (-1)^n \tilde{X}_2^n\} + \sum_{j=1}^{n-1} (-1)^j C_n^j \mathbb{E}\{\tilde{X}_1^{n-j}\}\mathbb{E}\{\tilde{X}_2^j\} \\ &= 2\mu_n + \sum_{j=1}^{n-1} (-1)^j C_n^j \mu_j \mu_{n-j} = 2\mu_n + \sum_{j=2}^{n-2} (-1)^j C_n^j \mu_j \mu_{n-j}\end{aligned} \tag{A.22}$$

hence

$$\mu_n = \frac{1}{2}\Big[\mathbb{E}\{(X_1 - X_2)^n\} - \sum_{j=2}^{n-2} (-1)^j C_n^j \mu_j \mu_{n-j}\Big] \tag{A.23}$$

which yields (4.6). □

PROOF OF PROPOSITION 4.2 To show (4.13), we first observe that the result holds for $k = 1$, 2, 3 [by (2.5) and Propositions 3.2 - 3.4]. We can then proceed by mathematical induction. Suppose that (4.13) holds for values of k such that $1 \le k < n$, *i.e.*

$$\mathbb{E}\{\bar{\mu}_k(X_1, \ldots, X_k)\} = \mu_k\,, \text{ for } 1 \le k < n\,. \tag{A.24}$$

For $k \ge 4$, we replace $(x_1, \ldots, x_{k+1})$ by $(X_1, \ldots, X_{k+1})$ in (4.11) and we take the expected value:

$$
\begin{aligned}
\mathbb{E}\{\bar{\mu}_{k+1}(X_1,\ldots,X_{k+1})\} &= \mathbb{E}\{(X_1-X_3)(X_1-X_2)^k\} \\
&\quad -\sum_{j=2}^{k-1}(-1)^j C_k^j \mathbb{E}\{\bar{\mu}_j(X_1,\ldots,X_j)\bar{\mu}_{k+1-j}(X_{j+1},\ldots,X_{k+1})\} \\
&= \mathbb{E}\{(X_1-X_3)(X_1-X_2)^k\} \\
&\quad -\sum_{j=2}^{k-1}(-1)^j C_k^j \mathbb{E}\{\bar{\mu}_j(X_1,\ldots,X_j)\}\mathbb{E}\{\bar{\mu}_{k+1-j}(X_{j+1},\ldots,X_{k+1})\} \\
&= \mathbb{E}\{(X_1-X_3)(X_1-X_2)^k\} - \sum_{j=2}^{k-1}(-1)^j C_k^j \mu_j \mu_{k+1-j} \qquad \text{(A.25)}
\end{aligned}
$$

where the last identity follows from the independence of $(X_1,\ldots,X_j)$ and $(X_{j+1},\ldots,X_{k+1})$. By (4.5), this implies

$$
\mathbb{E}\{\bar{\mu}_{k+1}(X_1,\ldots,X_{k+1})\} = \mu_{k+1}\,. \qquad \text{(A.26)}
$$

This establishes (4.13). If $k+1$ is even, we get from (4.12) and (4.6):

$$
\begin{aligned}
\mathbb{E}\{\tilde{\mu}_{k+1}(X_1,\ldots,X_{k+1})\} &= \frac{1}{2}\big[\mathbb{E}\{(X_1-X_2)^{k+1}\} \\
&\quad -\sum_{j=2}^{k-1}(-1)^j C_{k+1}^j \mathbb{E}\{\bar{\mu}_j(X_1,\ldots,X_j)\bar{\mu}_{k+1-j}(X_{j+1},\ldots,X_{k+1})\}\big]
\end{aligned}
$$

$$
= \frac{1}{2}\big[\mathbb{E}\{(X_1-X_2)^{k+1}\} - \sum_{j=2}^{k-1}(-1)^j C_{k+1}^j \mathbb{E}\{\bar{\mu}_j(X_1,\ldots,X_j)\}\mathbb{E}\{\bar{\mu}_{k+1-j}(X_{j+1},\ldots,X_{k+1})\}\big]
$$

$$
= \frac{1}{2}\big[\mathbb{E}\{(X_1-X_2)^{k+1}\} - \sum_{j=2}^{k-1}(-1)^j C_{k+1}^j \mu_j\mu_{k+1-j}\big] = \mu_{k+1}\,. \qquad \text{(A.27)}
$$

This completes the proof of (4.14). □

PROOF OF PROPOSITION 5.1 To get (5.1), we develop $\mathbb{E}\{(X_1Y_2 - X_2Y_1)^2\}$ and use the independence assumptions:

$$\begin{aligned}
\mathbb{E}\{(X_1Y_2 - X_2Y_1)^2\} &= \mathbb{E}\{X_1^2Y_2^2 + X_2^2Y_1^2 - 2X_1X_2Y_1Y_2\} \\
&= \mathbb{E}\{X_1^2Y_2^2\} + \mathbb{E}\{X_2^2Y_1^2\} - 2\mathbb{E}\{X_1Y_1X_2Y_2\} \\
&= \mathbb{E}\{X_1^2\}\mathbb{E}\{Y_2^2\} + \mathbb{E}\{X_2^2\}\mathbb{E}\{Y_1^2\} - 2\mathbb{E}\{X_1Y_1\}\mathbb{E}\{X_2Y_2\}. \quad \text{(A.28)}
\end{aligned}$$

□

PROOF OF COROLLARY 5.2 (5.6) follows on replacing X_i and Y_i by the corresponding centered variables $\tilde{X}_i$ and $\tilde{Y}_i$ in Proposition 5.1. (5.7) and (5.8) are obtained as special cases of (5.6). □

PROOF OF PROPOSITION 5.3 To get (5.1), we develop $\mathbb{E}\{(A_1B_2 - A_2B_1)(C_1D_2 - C_2D_1)\}$ and use the independence assumptions:

$$\begin{aligned}
\mathbb{E}\{(A_1B_2 - A_2B_1)(C_1D_2 - C_2D_1)\} &= \mathbb{E}\{A_1B_2C_1D_2\} + \mathbb{E}\{A_2B_1C_2D_1\} \\
&\quad -\mathbb{E}\{A_1B_2C_2D_1\} - \mathbb{E}\{A_2B_1C_1D_2\} \\
&= \mathbb{E}\{A_1C_1\}\mathbb{E}\{B_2D_2\} + \mathbb{E}\{B_1D_1\}\mathbb{E}\{A_2C_2\} \\
&\quad -\mathbb{E}\{A_1D_1\}\mathbb{E}\{B_2C_2\} - \mathbb{E}\{B_1C_1\}\mathbb{E}\{A_2D_2\}. \quad \text{(A.29)}
\end{aligned}$$

Suppose now that $(A_1, B_1, C_1, D_1)'$ and $(A_2, B_2, C_2, D_2)'$ have the same covariance matrix.

On replacing Z_i by $\tilde{Z}_i$ in the above identity, we can then write:

$$\begin{aligned}\mathbb{E}\{(\tilde{A}_1\tilde{B}_2-\tilde{A}_2\tilde{B}_1)(\tilde{C}_1\tilde{D}_2-\tilde{C}_2\tilde{D}_1)\} &= 2\mathbb{E}\{\tilde{A}_1\tilde{C}_1\}\mathbb{E}\{\tilde{B}_1\tilde{D}_1\}-2\mathbb{E}\{\tilde{A}_1\tilde{D}_1\}\mathbb{E}\{\tilde{B}_1\tilde{C}_1\} \\ &= 2\mathsf{C}[A_1,C_1]\mathsf{C}[B_1,D_1]-2\mathsf{C}[A_1,D_1]\mathsf{C}[B_1,C_1] \qquad \text{(A.30)}\end{aligned}$$

from which (5.11) follows. □