Reprinted from

JOURNAL OF Econometrics

Journal of Econometrics 90 (1999) 317–336

Ordering univariate distributions by entropy and variance

Nader Ebrahimi[a], Esfandiar Maasoumi[b], Ehsan S. Soofi[c,*]

[a] Division of Statistics, Northern Illinois University, DeKalb, IL 60115, USA
[b] Department of Economics, Southern Methodist University, Dallas, TX 75275-0496, USA
[c] School of Business Administration, University of Wisconsin-Milwaukee, P.O. Box 742, Milwaukee, WI 53201, USA

ELSEVIER

ELSEVIER

Journal of Econometrics 90 (1999) 317–336

JOURNAL OF
Econometrics

Ordering univariate distributions by entropy and variance

Nader Ebrahimi[a], Esfandiar Maasoumi[b], Ehsan S. Soofi[c,*]

[a] *Division of Statistics, Northern Illinois University, DeKalb, IL 60115, USA*
[b] *Department of Economics, Southern Methodist University, Dallas, TX 75275-0496, USA*
[c] *School of Business Administration, University of Wisconsin-Milwaukee, P.O. Box 742, Milwaukee, WI 53201, USA*

Received 1 November 1996; received in revised form 1 April 1998

Abstract

This paper examines the role of variance and entropy in ordering distributions and random prospects. There is no universal relation between entropy and variance orderings of distributions. But we place their relationship in the context of a stronger ordering relation known as *dispersion ordering*. Further, some conditions are identified under which variance and entropy order similarly when continuous variables are transformed. We also analyze parametric changes which do not disturb the agreement between these rankings. The results are conveniently tabulated in terms of distribution parameters. © 1999 Elsevier Science S.A. All rights reserved.

JEL classification: C10; C11; C40

Keywords: Risk; Volatility; Entropy; Information theory; Ordering relations

1. Introduction

The principal activity in econometrics is assessing distribution functions and random prospects based on partial information. Estimation, tests of hypotheses, model selection, portfolio analysis, and inequality/poverty evaluation are but a few examples. Partial rankings of distributions can be obtained by stochastic dominance and similar order relations. The advantage of partial ranking is the

* Corresponding author. E-mail: esoofi@csd.uwm.edu

PII: S0304-4076(98)00046-3

avoidance of strong cardinalization as is necessary when a criterion such as variance or entropy is used to rank distributions.

But variance and other indices continue to be popular because of simplicity and the *complete ordering*. The historical development has resulted in variance playing the central role in measuring dispersion, uncertainty, evaluating fit, and much more. This has made variance the implicit 'index' of choice in assessing and ranking random variables. But much axiomatic work points to other indices, particularly entropy classes, as superior measures of information, see Maasoumi (1993) for a survey. The simplicity of variance, however, remains a major attraction.

The literature in economics, econometrics, and statistics records a major rise in the use of information theory concepts and measures in the last decade or so. The axiomatic appeal and the role played by entropy as a criterion function in deriving optimal measures and most 'data compatible' maximum entropy (ME) distributions, explain the recent abundance of entropy-based methods in econometrics and other areas; see, e.g., Holm (1993), Maasoumi (1993), Ryu (1993), Stutzer (1995), Golan et al. (1996), Fomby and Hill (1997), Soofi (1997), and Zellner (1997), and references therein. Applications of ME procedures in Bayesian econometrics are noteworthy. In the Bayesian method of moments (BMOM) approach, for example, the ME is the vehicle for generating post-data distributions for the *structural parameters* of econometric models and for prediction. For applications of BMOM see Green and Strawderman (1995), Zellner (1996a,b, 1997), Zellner and Sacks (1996), Zellner et al. (1997) and Tobias and Zellner (1997).

Interest in relating entropy to variance dates back to Shannon (1948) who proposed comparison of continuous random variables according to the entropy power fraction defined as the variance of a Gaussian random variable with a given entropy. Studying relationships between entropy and other moments has been a matter of interest in various contexts. Wyner and Ziv (1969) provided a bound on entropy in terms of a single moment of a continuous random variable. This entropy-moment inequality, for which the variance is a special case, has played an important role in the development of prediction theory (Shepp et al., 1979). Maasoumi and Theil (1979) gave approximations for two entropy-based income disparity measures in terms of the first four moments of the underlying income distributions. Chandra and Singpurwalla (1981) discussed entropy ordering in the context of some notions common between economics and reliability analysis. Mukherjee and Ratnaparkhi (1986) presented some relationships between the entropy and variance for a number of distributions, graphically.

Many of the well-known families of distributions have been characterized as the unique maximum entropy solutions. No such characterization is available in terms of variance. A maximum entropy density is spread out maximally by construction with entropy measuring the spread. But by virtue of our basic training, the spread of a distribution in terms of its variance is the more familiar concept. For example, Cressie (1993) begins with entropy as a measure of disorder to motivate statistics for spatial data and refers to its apparent relation with the variance as a more familiar measure to statisticians. There is no universal relationship between entropy and variance orderings of distributions. Some patterns can be established, however, for wide families of distributions. One way of analyzing the relationship between these two concepts is to study the effect of parameter variations in a family of densities.

There is not much work readily available that explores the similarities and dissimilarities of entropy with other measures in ordering distributions. Thus further light may be shed on these measures when viewed as indices for ranking random variables. In particular, we examine the relation between stronger partial rank orders and the entropy/variance orders and present some results that identify conditions for the equivalence of these orderings. We also analyze the effect of certain transformations of continuous random variables on the relation between entropy and variance. Finally, for well-known families of distributions, parametric variations are studied for their impact on these patterns of disagreement or agreement. The latter results are conveniently tabulated in terms of distribution parameters.

The outline of this paper is as follows. In Section 2, we point out some differences between entropy and variance, explore the difference in terms of an approximation of the density, and define our basic notations. In Section 3, we discuss the equivalence of entropy and variance orderings implied by a more general partial order relation. Conditions for transformations of continuous random variables that preserve the equivalence of variance and entropy orderings are also presented. The section ends with some results on the equivalence of entropy and variance orderings for well-known families of continuous distributions. This expands the entropy table of Verdugo Lazo and Rathie (1978). In Section 4, we examine the equivalence issue for some well-known discrete families of distributions. Concluding remarks are given in Section 5.

2. Preliminaries

Entropy of a random variable X with a probability distribution P is defined by

$$H(X) \equiv H[p(x)] = -\int_{-\infty}^{\infty} \log p(x)\,\mathrm{d}P(x),$$

where $p(x) = \mathrm{d}P(x)$ is the probability density (mass) function for the absolutely continuous (discrete) distribution P.

Entropy is a measure of disparity of the density $p(x)$ from the uniform. It measures uncertainty in the sense of the 'utility' of using $p(x)$ in place of the ultimate uncertainty of the uniform distribution (Good, 1968). Variance measures an average of distances of outcomes of the probability distribution $P(x)$ from the mean. Although both entropy and variance are measures of dispersion and uncertainty, the lack of a simple relationship between orderings of a distribution by the two measures emanates from quite substantial and subtle differences. Both measures reflect 'concentration' but their respective metrics for concentration are different. Unlike variance which measures concentration only around the mean, entropy measures diffuseness of the density irrespective of the location(s) of concentration.

In terms of mathematical properties, entropy is non-negative in the discrete case. For the discrete case, $H(X)$ is invariant under one-to-one transformations of X, but the variance is not. For the continuous case, neither the entropy nor the variance is invariant under one-to-one transformations of X. The entropy of a continuous random variable X takes values in $[-\infty, +\infty]$, and $E(X)^2 < \infty$ implies $H(X) < \infty$, but the converse may not hold.

It can be shown that, given $E|X|^k$,

$$H(X) \leqslant \frac{1}{k}\log\frac{2^k e\Gamma^k(1/k)E|X|^k}{k^{k-1}}, \quad k > 0, \tag{1}$$

where $|X|$ denotes the absolute value of X (Wyner and Ziv, 1969). The equality in Eq. (1) is attained by the maximum entropy distribution with density

$$p^*(x) = C(\eta)e^{-\eta|x|^k},$$

where the model parameter η is obtained as the Lagrange multiplier for satisfying the constraint $E|X|^k \leqslant \theta$ and $C(\eta)$ is the normalizing constant.

For the case of $k = 2$, relation (1) gives

$$\frac{e^{2H(X)}}{2\pi e} \leqslant \mathrm{Var}(X). \tag{2}$$

The ratio in Eq. (2) is the entropy power fraction proposed by Shannon (1948) for comparison of continuous random variables. The equality in Eq. (2) holds if and only if $P(x)$ is normal.

2.1. A Legendre series expansion[1]

Approximating the density function through a Legendre series expansion function provides significant insights about entropy and its relation to variance and higher-order moments.

[1] This approach was generously suggested by a referee.

A smooth and continuous density can be well approximated as

$$p(x) \approx a_0 G_0(x) + a_1 G_1(x) + \cdots + a_N G_N(x), \tag{3}$$

where $G_i(x)$, $i = 1, \ldots, N$ are Legendre polynomials:

$$G_0(x) = 1, \quad G_1(x) = x, \quad G_2(x) = 0.5(3x^2 - 1), \ldots .$$

Note that

$$\int_{-1}^{+1} G_i(x) G_j(x)\,\mathrm{d}x = \frac{2\delta_{ij}}{2i+1},$$

where δ_{ij} is the Kronecker's delta, and $x \in [-1, +1]$. One might obtain a_0 and a_1 to satisfy the normalization restriction and mean zero restriction.

Since

$$x^2 = \tfrac{1}{3}[2G_2(x) + G_0(x)],$$

variance is approximated by

$$V(x) = \int x^2 p(x)\,\mathrm{d}x \approx \tfrac{1}{3}[\tfrac{4}{5}a_2 + 2a_0].$$

This approximations reveals that variance increases if and only if a_2 increases. Other a_i, $i \geqslant 3$, do not influence variance.

Now, employing Eq. (3), it can be verified that the derivative of H with respect to a_2 is

$$\frac{\partial H}{\partial a_2} \approx -\int G_2(x) \log[a_0 G_0(x) + a_1 G_1(x) + \cdots + a_N G_N(x)]\,\mathrm{d}x.$$

Entropy increases with variance if this last expression is positive. That is, the variation of entropy depends on many more parameters than just a_2.

The Legendre series expansion reveals that entropy may be related to high-order moments of a distribution, which unlike the variance, could offer a much closer characterization of $p(x)$. In general, only when $p(x)$ is fully characterized by the first two moments (like normal) or when a quadratic approximation is satisfactory, then ordering distributions by variance alone is justifiable. Even in such situations, however, there is no loss in ordering by entropy since the two rankings would seem to agree in such cases.

Despite of the lack of a universal relationship between entropy and variance orderings, we can, nevertheless, identify conditions under which the two measures agree, and establish patterns of agreement for many important classes of distributions in Sections 3 and 4.

2.2. Notations

We use Ω to denote the class of probability distributions for which orderings of variance and entropy are under consideration. We discuss the problem in terms of two random variables X_1 with distribution $P_1 \in \Omega$ and X_2 with distribution $P_2 \in \Omega$. Variances and entropies of the two distributions will be denoted by V_1, V_2 and H_1, H_2, respectively. The *variance ordering* $V_1 \leqslant V_2$, will be denoted as $P_1 \overset{V}{<} P_2$ or $X_1 \overset{V}{<} X_2$. The *entropy ordering* $H_1 \leqslant H_2$, will be denoted as $P_1 \overset{E}{<} P_2$ or $X_1 \overset{E}{<} X_2$. When Ω_θ is a class of distributions indexed by a real parameter θ, then we write $V \searrow \theta\ (V \nearrow \theta)$ when variance is decreasing (increasing) in θ over the region R_θ in which $V(X|\theta) < \infty$. We write $H \searrow \theta\ (H \nearrow \theta)$ when entropy is decreasing (increasing) in θ whenever $H(X|\theta) < \infty$. When variance and entropy order similarly, we use notation $P_1 \overset{EV}{<} P_2$, $X_1 \overset{EV}{<} X_2$, $(V, H) \searrow \theta$, or $(V, H) \nearrow \theta$.

3. Continuous random variables

In this section first we discuss variance and entropy orderings implied by a stronger order relation. Then we provide some results for a few transformations of random variables. Finally, we examine entropy and variance orderings for many well-known parametric families. The entropies of most well-known absolutely continuous distributions are tabulated in Verdugo Lazo and Rathie (1978). We add a few more parametric families to their list. Three groups of parametric families are discussed: *Location-scale families*; *Shape-scale families*; and *Student t*, *F*, and *Beta* families.

3.1. Dispersion orderings

Two distribution functions P_1 and P_2 are said to have *dispersion ordering* (Bickel and Lehmann, 1976), denoted by $P_1 \overset{D}{<} P_2$ or $X_1 \overset{D}{<} X_2$, if and only if

$$P_2^{-1}(u) - P_2^{-1}(v) \geqslant P_1^{-1}(u) - P_1^{-1}(v) \quad \text{for all } 0 < v < u < 1, \tag{4}$$

where $P_i^{-1}(u) = \sup\{x\colon P_i(x) \leqslant u\}$, $i = 1, 2$ define the quantiles.

Dispersion ordering implies both variance ordering and entropy ordering; (see Oja 1981). That is, $P_1 \overset{D}{<} P_2$ is a sufficient condition for $P_1 \overset{EV}{<} P_2$. This is a strong stochastic order relation and can be a useful vehicle for determining when variance, entropy and many other orderings concur. Direct verification of dispersion ordering is difficult, but many results are available in the reliability literature which can be utilized for this purpose.

Example 3.1. Consider three duration random variables X_1, X_2, and X_3 with the following probability distribution functions:

$$P_1(x_1) = 1 - e^{-x_1};$$
$$P_2(x_2) = 1 - e^{-2x_2};$$
$$P_3(x_3) = 1 - \alpha e^{-x_3} - (1 - \alpha)e^{-2x_3},\ 1 < \alpha < 2.$$

Comparison of the entropies and variances of the two exponential distributions are straightforward. However, the entropy of mixture of exponentials is not available in closed form. By Theorem (2.2b) of Shaked and Shantikumar (1994) one can show that $P_1 \overset{D}{<} P_3$ and $P_3 \overset{D}{<} P_2$. Thus, $P_1 \overset{EV}{<} P_3$ and $P_3 \overset{EV}{<} P_2$.

The following theorem extends the impact of scale on entropy and variance of a random variable to more general transformations.

Theorem 1. Let X be a random variable with an absolutely continuous distribution P_X. Let $Y = g(X)$ where $g(x)$ is a function with a continuous derivative $g'(x)$ in the support of P_X and $\mathrm{E}(Y^2) < \infty$. If $|g'(x)| \geqslant 1$ for all x in the support of P_X, then $X \overset{EV}{<} Y$.

Proof. From Eq. (4) we have $P_1 \overset{D}{<} P_2$ if and only if $P_2^{-1}(\alpha) - P_1^{-1}(\alpha)$ is nondecreasing in $\alpha \in (0, 1)$. Letting $\alpha = P_1(x)$ we obtain $P_1 \overset{D}{<} P_2$ if and only if $P_2^{-1}(P_1(x)) - x$ is nondecreasing in x.

(i) If $g'(x) > 1$, then $g(x)$ is a strictly increasing function and we have

$$P_Y(y) = Pr[g(X) \leqslant y] = Pr[X \leqslant g^{-1}(y)] = P_X(g^{-1}(y)) \tag{5}$$

and

$$P_Y^{-1}(P_X(x)) = \sup\{y: P_Y(y) \leqslant P_X(x)\}$$
$$= \sup\{y: P_X(g^{-1}(y)) \leqslant P_X(x)\}$$
$$= \sup\{y: g^{-1}(y) \leqslant x\}$$
$$= \sup\{y: y \leqslant g(x)\}$$
$$= g(x).$$

Noting that $g'(x) \geqslant 1$ if and only if $g(x) - x$ is nondecreasing, we have $P_Y^{-1}(P_X(x)) - x$ is nondecreasing in x and obtain $P_1 \overset{D}{<} P_2$ which gives the result.

(ii) If $g'(x) < -1$, then $g(x)$ is a strictly decreasing function and $-Y = -g(X)$ is strictly increasing and Eq. (5) holds with Y and g replaced by $-Y$ and $-g$, respectively. Noting that $g'(x) \leqslant -1$ if and only if $g(x) + x$ is nonincreasing and that $g(x) + x$ is nonincreasing if and only if $-g(x) - x$ is nondecreasing. Thus from part (i) $X \overset{EV}{<} -Y$. The result is obtained by the fact that $V(-Y) = V(Y)$ and $H(-Y) = H(Y)$. □

Note that the $X \overset{E}{<} Y$ part may be directly seen via the following well-known relationship:

$$H(Y) = H(X) - \mathrm{E}\left[\log\left|\frac{\mathrm{d}}{\mathrm{d}X}g^{-1}(Y)\right|\right]. \tag{6}$$

However, no such a direct relationship is available for variance.

Theorem 2. Let X be a random variable with an absolutely continuous distribution and with a moment generating functions $M_X(t)$. Define $Y_i = \exp(\theta_i X)$, $i = 1, 2$, such that $\theta_2 > \theta_1 > 0$. Then $Y_1 \overset{EV}{<} Y_2$ if the following two conditions hold:

(i) $\mathrm{E}(X) > \log(\theta_1/\theta_2)/(\theta_2 - \theta_1)$;
(ii) $\mathrm{E}(Y_2) - \mathrm{E}(Y_1) = M_X(\theta_2) - M_X(\theta_1) > 0$.

Proof. Use Eq. (6) with $g^{-1}(y_i) = \log(y_i)/\theta_i$ to obtain $H(Y_i) = H(X) + \theta_i\mathrm{E}(X) + \log\theta_i$, $i = 1, 2$. Now for $\theta_1 < \theta_2$, assumption (i) implies $Y_1 \overset{E}{<} Y_2$.

Let $\Delta(\xi) = P_2^{-1}(P_1(\xi)) - \xi$, where P_i is the probability distribution of Y_i, $i = 1, 2$. A result from Oja (1981), (p. 158) implies that if there exists a constant ξ_0 such that:

$$\Delta(\xi) \leqslant \mathrm{E}(Y_2) - \mathrm{E}(Y_1) \quad \text{for } \xi \leqslant \xi_0 \tag{7}$$

and

$$\Delta(\xi) \geqslant \mathrm{E}(Y_2) - \mathrm{E}(Y_1) \quad \text{for } \xi \geqslant \xi_0, \tag{8}$$

then $Y_1 \overset{V}{<} Y_2$. Note that $P_Y(y) = P_X(\log(y)/\theta_2)$ implies that $\Delta(\xi) = \xi^{\theta_2/\theta_1} - \xi$. As a function of ξ, $\Delta(\xi)$ is convex function, passes through the point (1, 0), $\Delta(\xi) \leqslant 0$ for $\xi \in (0, 1]$, and attains its minimum at a point in (0, 1]. Let $\delta = \mathrm{E}(Y_2) - \mathrm{E}(Y_1)$. Since $\xi > 0$, it is clear that for $\delta > 0$, Eqs. (7) and (8) are satisfied with $\xi_0 = \Delta^{-1}(\delta)$. $\square$

At this point we remark that $\delta > 0$ is also necessary for $Y_1 \overset{V}{<} Y_2$. To see this, note that for $\delta \in [\min \Delta(\xi), 0]$, the horizontal line $h(\xi) = \delta$ intersects $\Delta(\xi)$ at two points, say ξ_1 and ξ_2, $\xi_1 < \xi_2$. In this case, $\Delta(\xi) < \delta$ for $\xi_1 < \xi < \xi_2$ and $\Delta(\xi) > \delta$ for $0 < \xi < \xi_1$ and for $\xi > \xi_2$. Thus, Eqs. (7) and (8) cannot be satisfied when $\delta < 0$.

Corollary 1. Let X be a nonnegative random variable with an absolutely continuous distribution and $Y_i = X^{\theta_i}$, $\theta_2 > \theta_1 > 0$. If the assumptions (i) *and* (ii) *of Theorem 2 hold for* $\log X$, *then* $Y_1 \overset{EV}{<} Y_2$.

Proof. Use $y_i = \exp\{\theta_i \log x\}$, $i = 1,2$ in Theorem 2. $\square$

3.2. *Location-scale family*

A distribution is said to be in a location-scale family with location parameter α and scale parameter β if its density is in the form of:

$$p(x|\alpha, \beta) = \frac{1}{\beta} \mathrm{p}\left(\frac{x - \alpha}{\beta}\right).$$

Table 1 shows several well-known location-scale families of distributions. A random variable X with a distribution in a location-scale family may be written as $X = \beta Z + \alpha$, where Z has a distribution free from α and β. Thus, the variance and entropy of all location-scale distributions are independent of the

Table 1
Entropy and variance orderings for location-scale distributions

Family and density	Variance	Entropy	Orderings
Gaussian (Normal)			
$p(x) = \frac{1}{\sqrt{2\pi}\beta} e^{-1/2\beta^2(x-\alpha)^2}$	β^2	$\log\beta + \frac{1}{2}\log(2\pi e)$	$(V, H) \nearrow \beta$
Gumbel (Extreme value)			
$p(x) = \frac{1}{\beta} e\left[{}^{-(x-\alpha)/\beta - \exp\left(-\frac{x-\alpha}{\beta}\right)}\right]$	$\frac{\pi^2\beta^2}{6}$	$\log\beta + 1 + \gamma$ $\gamma = 0.5772\cdots$	$(V, H) \nearrow \beta$
Laplace (Double exponential) $p(x) = \frac{1}{2\beta} e^{-\lvert x-\alpha\rvert/\beta}$	$2\beta^2$	$\log\beta + \log(2e)$	$(V, H) \nearrow \beta$
Logistic $p(x) = \frac{1}{\beta}(1 + e^{-(x-\alpha)/\beta})^{-2} e^{-(x-\alpha)/\beta}$	$\frac{\pi^2\beta^2}{3}$	$\log\beta + 2$	$(V, H) \nearrow \beta$
Uniform $p(x\mid\alpha, \beta) = \frac{1}{\beta}, \alpha - \frac{\beta}{2} < x < \alpha + \frac{\beta}{2}$	$\frac{\beta^2}{12}$	$\log\beta$	$(V, H) \nearrow \beta$

location parameter α and can be written as

$$V(X|\alpha, \beta) = \beta^2 V(Z) \quad \text{and} \quad H(X|\alpha, \beta) = \log\beta + H(Z),$$

where $V(Z)$ and $H(Z)$ are constants independent of α and β. Thus, within each location-scale family we have $(V, H) \nearrow \beta$.

Between the location-scale families the equivalence does not always hold. If the scale is adjusted so that all variances are the same, then the Gaussian distribution has the maximum entropy. Writing the entropy in terms of variance, we find that the following relationship for the orderings across the families of distributions listed in Table 1 holds.

$$\textit{Uniform} \overset{V}{<} \textit{Gumbel} \overset{V}{<} \textit{Laplace} \overset{V}{<} \textit{Logistic} \overset{V}{<} \textit{Gaussian}$$

implies

$$\textit{Uniform} \overset{E}{<} \textit{Gumbel} \overset{E}{<} \textit{Laplace} \overset{E}{<} \textit{Logistic} \overset{E}{<} \textit{Gaussian}.$$

In particular, this entropy ordering holds if variances of the distributions listed in Table 1 are scaled to be the same. It is easy to see that the converse is not true. For example, *Laplace* $\overset{E}{<}$ *Gaussian* when $e/\pi \leqslant V_G/V_L < 1$, where V_G and V_L are variances of Gaussian and Laplace distributions, respectively.

From Table 1, we note that for the same scale β, the variances and the entropies of the distributions shown in the table have the same ordering (normal and uniform tie on variance). For a given scale parameter, the logistic distribution has the largest variance and entropy whereas the uniform has the smallest.

3.3. Shape-scale family

A shape-scale distribution with a shape parameter α and scale parameter β is defined by having a density of the form

$$p(x|\alpha, \beta) = \frac{1}{\beta} p\left(\frac{x}{\beta}|\alpha\right).$$

Table 2 shows the shape-scale families of distributions, $p(x|\alpha, \beta)$. The densities are parameterized such that α is the shape parameter and β is the scale parameter. The variance and entropy are monotone increasing in β, so $(V, H) \nearrow \beta$. For most of these families, either $V(X|\alpha, \beta)$ or $H(X|\alpha, \beta)$ or both are complicated functions of α. The variance and entropy of some distributions over the parameter space may be ordered based on the algebra of the Gamma function $\Gamma(z)$, Psi (digamma) function $\psi(z) = \mathrm{d} \log \Gamma(z)/\mathrm{d}z$, and trigamma function $\psi_z(z) = \mathrm{d}\psi(z)/\mathrm{d}z$. But for some distributions, the algebra for establishing the orderings is not tractable and the results of Section 3.1 prove to be useful.

The scaled ($\beta = 1$) Gamma family has dispersion ordering with respect to α (see Shaked, 1982). Thus the entropy and variance are ordered in α. Since the variance is increasing in α, we conclude that the entropy is also increasing in α, so $(V, H) \nearrow \alpha$.

If X has a gamma distribution $p_X(x|\alpha, \beta)$, then $Y = 1/X$ has an Inverse gamma distribution $p_Y(y|\alpha, 1/\beta)$. It is clear that the variance of inverse gamma is decreasing in α for $\alpha > 2$. Now, using Eq. (6) one can easily show that if $Y_i = 1/X_i$, $i = 1,2$, then $Y_1 \overset{E}{<} Y_2$ if and only if

$$H(X_2) - H(X_1) \geqslant 2\mathrm{E}\left[\log\left(\frac{Y_2}{Y_1}\right)\right]. \tag{9}$$

Two gamma distributions with parameter $\alpha_1 < \alpha_2$ satisfy Eq. (9) so entropy of inverse gamma is decreasing in α. Thus, $(V, H) \searrow \alpha$ for the inverse gamma family.

From Theorem 2.6 of Shaked (1982), it can be shown that the generalized normal distribution is dispersion ordered in α. Thus, for scaled ($\beta = 1$) generalized normal family, $(V, H) \nearrow \alpha$.

Table 2

Entropy and variance orderings for shape-scale distributions

Family and density	Variance	Entropy	Orderings
Gamma; Exponential, $\alpha = 1$; Erlang, $\alpha = 1, 2, \dots$; Chi-square, $\alpha = \frac{1}{2}, 2/2, \dots \beta = 2$; $p(x\|\alpha, \beta) = \frac{1}{\beta\Gamma(\alpha)}\left(\frac{x}{\beta}\right)^{\alpha-1} e^{-x/\beta}$	$\beta^2\alpha$	$\log\beta + \log\Gamma(\alpha) + (1-\alpha)\psi(\alpha) + \alpha$	$(V, H) \nearrow \beta$ $(V, H) \nearrow \alpha$
Inverse gamma; Inverse Chi-square, $\alpha = 1/2, 2/2, \dots \beta = 1/2$ $p(x\|\alpha, \beta) = \frac{1}{\beta\Gamma(\alpha)}\left(\frac{x}{\beta}\right)^{-\alpha-1} e^{-(x/\beta)^{-1}}$	$\frac{\beta^2}{(\alpha-1)^2(\alpha-2)}, \alpha > 2$	$\log\beta + \log\Gamma(\alpha) - (1+\alpha)\psi(\alpha) + \alpha$	$(V, H) \nearrow \beta$ $(V, H) \searrow \alpha > 2$
Generalized-normal; Half-Normal, $\alpha = 1$; Rayleigh, $\alpha = 2$; Maxwell-Boltzmann, $\alpha = 3$; Chi, $\alpha = 1, 2, \dots$ $p(x\|\alpha, \beta) = \frac{1}{2\beta\Gamma(\alpha/2)}\left(\frac{x}{\beta}\right)^{\alpha-1} e^{-(x/\beta)^2}$	$\frac{\beta^2[\alpha - 2\Gamma^2(\alpha/2 + 1/2)]}{2\Gamma^2(\alpha/2)}$	$\log\beta + \log\frac{\Gamma(\alpha/2)}{2} + \frac{(1-\alpha)}{2}\psi\left(\frac{\alpha}{2}\right) + \frac{\alpha}{2}$	$(V, H) \nearrow \beta$ $(V, H) \nearrow \alpha$
Inverse generalized-normal; Inverse Chi, $\beta = \sqrt{2}, \alpha = 1, 2 \dots$ $p(x\|\alpha, \beta) = \frac{1}{2\beta\Gamma(\alpha/2)}\left(\frac{x}{\beta}\right)^{-\alpha-1} e^{-(x/\beta)^{-2}}$	$\beta^2\left[\frac{1}{\alpha-2} - \frac{\alpha\Gamma^2(\alpha-1)}{\Gamma^2(\alpha/2)}\right]$	$\log\beta + \log\frac{\Gamma(\alpha/2)}{2} - \frac{\alpha}{2}\psi\left(\frac{\alpha}{2}\right) + \frac{\alpha}{2}$	$(V, H) \nearrow \beta$ $(V, H) \searrow \alpha > 2$
Log-normal $p(x\|\alpha, \beta) = \frac{1}{\sqrt{2\pi}\alpha x} e^{-1/2\alpha^2(\log x - \log\beta)^2}$	$\beta^2[e^{2\alpha^2} - e^{\alpha^2}]$	$\log\beta + \log\alpha + \frac{1}{2}\log(2\pi e)$	$(V, H) \nearrow \beta$ $(V, H) \nearrow \alpha$
Pareto $p(x\|\alpha, \beta) = \frac{\alpha}{\beta}\left(\frac{x}{\beta}\right)^{-\alpha-1}, x > \beta$	$\frac{\beta^2}{\alpha(\alpha-1)^2(\alpha-2)}, \alpha > 2$	$\log\beta - \log\alpha + \frac{1}{\alpha} + 1$	$(V, H) \nearrow \beta$ $(V, H) \searrow \alpha > 2$
Triangular $p(x\|\alpha, \beta) = \begin{cases} \frac{2}{\alpha\beta}\left(\frac{x}{\beta}\right), & 0 \leqslant x \leqslant \alpha\beta \\ \frac{2}{(1-\alpha)\beta}\left(1 - \frac{x}{\beta}\right), & \alpha\beta \leqslant x \leqslant \beta \end{cases}$	$\frac{\beta^2(\alpha^2 - \alpha + 1)}{18}$	$\log\beta - \log 2 + \frac{1}{2}$	$(V, H) \nearrow \beta$ $V \searrow \alpha < 1/2$ $V \nearrow \alpha > 1/2$ H Constant $\forall\alpha$
Weibull $p(x\|\alpha, \beta) = \alpha\left(\frac{x}{\beta}\right)^{\alpha-1} e^{-(x/\beta)^\alpha}$	$\beta^2[\Gamma(1 + 2/\alpha) - \Gamma^2(1 + 1/\alpha)]$	$\log\beta - \log\alpha + \frac{(\alpha-1)\gamma}{\alpha} + 1$ $\gamma = 0.5772\cdots$	$(V, H) \nearrow \beta$ $(V, H) \searrow \alpha > \gamma$

The orderings for the inverse generalized-normal are obtained similarly to the inverse gamma. The orderings for the log-normal, Pareto, and triangular distributions are straightforward. Note that for the triangular distribution, the entropy is constant with respect to the shape parameter, but the variance decreases in α for $\alpha < \frac{1}{2}$ and increases for $\alpha > \frac{1}{2}$.

The orderings for the Weibull distribution are obtained by application of Corollary 1. If X is an exponential random variable with $\beta = 1$, then $Y = X^{1/\alpha}$ has the Weibull distribution with shape parameter α. For $\alpha_2 < \alpha_1$, $\mathrm{E}(Y_2) > \mathrm{E}(Y_1)$. Thus, Corollary 1 is applicable for $\theta_1 = 1/\alpha_1 < \theta_2 = 1/\alpha_2$. Since $\mathrm{E}[\log(X)] = -\gamma$, where $\gamma = 0.5772\cdots$ is the Euler constant, condition (i) of Theorem 2 constrains the parameters as $\gamma < \alpha_2 < \alpha_1$. This is in accord with the sign change of the derivative of the Weibull entropy. We find that $(V, H) \searrow \alpha$ for $\alpha > \gamma$.

3.4. Student-t, F, and beta families

Table 3 shows the orderings for three other distributions. For the *Student t* distribution, it is easy to see that $V(X|\alpha)$ is decreasing in α. Entropy ordering is obtained by using derivative $\mathrm{d}H(X|\alpha)/\mathrm{d}\alpha = H_\alpha(X|\alpha)$. Then $H_\alpha < 0$ is implied by the fact $\psi_\alpha(\alpha/2 + 1/2) < \psi_\alpha(\alpha/2)$ for all α.

The variance orderings of F distribution in terms of both parameters are clear. The entropy ordering of F distribution in terms of β is obtained by taking the derivative of H with respect to β and using the first inequality in the following result:

$$\frac{1}{z} < \psi_z(z) < \frac{1}{z-1} \quad \text{for } z > 1;$$

Mitrinovic (1970), (p. 228). Entropy ordering of F with respect to α is also obtained using the derivative with respect to α. We see that $(V, H) \searrow \beta$ for $\beta > 4$, but $V \searrow \alpha$ for $\alpha > 2$, and $H \nearrow \alpha$.

The most complicated case is the case of beta distribution. The variance and entropy of the beta family of distributions, *Beta*(α, β), is shown in Fig. 1. The variance and entropy of *Beta*(α, β) are symmetric functions of the parameters (α, β). Note that the peak of the variance surface is near the origin of $\alpha\beta$-plane. The variance surface sharply descends to a plain with mild slopes. Whereas, the entropy surface first ascends sharply, reaches to its peak at $(\alpha, \beta) = (1, 1)$, then forms a plateau of mild descending slopes. Using the derivatives, we find: $(V, H) \nearrow \alpha$ for $(\alpha, \beta) \in R_\alpha$,$(V, H) \nearrow \beta$ for $(\alpha, \beta) \in R_\beta$, $(V, H) \searrow \alpha$ for $(\alpha, \beta) \in S_\alpha$, and $(V, H) \searrow \beta$ for $(\alpha, \beta) \in S_\beta$. The regions R_α, R_β, S_α, and S_β are defined in Table 3. Fig. 2 shows R_α and S_α in $\alpha\beta$-plane. Note that $R_\alpha(S_\alpha)$ is above (below) the stationary points of the variance (entropy) with respect to α. The areas R_β and

Table 3

Entropy and variance orderings for student t, F, and beta distributions

Family and density	Variance	Entropy	Orderings
Student t; $\alpha = 1, 2, \ldots$ Cauchy; $\alpha = 1$ $p(x\|\alpha) = \dfrac{1}{\alpha^{1/2}B(1/2,\ \alpha/2)}\left(1+\dfrac{x^2}{\alpha}\right)^{-(\alpha+1)/2}$	$\dfrac{\alpha}{\alpha-2},\ \alpha > 2$	$\log[\alpha^{1/2}B(1/2, \alpha/2)]$ $+\dfrac{1+\alpha}{2}\left[\psi\left(\dfrac{\alpha}{2}+\dfrac{1}{2}\right)-\psi\left(\dfrac{\alpha}{2}\right)\right]$	$(V, H)\searrow \alpha > 2$
F; $\alpha = 1, 2, \ldots$ $\beta = 1, 2, \ldots$ $p(x\|\alpha, \beta) = \dfrac{\alpha^{\alpha/2}\beta^{\beta/2}\qquad x^{\alpha/2-1}}{B(\alpha/2,\ \beta/2)(\beta+\alpha x)^{(\alpha+\beta)/2}}$	$\dfrac{2\beta^2}{\alpha(\alpha+\beta-1)(\beta-2)(\beta-4)},\ \beta > 4$	$\log\dfrac{\alpha B(\alpha/2,\ \beta/2)}{\beta}+\left(1-\dfrac{\alpha}{2}\right)\psi\left(\dfrac{\alpha}{2}\right)$ $-\left(1+\dfrac{\beta}{2}\right)\psi\left(\dfrac{\beta}{2}\right)+\dfrac{\alpha+\beta}{2}\psi\left(\dfrac{\alpha+\beta}{2}\right)$	$V\searrow\alpha,\ H\nearrow\alpha$ $(V, H)\searrow\beta > 4$
Beta $\alpha > 0, \beta > 0$ $p(x\|\alpha, \beta) = \dfrac{1}{B(\alpha,\ \beta)}x^{\alpha-1}(1-x)^{\beta-1},\ 0 \leqslant x \leqslant 1$	$\dfrac{\alpha\beta}{(\alpha+\beta+1)(\alpha+\beta)^2}$	$\log[B(\alpha, \beta)] - (\alpha-1)[\psi(\alpha)-\psi(\alpha+\beta)]$ $-(\beta-1)[\psi(\beta)-\psi(\alpha+\beta)]$	$(V, H)\nearrow\alpha, (\alpha, \beta)\in R_\alpha$ $(V, H)\searrow\alpha, (\alpha, \beta)\in S_\alpha$ $(V, H)\nearrow\beta(\alpha, \beta)\in R_\beta$ $(V, H)\searrow\beta, (\alpha, \beta)\in S_\beta$

$$R_\alpha = \left\{(\alpha,\ \beta){:}\alpha < \frac{\sqrt{(\beta+1)(9\beta+1)}-(\beta+1)}{4}\right\},\quad S_\alpha = \left\{(\alpha,\ \beta){:}\alpha > 1-\frac{(\beta-1)\psi_\alpha(\alpha+\beta)}{\psi_\alpha(\alpha)-\ \psi_\alpha\ (\alpha+\beta)}\right\},$$

$$R_\beta = \left\{(\alpha,\ \beta){:}\beta < \frac{\sqrt{(\alpha+1)(9\alpha+1)}-(\alpha+1)}{4}\right\},\quad S_\beta = \left\{(\alpha,\ \beta){:}\beta > 1-\frac{(\alpha-1)\psi_\beta(\alpha+\beta)}{\psi_\beta\ (\beta)-\psi_\beta(\alpha+\beta)}\right\}.$$

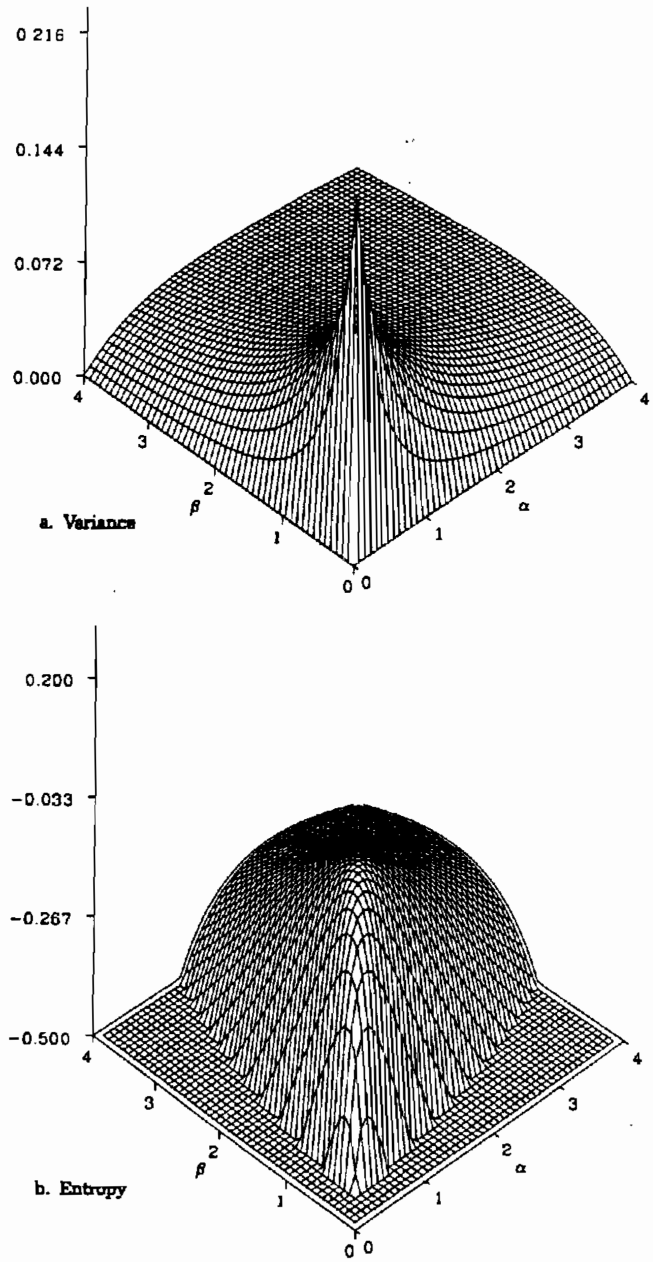

Fig. 1. Variance and entropy of beta family over the parameter space.

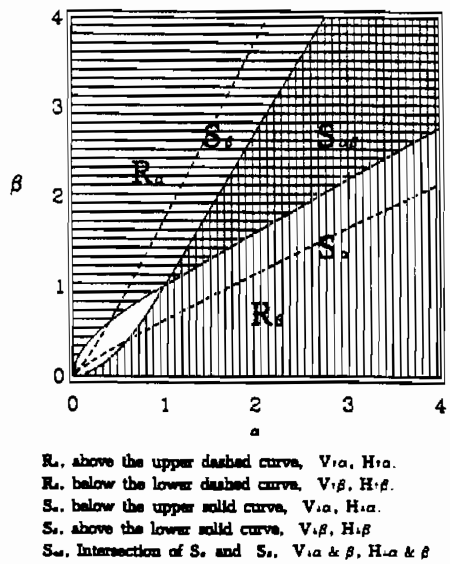

Fig. 2. Regions of variance and entropy orderings for beta family.

S_β are defined similarly. The central sections of the plain of the variance surface and plateau of entropy surface are over the region $S_{\alpha\beta} = S_\alpha \cap S_\beta$.

4. Discrete parametric families

In this section we examine entropy and variance orderings for the binomial, geometric, Poisson, and discrete uniform distributions. The results are summarized in Table 4.

Variance ordering for these distributions is straightforward. The entropy orderings for the geometric and uniform distributions are also straightforward.

The binomial entropy, $H(X|p)$, does not have a closed form. It is easy to see that $H(X|p) = H(X|1-p)$. Since $H(X|p)$ is symmetric around $p = \frac{1}{2}$ and is concave in p Marshall and Olkin, 1979, (p. 406), $H(X|p)$ is maximum at $p = \frac{1}{2}$; see Witsenhausen (1980). Thus, we conclude that $H \nearrow p$ when $p < \frac{1}{2}$, and $H \searrow p$ when $p > \frac{1}{2}$.

The entropy of the Poisson distribution, $H(X|\lambda)$ does not have a closed form and is a complicated function of the parameter λ. The entropy ordering of the Poisson distribution is obtained using the following result.

Lemma. If X is a Poisson random variable with $\mathrm{E}(X) = \lambda$, *then* $\mathrm{E}[\log(X+1)] > \log\lambda$.

Table 4
Entropy and variance orderings for discrete distributions

Family and probability function	Variance	Entropy	Orderings
Binomial			
$p(x\|n,p) = \binom{n}{x} p^x (1-p)^{n-x}$, $x = 0, 1, \ldots, n$	$np(1-p)$	$-np\log p - n(1-p)\log(1-p) - \log\Gamma(n+1) + \sum_{k=0}^{n}[\log\Gamma(k+1) + \log\Gamma(n-k+1)]p(k)$	$(V, H) \nearrow p < \frac{1}{2}$; $(V, H) \searrow p > \frac{1}{2}$
Geometric			
$p(x\|p) = p(1-p)^x$, $x = 0, 1, \ldots$	$\dfrac{1-p}{p^2}$	$-\log p - \dfrac{(1-p)\log(1-p)}{p}$	$(V, H) \searrow p$
Poisson			
$p(x\|\lambda) = \dfrac{e^{-\lambda}\lambda^x}{x!}$, $x = 0, 1, \ldots$	λ	$\lambda - \lambda\log\lambda + \sum_{k=0}^{\infty} \log\Gamma(k+1)p(k)$	$(V, H) \nearrow \lambda$
Uniform			
$p(x\|n) = \dfrac{1}{n}$, $x = 1, 2, \ldots, n$	$\dfrac{n^2-1}{12}$	$\log n$	$(V, H) \nearrow n$

Proof. The consecutive Poisson probabilities, $p_X(x) = P(X = x)$ and $p_X(x+1) = P(X = x+1)$ are related as follows:

$$(x+1)p_X(x+1) = \lambda p_X(x).$$

Using this fact we have

$$\sum_{x=0}^{\infty} p_X(x)\log(x+1) = \log\lambda + \sum_{x=0}^{\infty} p_X(x)\log\frac{p_X(x)}{p_X(x+1)}.$$

Next, we show that the summation in the right-hand side is positive by considering the Kullback–Leibler function $K(p_X{:}q_X)$, where q_X is the zero truncated Poisson distribution with parameter λ,

$$q_X(x) = \frac{1}{1-e^{-\lambda}}\frac{e^{-\lambda}\lambda^x}{x!}, \quad x = 1, 2, \ldots .$$

Noting that $q_X(k) = (1-e^{-\lambda})^{-1}p_X(x+1)$ for $k = x+1$, $x = 0, 1, 2, \ldots$, we have

$$K(p_X{:}q_X) = \log(1-e^{-\lambda}) + \sum_{x=0}^{\infty} p_X(x)\log\frac{p_X(x)}{p_X(x+1)} \geqslant 0.$$

Thus,

$$\sum_{x=0}^{\infty} p_X(x)\log\frac{p_X(x)}{p_X(x+1)} \geqslant -\log(1-\mathrm{e}^{-\lambda}) > 0. \qquad \square$$

Now, we show that the Poisson entropy $H(X|\lambda)$ is increasing in λ. After some simplifications we obtain

$$\frac{\mathrm{d}H(X|\lambda)}{\mathrm{d}\lambda} = -\log\lambda + \sum_{x=0}^{\infty}\log(x+1)\frac{\mathrm{e}^{-\lambda}\lambda^x}{x!}$$

$$= -\log\lambda + \mathrm{E}[\log(X+1)] > 0.$$

Thus, for the Poisson family we have $(V, H) \nearrow \lambda$.

5. Concluding remarks

Entropy and variance have been used to represent uncertainty, volatility and dispersion. We have identified patterns of agreement in orderings by variance and entropy, and situations in which they do not agree. General considerations suggest and specific approximation results given here show that entropy depends on much more information about a random variable than its variance.

Certain partial ranking relations, such as 'dispersion ordering' are stronger than either of entropy or variance orderings. Indeed, we show that the latter are implied by such general partial order. When this occurrence can be established exactly or by testing, ordering by variance and entropy will concur. We presented a few results that identified conditions under which variance and entropy order similarly. We then examined the impact of parametric changes on the entropy and variance of well known families of distributions and observed that agreement between these rankings is more frequent than disagreement.

Some limitations, however, should be noted. Our results pertain only to transformations of a random variable and to the known families of distributions. It seems that without imposing very stringent conditions on the density, specifying a very general condition for the equivalence of entropy and variance orderings is a formidable task.

We have only examined entropy and variance orderings of the parametric families of distributions with respect to variation of a single parameter. Examining entropy and variance orderings of distributions with respect to the variation of a vector of parameters is a subject of future study.

Finally, we have considered only the univariate case. It should be noted that entropy is a 'dimensionless' scalar measure of a distribution. This makes rankings in multivariable cases a natural extension. The same cannot be said of 'variance' and other moments which would have to be defined in terms of arbitrary scalar functions, such as the determinant or the trace of a covariance matrix. In the case of multivariate normal distribution, entropy ranks according to the determinant of the variance–covariance matrix. Extension to the multivariate case is an interesting and challenging research topic. Valuable references include the paper Atkinson and Bourguignon (1982) which compares multidimensional distributions according to the first- and second-order dominance relations and the paper by Ahmad and Gokhale (1989) which provides entropy expressions for several multivariate distributions.

Acknowledgements

We acknowledge with thanks the comments received from Arnold Zellner, an Associate Editor, and two referees on the previous drafts of this paper. Their comments enhanced our understanding of the topic and were very helpful in improving the presentations. We are, however, responsible for any remaining shortcomings.

References

Ahmad, N.A., Gokhale, D.V., 1989. Entropy expressions and their estimators for multivariate distributions. IEEE Transactions on Information Theory IT-35(3) 688–692.

Atkinson, A., Bourguignon, F., 1982. The comparison of multidimensioned distributions of economic status. Review of Economic Studies 49, 183–201.

Bickel, P.J., Lehmann, E.L., 1976. Descriptive statistics for nonparametric models. III. Dispersion. The Annals of Statistics 4, 1139–1158.

Chandra, M.T., Singpurwalla, N.D., 1981. Relationships between some notions which are common to reliability theory and economics. Mathematics of Operations Research 6, 113–121.

Cressie, N.A.C., 1993. Statistics for Spatial Data. Wiley, New York.

Fomby, T.B., Hill, R.C., 1997. Advances in Econometrics: Applying Maximum Entropy to Econometric Problems, vol. 12. JAI Press, Greenwich CT.

Golan, A., Judge, G., Miller, D., 1996. Maximum Entropy Econometrics. Wiley, New York.

Good, I.J., 1968. Utility of a distribution. Nature 219(5161).

Green, E., Strawderman, W., 1995. A Bayesian growth and yield model for slash pine plantation. Journal of Applied Statistics 23.

Holm, J., 1993. Maximum entropy lorenz curves. Journal of Econometrics 59, 377–389.

Maasoumi, E., 1993. A compendium to information theory in economics and econometrics. Econometric Reviews 12 (2), 137–181.

Maasoumi, E., Theil, H., 1979. The effect of the shape of the income distribution on two inequality measures. Economics Letters 4, 289–291.

Marshall, A.W., Olkin, I., 1979. Inequalities: Theory of Majorization and Its Application. Academic Press, New York.

Mitrinovic, D.S., 1970. Analytic Inequalities. Springer, New York.

Mukherjee, D., Ratnaparkhi, M.V., 1986. On the functional relationship between entropy and variance with related applications. Communications in Statistics-Theory and Methods 15, 291–311.

Oja, H., 1981. On location, scale, skewness and Kurtosis of univariate distributions. Scandinavian Journal of Statistics 8, 154–168.

Shaked, M., 1982. Dispersive ordering of distributions. Journal of Applied Probability 19, 310–320.

Shaked, M., Shantikumar, J.G., 1994. Stochastic Orders and their Applications. Academic Press, New York.

Shannon, C.E., 1948. A mathematical theory of communication. Bell System Technical Journal 27, 379–423.

Shepp, L.A., Slepian, D.., Wyner, A.D., 1979. On prediction of moving-average processes. Bell System Technical Journal 59, 367–415.

Soofi, E.S., 1997. Information theoretic regression methods. in: Fomby, T.B., Hill, R.C. (Eds.), Advances in Econometrics: Applying Maximum Entropy to Econometric Problems, vol. 12. JAI Press, Greenwich, CT, pp. 25–83.

Stutzer, M., 1995. A Bayesian approach to diagnostics of asset pricing models. Journal of Econometrics 68, 367–397.

Verdugo Lazo, A.C.G., Rathie, P.N., 1978. On the entropy of continuous probability distributions. IEEE Transactions on Information Theory IT-24(1), 120–122.

Witsenhausen, H.S., 1980. Some aspects of convexity useful in information theory. IEEE Transactions on Information Theory IT-26(3), 265–271.

Wyner, A.D., Ziv, J., 1969. On communication of analog data from bounded source spaee. Bell System Technical Journal 48, 3139–3172.

Zellner, A., 1996a. Bayesian method of moment/instrumental variable (BMOM/IV) analysis of mean and regression models. in: Lee, J., Johnson, W., Zellner, A. (Eds.), Modeling and Prediction: Honoring Seymour Geisser. Springer, Berlin.

Zellner, A., 1996b. Models, prior information, and Bayesian analysis. Journal of Econometrics 75, 51–68.

Zellner, A., 1997. The Bayesian method of moments (BMOM) theory and applications. in: Fomby, T.B., Hill, R.C. (Eds.), Advances in Econometrics: Applying Maximum Entropy to Econometric Problems, vol. 12. JAI Press, Greenwich, CT, pp. 85–106.

Zellner, A., Saeks, B., 1996. Bayesian method of moment (BMOM) analysis of the multiple regression model with autocorrelated errors. H.G.B. Alexander Research Foundation, University of Chicago.

Zellner, A., Tobias, J., Ryu, H.K., 1997. Bayesian method of moments (BMOM) analysis of parametric and semiparameteric regression models. Manuscript presented at the Fourth World Meeting of the Int. Society for Bayesian Analysis, Istanbul, Turkey.