

Directional entropy and tail uncertainty, with applications to financial hazard

Roger J. Bowden*

“Mine is a long and sad tale”, said the Mouse, turning to Alice and sighing. “It is a long tail certainly,” said Alice, looking down with wonder at the Mouse’s tail; “but why do you call it sad?” And she kept on puzzling about it while the mouse was speaking...

Abstract

Financial risk management metrics such as value at risk (VaR) can be illuminated by means of a regime-specific concept of directional entropy. This enables a change of measure via a rescaling function to an equivalent logistic distribution, one that has the same total and directional entropies at the chosen critical point. VaR rescaling adjusts the critical probability to capture the long tail entropy, and the scaling function can be used as a comparative metric for tail length. Directional entropy can be used to identify regions of maximal exposure to new information, which can actually increase entropy rather than collapse it.

JEL numbers: C19,44; D81; G11,21,22,32.

Key words: Conditional value at risk, entropy, expected shortfall, logistic distribution, measure shift, VaR.

* School of Economics and Finance, Victoria University of Wellington, New Zealand, and Kiwicap Research Ltd, ph. + 64 472 4984/4986, fax 4983. Email roger.bowden@kiwicap.co.nz or roger.bowden@vuw.ac.nz. Thanks go to seminar participants at the University of South Australia, also to *QF* referees for comments that improved the paper’s focus.

1. Introduction

In fields such as financial risk management and option pricing, concern attaches not so much to total uncertainty over the entire range of potential outcomes, as in its concentration over designated ranges, commonly the left or right hand tails. Traditional uncertainty or spread metrics, such as differential entropy or moment based statistics, do not capture the regime specific exposures. In addition, they can fail for very long tailed distributions because the relevant moments do not exist. Indeed, the Cauchy does not even have a mean, which can be a problem for other members of the Levy stable family, distributions which have found applications for financial data (see Tokat *et al* 2003 for the stable Paretian). It is not clear in such cases just what is meant by a long tail. The question of just how long is a tail, or whether one kind of distribution has an intrinsically longer tail than another, requires a more general metric to resolve it.

As an instance, value at risk (VaR) has become a favoured tool of financial regulators, in bank capital adequacy (see Berkowitz and O'Brian (2002), Alexander *et al* (2006)) and other fields such as insurance and fund management. Managers fix a tolerance probability for losses in excess of a stated amount, the VaR critical point. The basic welfare reference is therefore to a cumulative probability, i.e. the value of the distribution function at the VaR critical point. This has attracted criticism on the grounds that two distributions can have the same tail probability, but if one has a much longer tail than the other, then the firm is exposed to greater potential losses once the critical point is breached. In other words, the VaR probability underestimates the sting in the tail.

Conditional tail moments have been suggested as a criterion. Conditional value at risk, or expected shortfall, refers to the censored mean in the tail and has become popular in insurance circles, bank capital adequacy and other portfolio contexts (Uryasev 2000, Palmquist *et al* 2001, Rockafellar and Uryasev 2002, Scaillet (2003), Alexander and Baptista 2004, Bertsimas *et al* 2004). Other proposals have suggested specific higher order moments (Yamai and Yoshida 2002). On the other hand, Bowden (2006) has pointed out an inconsistency between VaR and CVaR, in that tolerances for the one cannot be set independently of the other. Moreover, the required moments must actually exist; thus in the case of CVaR, the Cauchy distribution has an undefined lower conditional mean.

Risk management tools such as VaR or CVaR (collectively, generalised value at risk GVaR) can be viewed as analogous to put options written by the manager (Artzner *et al* 1999, Bowden 2006). Similar remarks to the above can therefore apply to call options and warrants

that benefit in this case from exposure to the upper tail. Venture capital investments are very often structured in just this form, essentially as real options. The task in such contexts becomes one of identifying investments that benefit from upper directional entropy, including choosing strike prices that will result in maximal exposure to changes in entropy as information unfolds.

Such applications share a more or less common theme, namely the importance of understanding and measuring uncertainty over the different regions of the support axis for the underlying distribution. Entropy is commonly proposed as a general uncertainty metric. But standard differential (or ‘total’) entropy has general but not sufficiently specific relevance, for it is a property of the entire distribution. The agenda for the present paper is to develop an entropy concept that can apply to designated support regions, e.g. the upper or lower tails, and to show how this can be applied to some common problems in financial economics and risk management, notably those concerned with GVaR.

The paper develops an entropy concept called directional entropy, which is region specific but more informative than simply partitioning the differential entropy integral, or the conditional entropy in each region. Directional entropy is a metric for the uncertainty that arises specifically from the subject region, so that upper and lower directional entropies for a simple binary support partition will add up to total entropy.

One distribution may have a longer apparent tail than another simply because its overall variation is greater. However this is not very useful in comparing different distribution families. What turns out to be relevant is not so much the total entropy as the ratio of directional entropy to total entropy. This defines a change of measure and it is the comparison of this to the original distribution function value at the chosen critical point that provides a general non moment-based comparative measure for tail length.

Directional entropy is the entropy concept most closely related to contexts such as GVaR, or option pricing. The logistic distribution plays a key benchmarking role in this relationship for two reasons, first the linearity of its log odds function and second, because the tail probability is a sufficient proxy for expected shortfall, or option value. Comparisons or corrections are vectored by conversion of the subject distribution to an equivalent logistic distribution, one that has the same total entropy and directional entropy at the given marker value e.g. strike price or VaR critical point. The conversion is achieved by a scaling function that can be interpreted as a Radon-Nikodym measure for an implicit change of measure.

Conversion to logistic-equivalent measure (L_q) allows rescaling of tail probabilities in a way that is useful in GVaR and related contexts. If a distribution of investment outcomes has

a standard 5% VaR tail probability but a long tail, then one can convert it to the equivalent logistic and look at the resulting tail probability. If this comes out at 10%, say, it will suggest that the natural VaR probability (5%) is misleading as a hazard indicator and a more conservative stance should be taken. The scaling function in such a case would be greater than unity at the chosen critical point, indicating a longer tail than the equivalent logistic benchmark.

Because directional entropy is likely to be monotonically related to most sensible tail metrics, it becomes an effective hazard indicator once converted to equivalent probability terms via the L_q measure. Thus CVaR can be handled via that for the equivalent L_q measure. This gives a value proxy even when the natural distribution does not technically admit a CVaR, as noted above.

The paper is framed within the context of VaR and related concepts in financial risk management. However, many of the same principles apply to option pricing. L_q measure plays a useful convenience role in this, for the formulas for option pricing are particularly simple in the case of the logistic distribution. More important in this context is the response of value to flows of information, for in contexts such as venture capital and real options, tail value for a given strike price can actually increase as new information accumulates. Directional entropy enables the user to see more clearly where the maximal informational value gain is likely to occur and to set strike prices to take advantage.

The scheme of the paper is as follows. Section 2 is introductory, reviewing the connection between the log odds ratio and entropy theory as the mathematical expectation of a scaling function derived from the log odds function. Directional entropy is developed in section 3, which introduces the change of measure via the scaling function derived from the log odds function. Critical probability adjustment in VaR is an immediate application in section 4, and there is a further application to CVaR. Section 5 offers some concluding remarks, including potential applications of directional entropy to option value, and computational procedures.

2. Information measures: log odds and entropy

Statistical entropy is most naturally approached via the log odds function, so the first part of this section reviews some of this theory, expositing the general nature of the connection. It proceeds to introduce one of the key concepts, namely a scaling function based on the log derivative, and to relate this to differential or total entropy.

The starting point is a probability space $(\Omega, \mathfrak{F}, P)$ on which is defined a real-valued random variable $x(\omega)$, measurable with respect to \mathfrak{F} , with probability distribution function F . In financial applications the variable x might be taken as either returns, or financial value. It is assumed, at least initially, that the support of the distribution is the whole real line $-\infty < x < \infty$. In practice, this might not be technically true, e.g. where a boundary constraint such as limited liability might apply. However, this need not be unduly limiting. For instance, one could imagine that the success or otherwise of a financial venture depends on some underlying latent variable x . For positive values of the latent variable, returns are directly proportional to x . But if $x < 0$, then the project fails and negative values of this latent variables are projected on to the value zero (or some limited loss of capital) in return space. This means that the probability density of returns will be a mixed in character¹, with a spike at zero generated by negative values of the underlying latent variable. It is convenient in what follows to merge the latent variable x with observed return or value, and to assume it has a density $f(x)$.

2.1 Properties of the log odds function

Let X be any specific value of the random variable x , often denoted as a ‘marker value’ in what follows. Denote the event $\{x \leq X\}$ as $R_L(X)$, or just R_L where the regime marker value X is understood; similarly, $\{x > X\} = R_U(X)$ or just R_U . An initial question is how much information the given distribution function $F(x)$ gives us about the prospects $R_L(X)$ and $R_U(X)$. One measure of the information is the odds ratio. The magnitude

$$\lambda(X) = \ln \frac{1-F(X)}{F(X)} \quad (1)$$

will be referred to as the log odds of x being in $R_U(X)$ versus $R_L(X)$. Its complement $\bar{\lambda}(X) = -\lambda(X)$ will be the log odds ratio for the lower regime R_L . If X_m is the median of the distribution, the sign of the log odds will depend on whether $X < X_m$ (λ is positive) or $X > X_m$ (λ is negative). Measured thus, the log odds of regime R_U (i.e. $x > X$) diminish as X becomes larger: for any proposed log odds function, we must have $\lambda'(X) \leq 0$ and also

$\lambda(\infty) = -\infty$; $\lambda(-\infty) = \infty$, in an obvious sense. The derivative $\lambda'(X)$ will sometimes be referred to as the *rate of decay* of the log odds at point X , with the upper regime in mind. Mathematically, many of the formulas that follow are more naturally written in terms of $\bar{\lambda}(X)$ rather than $\lambda(X)$, but interpretive elements such as decay rates are more naturally oriented from left to right so the negative sign will be carried throughout. Although the

original logarithmic base for entropy theory was 2, natural logs will be assumed in what follows.

The density $f(x)$ corresponding to any given distribution function $F(x)$ can always be written in terms of its generic log odds function $\lambda(x)$ as

$$f(x) = -\lambda'(x)F(x)(1 - F(x)). \quad (2)$$

Written in the form

$$-\lambda'(x) = \frac{f(x)}{F(x)(1 - F(x))}, \quad (3)$$

the derivative of the log odds function is analogous to a signal to noise ratio. The denominator product can be viewed in the light of fuzzy logic as the product of fuzzy membership functions for R_U and R_L , and incorporates the categorisation doubt as to which regime the value x belongs. It has a maximum at the median X_m , diminishing thereafter. The numerator can loosely be regarded as conveying a degree of information about the interval $(x, x+dx)$. For symmetric densities, the signal to noise ratio ($-\lambda'(x)$) has a minimum at the median and is itself symmetric about that point.

The logistic constitutes a benchmark distribution for entropy theory and related topics. It is one for which the log odds function $\lambda(x)$ is linear in x : $\lambda'(x) = -\frac{1}{\beta}$; $\beta > 0$. The density is symmetric about the common mean and median $X_m = \mu$, while β is a scale or dispersion parameter such that $\sigma = \beta\pi / \sqrt{3}$ is the standard deviation. A convenient standardisation is $\tilde{x} = \frac{x - \mu}{\beta}$. The kurtosis coefficient is 4.2 versus the normal 3 and the total entropy is given by $2 + \ln \beta$. The logistic is one of the family of extreme value distributions, obtained as the limiting maxima of a collection of identically distributed random variables (Johnson *et al* 1994). This makes them potentially relevant in the analysis of rare events, which may depend upon a cluster of good or bad outcomes from a sequence of stages. Extreme value distributions have been used for financial returns data by Suárez and Menéndez (2005).

The logistic has moderately fat tails, relative to the normal density. Other fat tailed distributions commonly used for financial data are Student's t and various members of the Levy stable family. Jorion (2007 ch.14) is a useful survey, also Glasserman *et al* (2002) for the multivariate t distribution, Tokat *et al* (2003) for the Paretian stable. The logistic and the univariate t distributions will be used to illustrate the concepts in the present paper, along with the Cauchy and normal as two members of the Levy stable family. Comparisons will be

benchmarked on a common total entropy, chosen as 1.25, which is consistent with a 15 day standard deviation of 0.85% for a normal distribution, though a slightly higher entropy of 1.68 is used for the Student's t to be consistent with a choice of 4 degrees of freedom (e.g. Jorion (2007)). For illustrative purposes, all the distributions will be centred at zero.

Figure 1 depicts the log odds functions for the four distributions. The dotted straight line is the Logistic benchmark and it will be observed that the normal displays a steeper decay away from the median, taken as zero for all four distributions. The Cauchy exhibits much slower log odds in the tails, though a steeper sigmoid curve in the vicinity of the median, capturing the much sharper density peak at this point.

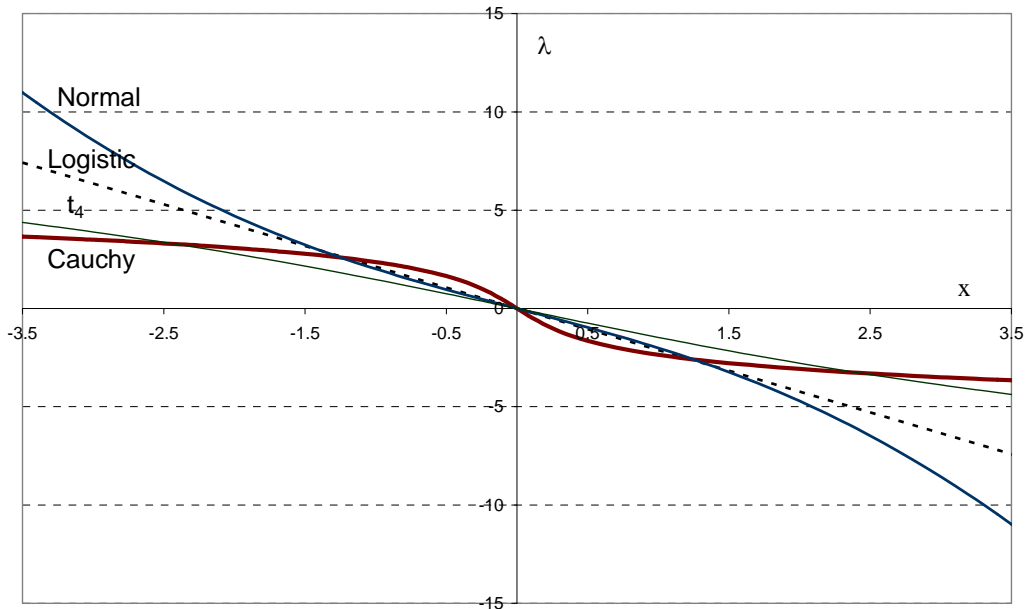


Figure 1: Comparative log odds functions

2.2 Differential entropy and the scaling function

For a continuous distribution, differential entropy is defined as

$$\kappa = - \int_{-\infty}^{\infty} f(x) \ln f(x) dx = -E[\ln f(x)]. \quad (4)$$

Continuous entropy metrics appear in various forms. Kullback-Leibler entropy measures the entropy gain of one distribution relative to another (Kullback-Leibler (1951), Kullback (1968)), while conditional and mutual entropy refer to the gains in information arising from associated information (Pinsker (1964)). The present contribution develops a directional

entropy concept associated with the kinds of partition that characterise VaR and associated concepts. The starting point is the parent differential entropy, which for brevity will sometimes be referred to as total entropy.

Expression (2) can be used to write differential entropy in terms of the log odds function. The following result serves to introduce a function of subsequent importance.

Lemma 1

Differential entropy satisfies

$$\kappa = E[2 - \ln(-\lambda'(x))]. \quad (5)$$

Equivalently, if we define a function:

$$\xi(x) = \frac{1}{\kappa} [2 - \ln(-\lambda'(x))],$$

then $E[\xi(x)] = 1$.

Proof: Apply property (2) to $\kappa = -E[\ln(f(x))]$, and use $E[\ln(F(x))] = E[\ln(1 - F(x))] = -1$.

□

The function $\xi(x)$ of Lemma 1 will be referred to the ‘density scaling function’ or just the scaling function where the context is clear. Later development identifies it with a Radon-Nikodym derivative for a change of measure, under suitable regularity conditions. Its slope and shape will depend on the particular distribution being considered. If the density $f(x)$ is symmetric, then the associated scaling function $\xi(x)$ is also symmetric about the median.

Figure 2 depicts scaling functions for the logistic, normal, t and Cauchy distributions, with parameters as earlier given. Those for the logistic and t_4 distributions are quite close together, but note the inflation of the Cauchy in the tails, relative to the constant logistic as benchmark. The scaling function $\xi(x)$ can be negative over certain ranges for some distributions, typically associated with very low values of the denominator ‘doubt’ product $F(x)(1-F(x))$ in expression (3). The normal distribution is a case in point; the scaling function for a standard normal turns negative at $x = \pm 7.25604$, corresponding to tail probabilities of 1.99E-13.

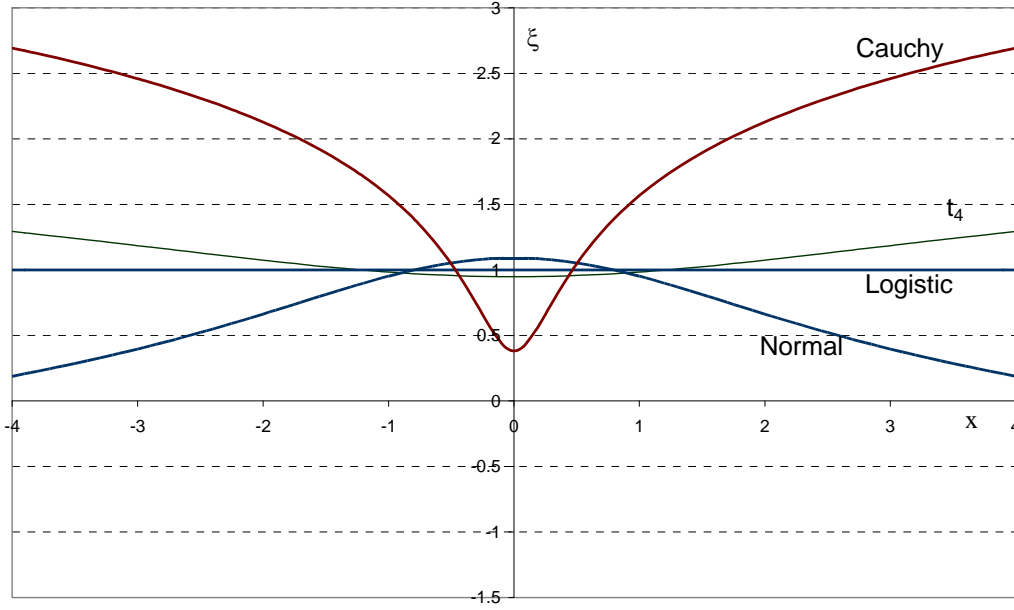


Figure 2: Scaling functions compared

Differential or total entropy is a property of the distribution as a whole. It says little about the precise way that uncertainty is distributed along the support axis. Given a marker point X , one might think of partitioning the entropy integral into upper and lower half integrals. This has no real informational meaning, and is open to a mathematical objection that the adding up constraint $F(\infty) = 1$ is not adequately incorporated in the partitioning. However, for applications such as GVaR it is necessary to consider just how uncertainty is distributed along the support axis.

A first approach to this problem might be to measure the binary entropy associated with a categorical random variable that signals whether the value x belongs to the upper or lower regime $R_L(X)$ or $R_U(X)$, for a given marker value X :

$$\kappa_I(X) = -[F(X) \ln F(X) + (1 - F(X)) \ln(1 - F(X))]. \quad (6)$$

This is called ‘positional’ or ‘locational’ entropy in Bowden (2007). It can be related to the conditional expectations of the log odds function:

$$\kappa_I(X) = F(X)E[\lambda(x) | x \leq X] = -(1 - F(X))E[\lambda(x) | x > X]. \quad (7)$$

Locational entropy has a number of applications, including the derivation of spread and asymmetry distribution diagnostics that do not depend upon benchmarks such as the Gaussian; or any assumption that relevant moments necessarily exist, which can be a problem

very long tail densities. However, it is not well adapted to the current context. One can certainly expect higher total entropy distributions of the same basic shape to have higher locational entropy at any given marker point X away from the median. Nevertheless, locational entropy is a purely positional concept, lacking a complete reference to total entropy, simply because it is always normalised to have value $\ln 2$ at the median (put $X = X_m$ in expression (6)). But an option written at a specific strike price will depend for its value not only on how far from the median it is set, but also upon which direction, and upon total uncertainty in the distribution. Directional entropy provides an inherently asymmetric entropy concept that encompasses both aspects, and is the subject of the next section.

3. Directional entropy and the scaling factor

This section introduces the key concept of directional entropy. As with locational entropy, it emerges naturally from the log odds function; locational entropy from $\lambda(x)$, and directional entropy from its derivative $\lambda'(x)$. The relationship with the log odds function facilitates a change of measure concept that converts natural to logistic measure, for which the log odds function is just a constant.

3.2 Directional entropy

From section 2.1 Lemma 1, the function $\xi(x) = (2 - \ln(-\lambda'(x))) / \kappa$ is such that $E[\xi(x)] = 1$. Given a marker value X and lower and indicator functions I_L, I_U for the respective regimes $R_L(X)$ and $R_U(X)$, define conditional expectations in each regime as:

$$\bar{\xi}_L(X) = E[I_L \xi] / E[I_L] = E[\xi(x) | x \leq X] ; \quad (8a)$$

$$\bar{\xi}_U(X) = E[I_U \xi] / E[I_U] = E[\xi(x) | x > X] . \quad (8b)$$

The functions $\bar{\xi}_L(x), \bar{\xi}_U(x)$ will be referred to as the lower and upper distribution scaling functions because they turn out to inflate or deflate the distribution function, while the density scaling function $\xi(x)$ operates on the density.

Proposition 1 shows that total differential entropy can be decomposed into components derived from each regime².

Proposition 1

Given a marker value X of the random variable x , let

$$\kappa_L(X) = \kappa \bar{\xi}_L(X) F(X) . \quad (9a)$$

$$\kappa_U(X) = \kappa \bar{\xi}_U(X) (1 - F(X)) ; \quad (9b)$$

Then total entropy divides as

$$\kappa = \kappa_L(X) + \kappa_U(X). \quad (10)$$

Proof: Follows by splitting the domain of integration implicit in Lemma 1 expression (5).

□

The functions $\kappa_L(X)$ and $\kappa_U(X)$ for a given market value X will be called the lower and upper directional entropies, respectively. The underlying reference is to the derivative $\lambda'(x)$ of the log odds function, as distinct from locational entropy, which refers to the log odds function $\lambda(x)$ itself, as in expression (7).

The decomposition (10) splits total entropy into a forward looking component $\kappa_U(X)$ and a backward looking component $\kappa_L(X)$. It follows that as the marker point X changes, there is an entropy tradeoff between the lower and upper regimes:

$$\kappa_L'(X) = \kappa \xi(X) f(X) = -\kappa_U'(X). \quad (11)$$

Note that if the density $f(x)$ is symmetric about the median X_m , then $\bar{\xi}_L(X_m) = \bar{\xi}_U(X_m) = 1$, which can provide a useful anchor point for the numerical evaluation of the conditional expectations $\bar{\xi}_L(x)$, $\bar{\xi}_U(x)$.

The logistic distribution constitutes a benchmark, for in this case the log odds function is linear, $\lambda'(x) = -1/\beta$ and $\xi(x) = 1$ for all x , meaning that $\bar{\xi}_L(X) = \bar{\xi}_U(X) = 1$ also. Hence for any given marker value X

$$\kappa_L(X) = \kappa F(X); \quad \kappa_U(X) = \kappa(1 - F(X)). \quad (12)$$

Directional entropy for the logistic is a simple function of the regime probability and total entropy. In a VaR context, lower directional entropy and critical probabilities effectively coincide. For different points along the support axis, the total entropy is just a constant, and tail probabilities ($F(X)$ for the lower, $1-F(X)$ for the upper) are sufficient statistics for the directional entropy. Compared across different logistic scale parameters, tail probabilities would be insufficient to capture the potential for risk, but the adjustment via total entropy is nevertheless straightforward.

For other distributions, lower and upper directional entropy diverge in more intrinsic fashion from $F(X)$ or $1-F(X)$ as the marker point X varies. If directional entropy is thought to be more closely related to welfare objectives, then tail probabilities and hence VaR, are insufficient welfare metrics in their natural form. However, a change of measure can be used to generate a closer correspondence and this is the topic of ensuing development.

For many purposes it will be convenient to renormalise by dividing directional entropy by total entropy. Given a subject distribution function $F(x)$ with differential entropy κ , set the marker X at a generic value x , and define a function $F_q(x)$ by either of the equivalent relationships:

$$F_q(x) = \frac{1}{\kappa} \kappa_L(x) = \bar{\xi}_L(x) F(x) \quad (13a)$$

$$1 - F_q(x) = \frac{1}{\kappa} \kappa_U(x) = \bar{\xi}_U(x) (1 - F(x)). \quad (13b)$$

Indeed if the function $\xi(x)$ is non-negative, the renormalisation (13) yields a proper distribution function in its own right. The fat tail distributions used in the present paper fall into this category. In the particular case of a logistic distribution, the measure shifts $\bar{\xi}_L(x), \bar{\xi}_U(x)$ are identically unity, which means that $F_q(x) = F(x)$, all x ; the logistic distribution remains invariant under the informational rescaling. The following result summarises.

Proposition 2

- (a) Suppose $\xi(x) \geq 0; -\infty < x < \infty$. Then $F_q(x)$ as defined by (13) is a probability distribution function. If it is differentiable, then its density $f_q(x) = \xi(x)f(x)$.
- (b) Let X be a given marker value and let $F_g(x; \mu_X, \beta_X)$ be a logistic distribution with centering and scale parameters μ_X, β_X chosen such that the total and directional entropies of F and F_g are the same at $x = X$, i.e. $\kappa_g = \kappa, \kappa_{gL}(X) = \kappa_L(X)$. Then $F_g(X; \mu_X, \beta_X) = F_q(X)$.

Proof:

(a) Note that ξ is non-negative and $E[\xi] = 1$, so that ξ qualifies as a Radon-Nikodym derivative on \mathfrak{R} . Hence given a marker value X , we can define a new measure such that $F_q(X) = E_x[\xi I_L(X)] = F(X)E[\xi(x) | x \leq X]$, which is expression (13a). The density follows by differentiation.

(b) $F_q(X) = \frac{1}{\kappa} \kappa_L(X) = \frac{1}{\kappa_g} \kappa_{gL}(X) = F_g(X)$, from the logistic property (12).

□

In what follows, F_q will be referred to as the distribution function for L_q measure, where L_q refers to ‘logistic equivalent’. It may be noted that if the parent density $f(x)$ is symmetric, then

so is $f_q(x)$. This follows because, as earlier noted, the scaling function $\xi(x)$ is then symmetric about the median. As earlier noted, if the underlying distribution is logistic to begin with, then F and F_q are one and the same.

Part (b) of the proposition is the basis for applications. The interpretation is as follows. Given a marker value X , the ordinate $F_q(X)$ is equal to that of a logistic with (i) the same total entropy as F ; and (ii) the same directional entropy at X , i.e. the total and directional entropies coincide at X . This does not mean that the L_q distribution is itself necessarily globally logistic in form; the effective logistic parameters β, μ are local and will move with the marker point X . Given X , and the given parent total entropy κ , the local logistic parameters β_X, μ_X can be obtained as:

$$\beta_X = e^{\kappa-2}; \quad \mu_X = X + \beta_X \ln \lambda_q(X) \quad (14)$$

where $\lambda_q(X) = (1 - F_q(X)) / F_q(X)$ is the odds function for F_q evaluated at X . The idea has similarities to an envelope relationship, although the local logistics are not in general tangent. Figure 5 of section 4 illustrates for the Cauchy distribution. With a long tailed distribution of this type there is a considerable divergence between F and F_q , as there is for some other members of the Levy stable family and for longer tailed members of the extreme value family such as the Gumbel. For Student's t distribution, the divergence is much smaller; figure 6 illustrates for the t distribution.

The density scaling function ξ plays a role very similar to the state price deflator of financial asset pricing, which also entails a change of measure. Thus if $\xi(x) > 1$ for $x < X$, this will mean that the equivalent logistic at X will have greater left hand tail probability to match the entropy of the subject distribution, so in this case ξ is a local inflator. For a normal distribution, on the other hand, the reverse is true in the more distant tails, and ξ will be a local deflator in such zones.

Scaling function values using the logistic reference could also become distribution diagnostics, of the same character as the common normal-based kurtosis or symmetry measures. Thus if the distribution scaling function has value $\bar{\xi}_L(X) > 1$ at a critical point X , this can be taken to mean that the distribution has a longer left hand tail at X than does the logistic, for then $F_q(X) > F(X)$. Or if the distribution is known to be symmetric, then $\xi(X_m) < 1$ at the median would tell us that the density needs to be inflated at that point to achieve the same directional half entropy as the logistic, so it is not as pointy headed as the logistic.

Similarly, $\xi(X_\alpha) > 1$ at a chosen VaR critical point for significance probability α would indicate a fat tail relative to the equivalent logistic.

As earlier remarked, in some cases the scaling function ξ can be negative over certain ranges. The problem, if it occurs, arises from very high values of the signal to noise ratio (expression (3) section 2.1), so that over the small interval dx , information rises very quickly from left to right, or falls too fast in the reverse direction; equivalently, uncertainty declines very quickly relative to the length of the interval. As expression (11) indicates, if $\xi(x) < 0$ over any zone, lower directional entropy can actually fall over such zones as x rises, which is counterintuitive. However, the point of the exercise is to obtain a logistic standardisation that has meaning in terms of the average uncertainty over the zone of interest. Thus for the asymmetric Gumbel distribution, another member of the extreme value family, negative ξ values occur at the short tail end, which is not typically the zone of interest for value at risk or investment purposes. Alternatively, the zone of negative values may be negligible in probability, as for the normal distribution. If the region in which $\xi(x)$ has negative values has very low measure, then outside this region, F_q as defined by expressions (13a,b) will be close to an L_q distribution function.

4. Applications to VaR and CVaR

The applications explored in this section make use of a convenient property of the logistic, namely that conditional expectations, and hence CVaR or option values, are expressible in terms of the relevant tail probabilities. In other words, for the logistic, VaR and CVaR are effectively just a rescaling of the same metric. This convenient property is not true for general distributions. However, the L_q measure allows us to translate to an equivalent logistic and in this way adjust the original distribution, even where the latter may not admit a CVaR calculation. The general idea is to give the directional entropy metric an interpretation either in terms of tail probabilities (VaR) or in terms of value (CVaR or related concepts). The ensuing development illustrates the methodology.

4.1 VaR rescaling

Value at risk by itself is an incomplete welfare metric in comparing return distributions that differ materially in their tail properties. Directional entropy captures the residual uncertainty in the tail. By converting the natural measure to equivalent logistic measure L_q , directional entropy becomes converted to an equivalent probability that does take into account different rates of decay in the tail past the VaR critical point. One asks what the VaR probability would

be for a logistic distribution that has the same directional entropy as the subject distribution at the given point. If the given distribution has a longer tail than the logistic benchmark, measured as directional entropy, this is reflected in an adjustment of one or other of the effective critical point or the effective probability.

Figure 5 illustrates for the Cauchy distribution earlier specified. A 90% Cauchy VaR means here a lower 10% critical point at -0.855. This has an L_q equivalent tail probability of 22.7% , more than double the risk for the local logistic, in probability terms. Alternatively, if one wanted to fix the probability, the L_q effective 10% critical point is now at -1.314 as compared with the original -0.855, a 54% leftward shift. Similarly, a 95% Cauchy VaR (-1.754) corresponds to an L_q equivalent of 14.20% ; or the L_q effective 95% VaR of -2.295, which is a 31% left shift. The correction gets proportionately more minor at the 99% level and beyond.

As figure 6 suggests, any VaR adjustment will be quite minor for a Student's t_4 distribution; one could say that the t_4 distribution is just slightly more long tailed than the logistic. In the case of distributions with less mass in their tail than the logistic benchmark, the adjustment will go the other way, i.e. the effective VaR point will move to the right instead of the left; the normal distribution is an example, though the adjustment is small in this case.

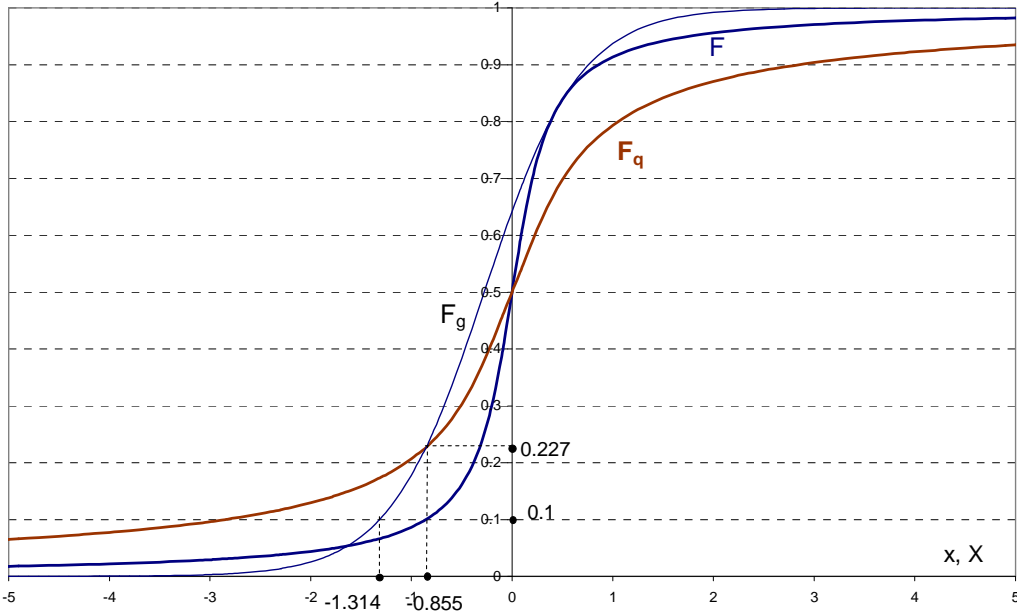


Figure 5: The Cauchy distribution: L_q equivalent and VaR adjustments

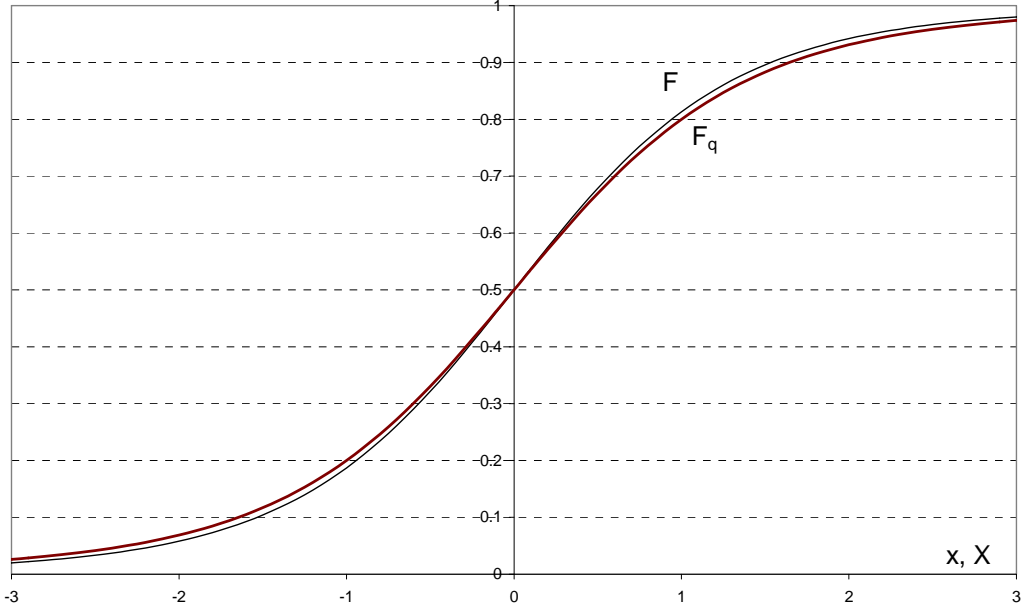


Figure 6: A t_4 distribution and its L_q equivalent

4.2 Conditional value at risk

The conditional value at risk (CVaR) refers to the expected loss, given that the VaR point (here X) has been breached:

$$CVaR(X) = E[x | x \leq X] = \frac{1}{F(X)} \int_{-\infty}^X xf(x)dx,$$

provided the expectation exists. It can be expressed as

$$CVaR(X) = X - \frac{\Phi(X)}{F(X)}; \quad \Phi(X) = \int_{-\infty}^X F(x)dx. \quad (15)$$

The function $\Phi(X)$ is the distribution accumulator familiar from stochastic dominance theory: distribution A is second order stochastically dominant (SSD) over distribution B if and only if $\Phi_A(X) \leq \Phi_B(X)$, for all X . The conditional value at risk is sometimes expressed in terms of

$$\text{the expected shortfall } X - CVaR(X) = \frac{\Phi(X)}{F(X)}.$$

Lemma 2: If $F(x)$ is logistic with scale parameter β , then for any marker value X ,

$$CVaR(X) = X + \frac{\beta \ln(1 - F(X))}{F(X)}. \quad (16)$$

Proof: Using expression (2) with $\lambda'(x) = -1/\beta$, $\Phi(X) = \int_{-\infty}^X \frac{\beta f(x)}{(1 - F(x))} dx = -\beta \ln(1 - F(X))$.

□

As with VaR, the logistic provides a possible benchmark. Given a marker value X , one can define a local conditional value at risk as that of the L_q equivalent at X . Combining expressions (14) and (16), this can be obtained as

$$CVaR_q(X) = X + \frac{e^{\kappa-2} \ln(1-F_q(X))}{F_q(X)}, \quad (17)$$

where κ is the parent distribution total entropy. The basic idea is to endow directional entropy with an equivalent value dimension. Figure 7 compares the actual and L_q equivalent CVaR's for the t_4 distribution. The correspondence for CVaR is not as close as it is for VaR (see figure 6) and is certainly not global over all values³ of X . It is closer over the region relevant for a CVaR calculation, in this case the lower tail.

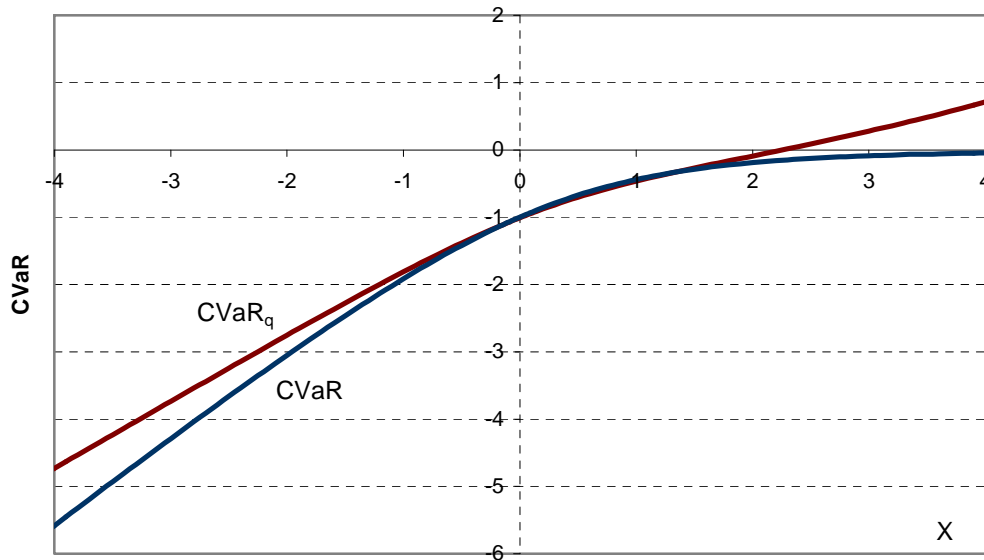


Figure 7: The $CVaR_q$ proxy for a t_4 distribution

If the primary welfare objective is couched in terms of the CVaR itself, there is in general no need to explore the $CVaR_q$ proxy, but it can be useful to do so in some circumstances. The first is where the underlying distribution has a very long tail that does not admit an actual CVaR, because the required conditional expectation may not exist, e.g. the Cauchy or Levy distributions. However, the L_q equivalent will continue to exist as $CVaR_q$, providing an embodiment to the directional entropy that has a conditional value at risk interpretation. Figure 8 illustrates for the Cauchy distribution. The CVAR proxy declines more or less linearly with the marker value X .

Another instance is where a portfolio manager is more agnostic that CVaR is the most appropriate or the only possible welfare metric; perhaps it is felt that higher order conditional moments are also of relevance, or on a more general level, the amount of information left in the tail. It is a reasonable conjecture that many, if not all, sensible welfare metrics are monotonically related to tail directional entropy. Thus if two alternative portfolios with similar upside potential are under consideration, the first of which dominates the second in lower tail directional entropy, then it might be better to choose the second.

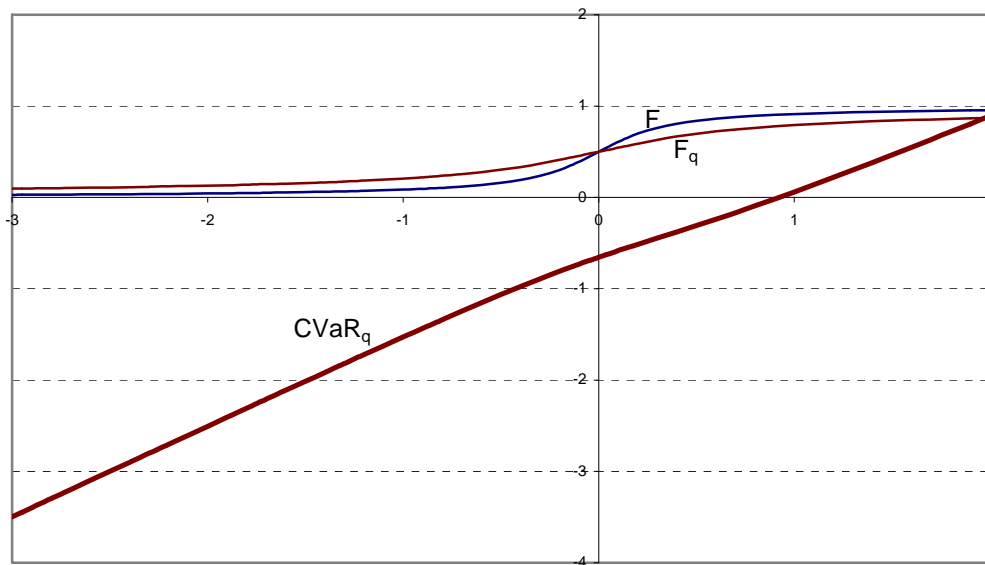


Figure 8: The CVaR proxy for a Cauchy distribution

5. Concluding remarks

The directional entropy concepts developed in the present paper provide a framework for assessing tail risk, and in particular a metric for tail length. Whether a given distribution has an intrinsically longer tail than another can be measured at a given marker point, in terms of the magnitude of the L_q correction and the distribution scaling factor. A measure change to an equivalent local logistic enables the user to assess the adequacy of tail probabilities such as VaR, as against more structured alternatives such as CVaR or higher order conditional tail moments. Some further remarks address matters of scope and operational procedure.

(a) Directional entropy can be viewed in a more dynamic setting where information is revealed through time. As fresh information comes to hand, this affects the values of financial assets whose future payoffs are linked in some way to the flow of news. Potential applications

of this kind occur in options pricing, particularly for the kind of real options that are implicit in contexts such as venture capital. One usually conceives of information as diminishing uncertainty, or entropy in the current context. This would mean that for any given strike price, a real option cast at any fixed strike price would move (further) either into or out of the money depending upon whether the new information was good or bad news for the option holder. The source of value in this case would be a progressive collapsing of the distributional spread.

But another possibility altogether is that the entropy could actually increase, in particular the directional entropy relative to the chosen strike price. Effects of this kind arise where the underlying distribution of possible payoffs is mixture distribution in character. Mixtures have been suggested as a good way to model asymmetric distributions (McLachlan and Peel 2000, also Suárez and Menéndez 2005 for financial data). Thus in venture capital, there are often two possible outcomes, corresponding to failure or else success for the project upon launch (initial public offering, IPO). If the project turns out to be a failure, the payoff range will be limited, and may even collapse to zero. If the project is a success, there is considerable blue sky potential, and the distribution of possible payoffs seen as of the time of the IPO has a long right hand tail. Most projects in venture capital fail, so the ex ante mixture distribution would be more heavily weighted towards the failure distribution. That information comes in stages over time has long been recognised in the practical pricing procedures adopted in the venture capital industry (Ruhnka and Young 1991, Chiampou and Kellett 1989). If the new information revises the mixing probabilities towards a success, then the reweighting opens up the blue sky outcome, and leads to an increase in the directional entropy of the right hand tail. In turn, this will increase the value of any option written out of the money with respect to the median of the original mixed distribution. It is possible to identify the strike price that will lead to a maximal shift in directional entropy and hence in potential option payoff. One could describe this as informational leverage.

(b) The logistic L_q benchmark is an infinite support distribution, which raises an issue as to the meaningfulness of using it as a benchmark for distributions with a naturally limited support, such as the exponential or Pareto. For such distributions, however, the risk management problem is typically at the long end rather than the short, which for the Pareto would be the right hand tail. In that case, the concern would shift to upper directional entropy rather than lower. The problem is in some respects the same as that referred to in section 3, where the scaling function can be negative over zones not of primary interest. For such purposes, what happens to the left of the marker point is virtually immaterial. One could

think in both cases of annealing the original distribution in the immaterial left hand region with a more regular one with the same total entropy, but without the scaling negativity problem. In effect one is operating from expressions (9b) and (13b) of section 3, so only values $x, f(x)$ for which $x > X$ are involved. The L_q standardisation to an upper VaR point, should this be the context, would remain meaningful.

(c) Concepts developed in the present paper do not appear to be any more demanding in their empirical implication than standard VaR. As noted in sections 3, transformation to equivalent logistic measure is accomplished via the lower or upper distribution scaling factors. An empirical estimate of the necessary scaling function $\xi(x)$ can be obtained by fitting a smoothing function e.g. via splines, to the empirical log odds table, and obtaining the derivative. Even if relatively few historical observations lie to the left of the desired VaR point, one can establish lower or upper limits for the desired conditional mean, and from this ascertain whether a significant correction needs to be made to value at risk probabilities.

(d) Applications to financial risk management have homologies with other areas of finance, arising from the interpretation of loss measures like CVaR as implied put options with payoffs in a financial distress zone to third parties such as competitors or liquidators. The value of expected shortfall options is readily computable for a logistic distribution in terms of the critical probability, together with the total entropy. This leads to potential applications of the L_q measure shift to the pricing of corporate bonds, which are commonly viewed as compound instruments made up of risk free bonds and put options written by the firm's debt holders.

References

- Alexander, G.J., and A.M. Baptista (2004), A comparison of VaR and CVaR constraints on portfolio selection with the mean-variance model, *Management Science*, 50, 1261-73.
- Alexander, G.J., and A.M. Baptista (2006), Does the Basel Accord reduce bank flexibility? An assessment of the value at risk approach, *Journal of Monetary Economics*, 53, 1631-60
- Artzner, P., Delbaen F., Eber J.M., and D. Heath (1999), Coherent measures of risk, *Mathematical Finance*, 9, 203-28.
- Berkowitz, J. and J. O'Brian (2002), How accurate are value at risk models at commercial banks?, *Journal of Finance*, 59, 1093-1111.

- Bertsimas, D., Lauprete G.J., and A. Samarov (2004), Shortfall as a risk measure: Properties, optimization and applications, *Journal of Economic Dynamics and Control*, 28, 1353-81.
- Bowden, R. (2005), Ordered mean difference benchmarking, utility generators, and capital market equilibrium, *Journal of Business*, 78, 441-67.
- Bowden, R. (2006), The generalized value at risk admissible set: constraint consistency and portfolio outcomes, *Quantitative Finance*, 6, 159-171.
- Bowden, R.J. (2007), Information, measure shifts and distribution diagnostics, working paper, School of Economics and Finance Victoria University of Wellington.
- Chiampou, G.F. and J.J. Kellett (1989), 'Risk return profile of venture capital, *Journal of Business Venturing*, 4, 1-10.
- Duffie, D. (1992), *Dynamic Asset Pricing Theory*, New Jersey: Princeton University Press.
- Glasserman, P., Heidelberger, P. and P. Shahabuddin 2002, Portfolio value at risk with heavy-tailed risk factors, *Mathematical Finance*, 12, 239-69.
- Harrison, M. and D. Kreps (1979), Martingales and arbitrage in multiperiod securities markets. *Journal of Economic Theory*, 20, 381-408.
- Johnson, N.L., Kotz, S. and N. Balakrishnan (1994), *Continuous univariate distributions* Vol. 1, New York, Wiley.
- Jorion, P. (2007), *Financial risk manager handbook*, 4th ed., Hoboken NJ: Wiley/GARP.
- Kullback, S. (1968), *Information theory and statistics*, New York: Dover.
- Kullback, S., and Leibler, R. A. (1951), On information and sufficiency, *Annals of Mathematical Statistics* 22: 79-86.
- McLachlan. G. and D. Peel (2000), *Finite mixture models*, New York: Wiley.
- Palmquist, J., P. Krokmal and S. Uryasev (2001), Portfolio optimization with conditional value-at-risk objective and constraints, *Journal of Risk*, 4, 21-41.
- Pinsker, M.S. (1964), *Information and information stability of random variables and processes*, San Francisco: Holden Day.
- Rockafellar. R. and S. Uryasev (2002), Conditional value at risk for general loss distributions, *Journal of Banking and Finance*, 26, 1443-1471.

- Ruhnka, J.C. and J.E. Young (1991), Some hypotheses about risk in venture capital investing, *Journal of Business Venturing*, 6, 115-33.
- Scaillet, O. (2004), Nonparametric estimation and sensitivity analysis of expected shortfall, *Mathematical Finance*, 14, 115-29.
- Suàrez, S. and S.C. Menéndez (2005), Computational tools for the analysis of market risk, in Menéndez, S.C., A.S. Callè and L. Seco (eds) *Risk management in Finance*, Proc. First Risk Lab international conference, Bilbao: Fundación B.B.V.A.
- Tokat, Y., Rachev, S.T. and E.S. Schwartz (2003), The stable non-Gaussian asset allocation: A comparison with the classical Gaussian approach, *Journal of Economic Dynamics and Control*, 27, 937-69.
- Uryasev, S. (2000), Conditional value-at-risk: Optimizing algorithms and applications, *Financial Engineering News*, 14, 1-5.
- Yamai, Y. and T. Yoshihara (2002), Comparative analyses of expected shortfall and value-at-risk (2): expected utility maximization and tail risk, *Monetary and Economic Studies*, 20, 2, 96-116.

Endnotes

¹ The convention in statistics is to refer to a distribution that contains both discrete and continuous values (jumps in $F(x)$) as ‘mixed’; while linear combinations of distribution functions are said to generate ‘mixtures’ of distributions.

² The two components are not the same as the two half integrals implicit in the expression (4) for differential entropy. The relationship is given by

$$-\int_{-\infty}^X f(x)\ln(f(x))dx = \kappa_L(X) + (1-F(X))\ln(1-F(X)) - F(X)\ln F(X),$$

with a complementary definition for

the other half integral $-\int_X^{\infty} f(x)\ln(f(x))dx$. Likewise, $\kappa_L(X)$ is not the same as the differential entropy

of the conditional distribution for $x \leq X$.

³ Intuitively, for X values greater than the median, the local logistic has to be centred well to the right of the median so that $F_g(X)$ at X coincides with $F(X)$. This means that the CVaR for F will contain values of X to the right of the median, to an increasing extent as X becomes larger.