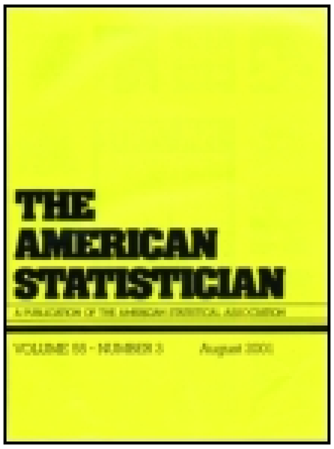

The American Statistician
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/utas20

The Meaning of Kurtosis: Darlington Reexamined

J. J. A. Moors [a]

[a] Department of Econometrics , Tilburg University , 5000LE , Tilburg , Netherlands

Published online: 27 Feb 2012.

To cite this article: J. J. A. Moors (1986) The Meaning of Kurtosis: Darlington Reexamined, The American Statistician, 40:4, 283-284

To link to this article: http://dx.doi.org/10.1080/00031305.1986.10475415

Theorem 2. Let (X, Y) be a pair of random variables whose distribution H is of the form (2.2) for some ϕ in Φ. Then

$$\tau(X, Y) = 4\int_0^1 \frac{\phi(t)}{\phi'(t)} dt + 1.$$

Proof. First note that $H(x, y) = 0$ for all (x, y) such that $\phi(x) + \phi(y) = 0$. Hence we can compute τ using (v) by integrating H over the region in which there is a density. That is,

$$\tau = \iint_{\phi(x)+\phi(y)<\phi(0)} H(x, y) \frac{-\phi''(H)\phi'(x)\phi'(y)}{[\phi'(H)]^3} dxdy.$$

Making the same transformations as in Theorem 1 and integrating yields the desired conclusion.

Figure 2 summarizes what the graph of $\phi(t)/\phi'(t)$ tells us about $H(x, y)$. The probability associated with the singular component and an estimate of Kendall's τ are readily available. The graph is of most interest when comparing two copulas. Note that Fréchet's lower bound gives $\phi(t)/\phi'(t) = t - 1$. As $\phi(t)/\phi'(t)$ approaches 0, we get a distribution close to Fréchet's upper bound.

Examples 1 and 2 (continued). In this case, it is easy to see that $\tau(X, Y) = \alpha/(\alpha + 2)$ for all $\alpha \geq -1$. In particular, note that the cases $\tau = -1$ and 0 correspond to the Fréchet lower bound and the independence distribution ($\alpha = -1, 0$, respectively), and τ approaches 1 as α tends to infinity. It is easy to verify directly that $\lim_{\alpha\to\infty} H_\alpha(x, y) = \min\{x, y\}$.

Example 3 (continued). Here, we see that $\tau(X, Y) = 1 - 2/\alpha$, which suggests that G_α approaches Fréchet's upper bound as α increases indefinitely. This is easy to check. Note also that $\tau = 0$ when $\alpha = 2$ but that

$$G_2(x, y) = \max\{0, 1 - \sqrt{(1 - x)^2 + (1 - y)^2}\}$$

is not the independence distribution. This illustrates the parenthetical comment made in exercise (iv).

5. COMMENT

Copulas of the form (2.2) serve other purposes besides those outlined in this article. Some of their theoretical uses are described in the book by Schweizer and Sklar (1983) and in a paper by Genest and MacKay (1986). In the latter, for example, it is shown how these copulas can be used to generate one-parameter families of bivariate distributions with prescribed marginals in such a way as to approach the Fréchet bounds "smoothly." Two examples of such families have been presented here (Examples 1 and 2). In general, it turns out that the convergence of a sequence H_n of copulas of the form (2.2) can be determined by simply looking at the graph of ϕ_n/ϕ_n'. See Genest and MacKay (1986) for details.

[*Received October 1985. Revised June 1986.*]

REFERENCES

Clayton, D. G. (1978), "A Model for Association in Bivariate Life Tables and Its Application in Epidemiological Studies of Familial Tendency in Chronic Disease Incidence," *Biometrika*, 65, 141–151.

Cook, R. D., and Johnson, M. E. (1981), "A Family of Distributions for Modelling Non-elliptically Symmetric Multivariate Data," *Journal of the Royal Statistical Society*, Ser. B, 43, 210–218.

——— (1986), "Generalized Burr-Pareto-Logistic Distributions With Applications to a Uranium Exploration Data Set," *Technometrics*, 28, 123–131.

Fréchet, M. (1951), "Sur les Tableaux de Corrélation Dont les Marges Sont Données," *Annales de l'Université de Lyon*, Série 3, 14, 53–77.

Genest, C., and MacKay, R. J. (1986), "Copules Archimédiennes et Familles de Lois Bidimensionnelles Dont les Marges Sont Données," *The Canadian Journal of Statistics*, 14, 145–159.

Oakes, D. (1982), "A Model for Association in Bivariate Survival Data," *Journal of the Royal Statistical Society*, Ser. B, 44, 414–422.

Schweizer, B., and Sklar, A. (1983), *Probabilistic Metric Spaces*, New York: North-Holland.

The Meaning of Kurtosis: Darlington Reexamined

J. J. A. MOORS*

There seems to be no universal agreement about the meaning and interpretation of kurtosis. An easy interpretation is given here: kurtosis is a measure of dispersion around the two values $\mu \pm \sigma$.

The concept of kurtosis seems to be rather difficult to interpret. Most statistical textbooks describe kurtosis in terms of peakedness, and some seek the explanation in heavy tails. Ben-Horim and Levy (1984) is an exception: it presents a more elaborate example featuring a bimodal distribution.

Bimodality as interpretation of kurtosis was introduced in Darlington (1970). Unfortunately, he pushed an otherwise correct argument one step too far. The present note reexamines his reasoning.

The kurtosis k will be defined here as the normalized fourth central moment; compare Kendall and Stuart (1969). So, any random variable X with expectation $E\{X\} = \mu$, variance $V\{X\} = \sigma^2$, and finite fourth moment has

$$k = E\{X - \mu\}^4/\sigma^4.$$

Introduction of the standardized variable $Z := (X - \mu)/\sigma$

*J. J. A. Moors is Senior Lecturer, Department of Econometrics, Tilburg University, 5000LE Tilburg, Netherlands. The author is grateful to two unknown referees, whose comments substantially improved earlier drafts of this article.

gives $k = E\{Z^4\} = V\{Z^2\} + [E\{Z^2\}]^2$, immediately implying the central formula

$$k = V\{Z^2\} + 1. \tag{1}$$

The inequality $k \geq 1$ follows at once; the minimum is attained for a symmetric two-point distribution. Further, (1) implies that k can be viewed as measure of the dispersion of Z^2 around its expectation 1, or equivalently, the dispersion of Z around the values -1 and $+1$. This was essentially Darlington's reasoning, who next concluded that k is a measure of bimodality. This last conclusion does not follow, however. Bimodal distributions can have large kurtosis; this occurs if the modes are not close to the points $Z = \pm 1$. This last phenomenon was illustrated by Hildebrand (1971) by means of a family of double gamma distributions, among others. As a consequence, Darlington's result did not receive the attention it deserves.

A valid interpretation may be formulated as follows: kurtosis measures the dispersion around the *two* values $\mu \pm \sigma$; it is an inverse measure for the concentration in these two points. High kurtosis, therefore, may arise in two situations: (a) concentration of probability mass near μ (corresponding to a peaked unimodal distribution) and (b) concentration of probability mass in the tails of the distribution. The existence of these two possibilities explains the present confusion about the interpretation of kurtosis. Since $V\{Z^2\}$ may be unbounded if heavy tails occur, (1) also indicates clearly why k is strongly affected by the tail behavior of the distribution. Darlington (1970) gave an additional argument by considering the contribution to k of additional point masses.

Sometimes, $k^*: = k - 3$ is presented as an alternative definition of kurtosis, since $k^* = 0$ for normal distributions. There is, however, nothing special about the kurtosis of normal distributions. A better alternative in the spirit of this note would be $k': = k - 1$. Of course, one could as well define a measure of concentration in $\mu \pm \sigma$ as $1 - k$ or $1/k$, the latter having the advantage of taking values between 0 and 1.

Since the derivation of (1) is a nice little exercise in dispersion, I expect to see it in any forthcoming textbook.

[*Received October 1985. Revised January 1986.*]

REFERENCES

Ben-Horim, Moshe, and Levy, Haim (1984), *Statistics: Decisions and Applications in Business and Economics* (2nd ed.), New York: Random House.

Darlington, Richard B. (1970), "Is Kurtosis Really 'Peakedness?'," *The American Statistician*, 24, No. 2, 19–22.

Hildebrand, David K. (1971), "Kurtosis Measures Bimodality?," *The American Statistician*, 25, No. 1, 42–43.

Kendall, M. G., and Stuart, A. (1969), *The Advanced Theory of Statistics* (Vol. 1), London: Charles W. Griffin.

Bias of S^2 in Linear Regressions With Dependent Errors

JEAN-MARIE DUFOUR*

Bounds are given for the expected value of the estimator of the error variance in linear regressions, when the errors are dependent or heteroscedastic. The bounds are valid irrespective of the covariance structure between the errors. Necessary and sufficient conditions to attain the bounds are supplied.

KEY WORDS: Variance estimation; Heteroscedastic errors; Bound on bias.

Recently, David (1985) showed the following interesting result: if $x_1, x_2, \ldots, x_n$ are random variables each with mean μ and variance σ^2, then

$$0 \leq E(s^2) \leq [n/(n-1)]\sigma^2, \tag{1}$$

where $s^2 = \Sigma_{i=1}^n (x_i - \bar{x})^2/(n-1)$. Of course, $E(s^2) = \sigma^2$ when $x_1, \ldots, x_n$ are uncorrelated. It is especially remarkable that the upper bound is valid for any form of dependence between $x_1, \ldots, x_n$. Does a corresponding result hold in a more general setup like the linear regression model? In this article a similar inequality is given for the usual estimator of the error variance in linear regressions when the errors have an arbitrary covariance structure.

Let

$$\mathbf{y} = X\boldsymbol{\beta} + \mathbf{u},$$

where $\mathbf{y}$ is an $n \times 1$ vector of observations on a dependent variable, X is an $n \times k$ fixed regressor matrix such that $1 \leq \text{rank}(X) = k < n$, $\boldsymbol{\beta}$ is a $k \times 1$ vector of coefficients, and $\mathbf{u} = (u_1, \ldots, u_n)'$ is an $n \times 1$ vector of disturbances (or errors) such that $E(\mathbf{u}) = 0$ and $E(\mathbf{uu}') = \Omega$. Ω is an arbitrary covariance matrix. Let

$$\hat{\boldsymbol{\beta}} = (X'X)^{-1}X'\mathbf{y}, \quad \hat{\mathbf{u}} = \mathbf{y} - X\hat{\boldsymbol{\beta}}, \quad s^2 = \hat{\mathbf{u}}'\hat{\mathbf{u}}/(n-k).$$

It is well known that $E(s^2) = \sigma^2$ when $\Omega = \sigma^2 I_n$, but this does not usually hold otherwise. Since $\hat{\mathbf{u}} = M\mathbf{u}$, where $M = I_n - P$ and $P = X(X'X)^{-1}X'$, we have

$$E(\hat{\mathbf{u}}'\hat{\mathbf{u}}) = E(\mathbf{u}'M\mathbf{u}) = \text{tr}(M\Omega) = \text{tr}(\Omega) - \text{tr}(P\Omega). \tag{2}$$

*Jean-Marie Dufour is Associate Professor of Statistics, Département de Sciences Économiques, Université de Montréal, C.P. 6128, Succursale A, Montréal, Québec, H3C 3J7, Canada, and Research Fellow at CORE, Université Catholique de Louvain, 34 voie du Roman Pays, 1348 Louvain-la-Neuve, Belgium. This work was supported by the Natural Sciences and Engineering Research Council of Canada, the Social Sciences and Humanities Research Council of Canada, the Fondation FCAR (Government of Québec) and the Centre de Recherche et Développement en Economique (Université de Montréal). The author thanks Herbert David, an associate editor, and two anonymous referees for several useful comments and suggestions.