

Directional entropy and tail uncertainty, with applications to financial hazard and investments

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

Contexts such as value at risk or venture capital require local uncertainty measures, as distinct from properties of the entire distribution such as differential entropy. Applications such as value at risk and options pricing can be illuminated by means of a regime-specific concept of directional entropy. The latter enables a change of measure to an equivalent logistic distribution, one that has the same total and directional entropies at the given marker, e.g. value at risk critical point or option strike price. This is done via a scaling function that can be interpreted as a Radon-Nikodym derivative and used in its own right as a risk metric. Value at risk rescaling adjusts the critical probability to capture the long tail risk. 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: Differential entropy, discovery processes, expected shortfall, logistic distribution, mutual information, value at risk rescaling, venture capital.

* 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.

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 e.g. the Levy, because the relevant moments do not exist. The problem arises in contexts such as value at risk, or at the other extreme, venture capital and related high risk investments.

As an instance, value at risk (VaR) has become a favoured tool of financial regulators, originally in bank capital adequacy, more recently extending into 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 just 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; welfare metrics should incorporate this contingency in some way. In other words, the VaR probability underestimates the sting in the tail. Conditional tail moments have been suggested as a criterion. Thus expected shortfall, or conditional value at risk, refers to the censored mean in the tail and has become popular in insurance circles (e.g. Uryasev 2000, Palmquist et al 2001, Rockafellar and Uryasev 2002, Alexander and Baptista 2004). Other proposals have suggested specific higher order moments (e.g. Yamai and Yoshida 2002). On the other hand, Bowden (2006) has pointed out an inconsistency between value at risk and conditional value at risk. Moreover, the required moments must actually exist, which is not automatically the case.

Venture capital and related investments such as patents are oriented towards the other end of the distribution – the right hand tail, for traditional call type structures. Options and warrants are obvious examples, and indeed, venture capital investments are very often structured in just this form. The task becomes one of identifying the investments that do have tail information content and structure, and of designing optimal exposures in the light of that structure. A problem here is just where to set the strike price for maximal leverage. Of relevance is the flow of information. That information comes in stages over time has long been recognised in the practical pricing procedures adopted in the venture capital industry (e.g. Ruhnka and Young (1991), Chiampou and Kellett (1989) and Sahlman (1990)). At each

of these stages, information about the product and its prospects will accumulate and the value of options in hand will correspondingly change, at times rapidly. So one needs to set strike values that will result in maximal value change with respect to information changes. It may be noted that discovery processes help to determine the structure of the tail as well as its length, so it should not be assumed in what follows that the tail to a distribution is necessarily monotone decreasing.

Applications such as value at risk, venture capital, or patent investment, 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. As earlier noted, 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 these may be applied to some common problems in financial economics and risk management. A checklist of contributions is as follows:

- (a) 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, once weighted by regime probabilities.
- (b) Directional entropy is the entropy concept most closely related to contexts such as value at risk, 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. The scaling factor can itself be used for metric purposes e.g. for kurtosis or asymmetry, as an alternative to the standard metrics based on the normal distribution.
- (c) The conversion to logistic-equivalent measure (L_q) allows rescaling of tail probabilities in a way that is useful in value at risk and related contexts. If a distribution of investment outcomes has a standard 10% 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 20% , say,

it will suggest that the natural VaR probability (10%) is misleading as a hazard indicator and a more conservative stance should be taken. Because directional entropy is monotonically related to most conceivable tail metrics, it becomes an effective hazard indicator once converted to equivalent probability terms via the L_q measure.

(d) Much the same principle applies 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. 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.

(e) Perhaps contrary to usual ways of thinking, the passage of time can actually increase directional entropy, not diminish it. Mixture distributions are a good way to think about discovery processes involved. In terms of venture capital usage, if the prior distribution consist of a mixture of a pessimistic ‘No-go’ and a blue sky ‘Go’ distribution, then favourable news is likely to increase ex post entropy, with favourable value outcomes for an option with a correctly set strike price.

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. The core idea of directional entropy is developed in section 3, which introduces the change of measure via the scaling function. Critical probability adjustment in value at risk is an immediate application in section 4, and there is a further application to indicative option pricing. Section 5 provides an informal discussion of a more dynamic context involving progressive information revelation. Mixture distributions provide a good way of understanding how entropy evolves and the way that this generates financial value. Section 6 offers some concluding remarks.

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

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

Written in the form

$$-\lambda'(x) = \frac{f(x)}{F(x)(1-F(x))}, \quad (2b)$$

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. 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 and Kotz (1970)). 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.

Although the logistic kurtosis coefficient is higher than that for the normal distribution, later analysis shows that in fact the logistic is not too far removed from the normal: its tails are fatter, but far from obese. By way of contrast, consider another member of the extreme value family, the Gumbel (or Fisher -Tippett) distribution. This is often used in statistical reliability theory, where part failure, if it occurs, is more likely early on, and once this 'probationary period' is over, survival time is likely extended. The density is therefore asymmetric with a longer right hand tail. An analogy exists with venture capital, where the 'Go' signal corresponds to early survival, and the blue sky prospects are enhanced. The Gumbel distribution and density functions are defined in terms of location and scale parameters μ, ϕ by

$$F(x) = \exp(-e^{-\tilde{x}}); f(x) = \frac{1}{\varphi} \exp(-\tilde{x} - e^{-\tilde{x}}) = \frac{1}{\varphi} e^{-\tilde{x}} F(x); \quad \tilde{x} = \frac{x - \mu}{\varphi}. \quad (3)$$

The mean is given by $\mu + \gamma\varphi$, where $\gamma \approx 0.57721$ is the Euler-Mascheroni constant. The mode is at μ , the median is given by $\mu - \varphi \ln(\ln 2)$, and the standard deviation is $\frac{\pi}{\sqrt{6}}\varphi$. The Gumbel log odds function $\lambda(x)$ is a diminishing convex function, asymptotically tangent from above as $x \rightarrow \infty$ to the line $\lambda = -\frac{x - \mu}{\varphi}$, so that $\lim_{x \rightarrow \infty} \lambda'(x) = -\frac{1}{\varphi}$. An more extreme example of a distribution with a long right hand tail is the Levy, for which $\lim_{x \rightarrow \infty} \lambda'(x) = 0$, so the log odds decay even less rapidly.

The logistic, normal and Gumbel distributions will be used throughout for illustrative comparisons. Figure 1 depicts the log odds functions for three distributions, the normal, logistic and Gumbel, all standardised to have the same Shannon entropy and median. The log odds derivative decays faster for the normal distribution indicating a shorter right hand tail. For the Gumbel the log odds rise quite quickly going away from the median in the left hand direction, indicating a short tail in this zone, but a much slower decay rate to the right.

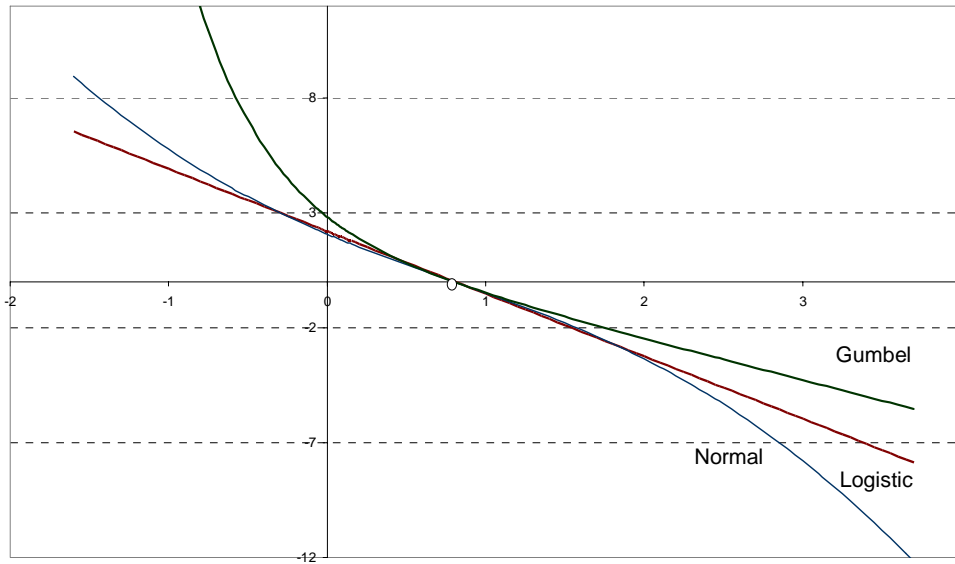


Figure 1: Comparative log odds functions and their decay.

2.2 Differential entropy

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)). Conditional and mutual entropy refer to the gains in information arising from associated information (Pinsker (1964)). Such extensions also figure in the present treatment, but the initial development runs in terms of the parent differential entropy concept, 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 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 a ‘scaling function’ in what follows. 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. Figure 2 illustrates for the normal, Gumbel and logistic, all benchmarked to have the same differential entropy $\kappa = 1.25$. Note that $\xi(x)$ 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 (2b). For skewed distributions like the Gumbel, this will occur in the shorter left hand tail, associated with very low cumulative probabilities at that point.

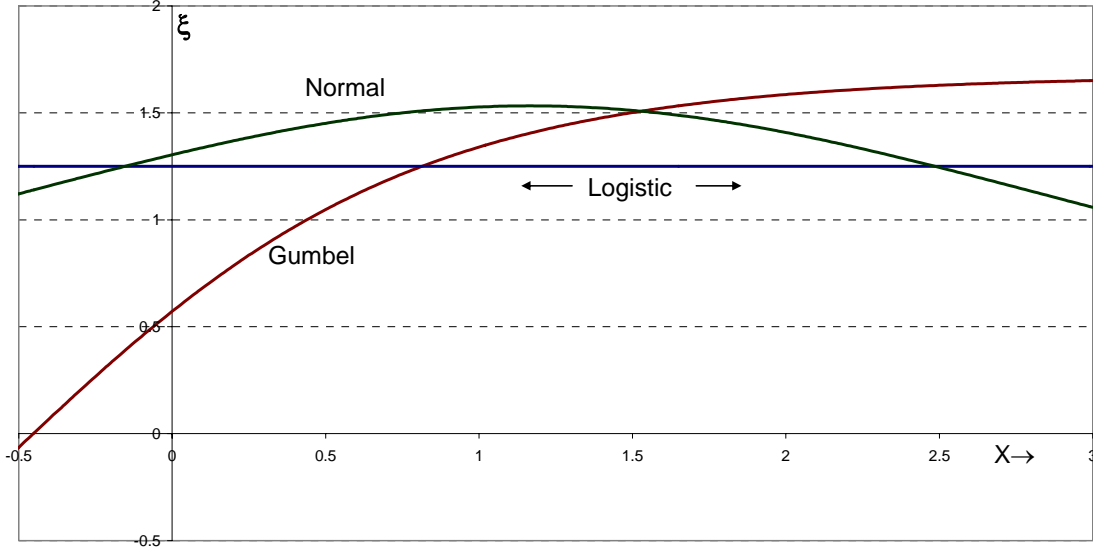


Figure 2: Scaling functions ξ compared

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 value at risk or options pricing 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 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 such as the Levy distribution. 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 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:

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

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

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

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).

□

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

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

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.

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) For a logistic distribution, $F_q(x) = F(x)$.

Proof: Note that ξ is non-negative and $E[\xi] = 1$, so that ξ qualifies as a Radon-Nikodym derivative on \mathbb{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. Part (b) follows because for a logistic, $\bar{\xi}_L(X) = 1$, for any marker value X .

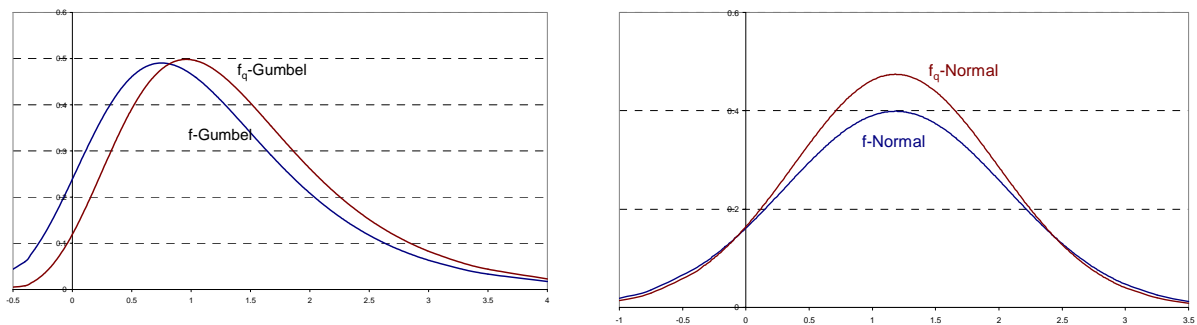
□

In what follows, F_q will be referred to as the distribution function for L_q measure or simply as the L_q distribution function, where L_q refers to ‘logistic equivalent’. 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 . In other words, one is effectively locating the two logistic degrees of freedom, namely β and μ , by the requirement that 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 . The idea has similarities² to an envelope relationship, although the local logistics are not in general tangent. If the underlying distribution is logistic to begin with, then F and F_q are one and the same. It may also be noted that if the parent density $f(x)$ is symmetric, then so is $f_q(x)$. This follows because for a symmetric distribution $F(X_m + \delta) = 1 - F(X_m - \delta)$, so that the log odds function $\lambda(x)$ is symmetric about the median X_m .

The function ξ plays a role very similar to the state price deflator of financial asset pricing, which also entails a change of measure. Because it can either deflate or inflate, it will be referred to in the current context as a scaling function. Thus if $\xi(x) > 1$ for $x > X$, this will mean that the equivalent logistic at X will have greater 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. For instance, if the distribution is known to be symmetric, then $\xi(X_m) > 1$ would tell us that it needs to be inflated at that point to achieve the same half entropy as the logistic,

so it is not as pointy headed as the logistic. Figure 4b illustrates. Or $\xi(X_\alpha) > 1$ at a chosen VaR critical point for significance probability α would indicate a fat tail relative to the equivalent logistic.

In some cases the scaling function ξ can be negative, if only over certain ranges. The problem, if it occurs, arises from very high values of the signal to noise ratio (expression (2b) 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 Gumbel distribution, 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. For the standard normal distribution, $\xi < 0$ at approximately $\tilde{X} = \pm 7.25604$, at which point $\tau_{L,U} = 1.5166E-23$. Thus in practice, the requirement that ξ be a non-negative function is not as restrictive as it might appear. The general issue is revisited in section VI. 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; one could call it ‘quasi- L_q ’. The quasi- L_q approximation has been employed in figures 4a,b, which compare the natural and L_q versions for both normal and Gumbel distributions, standardised to have the same total entropy. The symmetry of the normal density is reflected in that of the companion density. For the normal, the figure also indicates that the tail is restricted relative to an invariant. For the Gumbel, which is asymmetric with a long right hand tail, the companion density is shifted bodily to the right.



Figures 4a,b: L_q and natural densities compared

4. Applications to value at risk and option value

The applications that follow, in this and the next section, make use of a convenient property of the logistic, namely that expected shortfall or option values are expressible in terms of the relevant tail probabilities. This means that for the logistic distribution, applications such as value at risk and expected shortfall amount have a proxy relationship with each other. In a risk management context, it would add little to perform an expected shortfall, simply because the value at risk is an adequate substitute. This convenient property is not true for general distributions. However L_q measure allows us to translate to an equivalent logistic and in this way adjust the original. The ensuing development illustrates the methodology.

4.1 VaR rescaling

With long tail distributions, the measured value at risk is in general an understatement of the potential losses and a better proxy is based on the directional entropy. By converting the natural measure to equivalent logistic measure L_q , directional entropy becomes converted to an equivalent probability for value at risk purposes.

The technique can be illustrated with a reverse Gumbel, meaning one with density reflected backwards around the original median to capture a long left hand tail. In figure 5, the natural and L_q distribution functions are plotted on the same diagram. The lower 10% VaR point for the natural distribution is marked as $X_{10\%}$. Also marked is the lower 10% point $X_{q,10\%}$ for the companion distribution, which is the VaR point for the L_q distribution. This means that at $X_{q,10\%}$, the equivalent logistic would have the same directional entropy. Now as earlier remarked, directional entropy for a logistic is a satisfactory proxy for tail probability. This suggests that a correction is going to be necessary for the Gumbel, and the effective 10% VaR critical point is really $X_{q,10\%}$ rather than $X_{10\%}$. Alternatively, if the critical point is fixed at $X_{10\%}$, this would correspond to an equivalent tail probability of 21.8%, i.e. the scaling factor $\bar{\xi}_L$ is about 2.18 at the natural VaR critical point.

A similar construction applied to the lower tail of the normal distribution indicates that the normal and its logistic equivalent companion distributions are much closer together than for the Gumbel. For total entropy corresponding to that used in figure 5, the 10% normal VaR probability has to be lowered to about 8%, indicating that the normal-based VaR is a slight overestimate of directional entropy relative to the logistic benchmark.

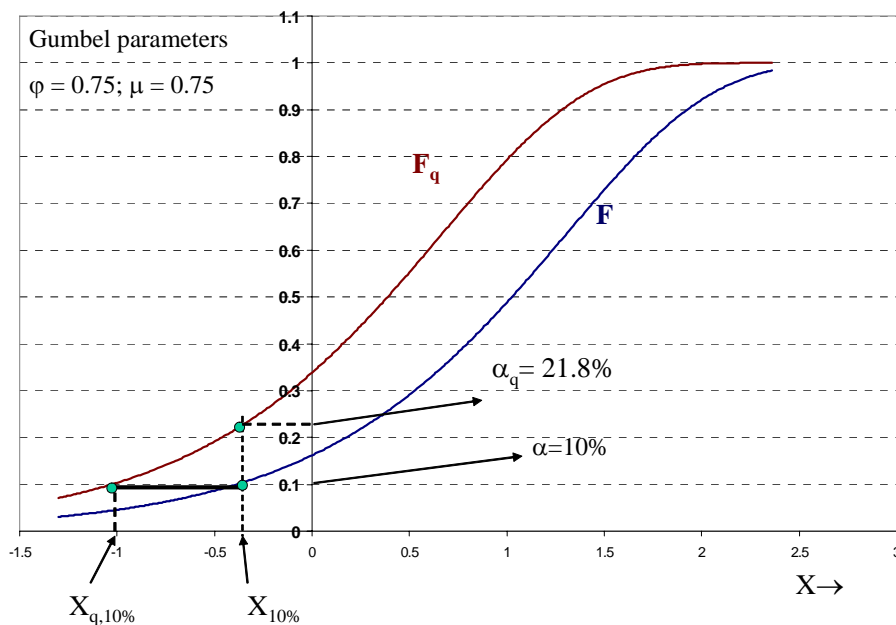


Figure 5: Adjusting value at risk via L_q measure

4.2 Indicative option value

Where distributions are badly asymmetric or fat tailed, directional entropy, and hence L_q measure, gives a better guide than does tail probability to the creation of investment value or that of similar location-dependent investments. Distributions that have higher upper directional entropy are likely to provide more productive opportunities for writing or investing in call options and related instruments such as warrants. In what follows, conventional call options are used extensively for illustrative purposes. These depend upon payoffs beyond the strike price, as distinct from binary or lump sum options. Locational entropy, which depends only on tail probability, would provide a better information-based guide to the value of instruments of the latter type.

Let x denote a value or return outcome for an underlying investment on which options or warrants are to be written. The function defined by

$$\pi(X) = \int_X^{\infty} (x - X) f(x) dx = (1 - F(X))(E[x | x > X] - X). \quad (14)$$

may be described as the core value for a single-period European call option with strike price or return value at X , relative to a designated probability measure with distribution function $F(x)$. The latter could refer any one of natural, risk-neutral (Harrison and Kreps (1979)), or even subjective probability measures, as in Kahneman and Tversky (1979), depending on the

application. For expositional convenience, time is discrete and the usual interest rate or discount factor is disregarded (it can otherwise be accommodated by adjusting parameters³).

The core option value can also be written

$$\pi(X) = \mu - X + \Phi(X) \quad (15)$$

where $\mu = E[X]$ and $\Phi(X) = \int_{-\infty}^X F(x)dx$. 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. However, it is quite possible to have distribution A SSD over distribution B, yet call options written on B are more valuable. There is indeed an options theory embedded in SSD but it more naturally refers to put options on the downside (Bowden 2005).

Once again, the logistic provides a natural benchmark:

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

$$\Phi(X) = -\beta \ln(1 - F(X)); \quad (16a)$$

$$\pi(X) = -\beta \ln(F(X)). \quad (16b)$$

Proof: Using expression (2), with $\lambda'(x) = -1/\beta$, we get

$$\Phi(X) = \int_{-\infty}^X F(x)dx = \int_{-\infty}^X \frac{\beta f(x)}{(1 - F(x))} dx = -\beta \ln(1 - F(X)), \text{ which is part (a). Part (b) follows by}$$

substituting in expression (15) and using $\tilde{X} = \ln(e^{\tilde{X}})$; $\tilde{X} = (X - \mu)/\beta$.

□

The above suggests that a reasonable guide to option value for an arbitrary distribution might be to find the equivalent logistic-equivalent (L_q) distribution as in section 3.2 and use expression (16b). The relationship with directional entropy and option value is not necessarily linear; one is simply saying that the more entropy there is in the tail, the higher option value is likely to be, i.e. the general relation is monotonic. In fact, the correspondence can be quite close, especially in the tail areas. Figure 6 is based on the Gumbel distribution, and compares:

- (i) The exact option price, based on the Gumbel; and
- (ii) An option price derived by using the L_q distribution and formula (17b), with $2 + \ln \beta$ set equal to the Shannon entropy κ of the Gumbel as in section 2.2.

The correspondence is remarkably close for strike prices greater than the median X_m , but diverges for smaller strike prices, though the relationship between prices (i) and (ii) remains monotonic. Section 5 discusses how this might be used.

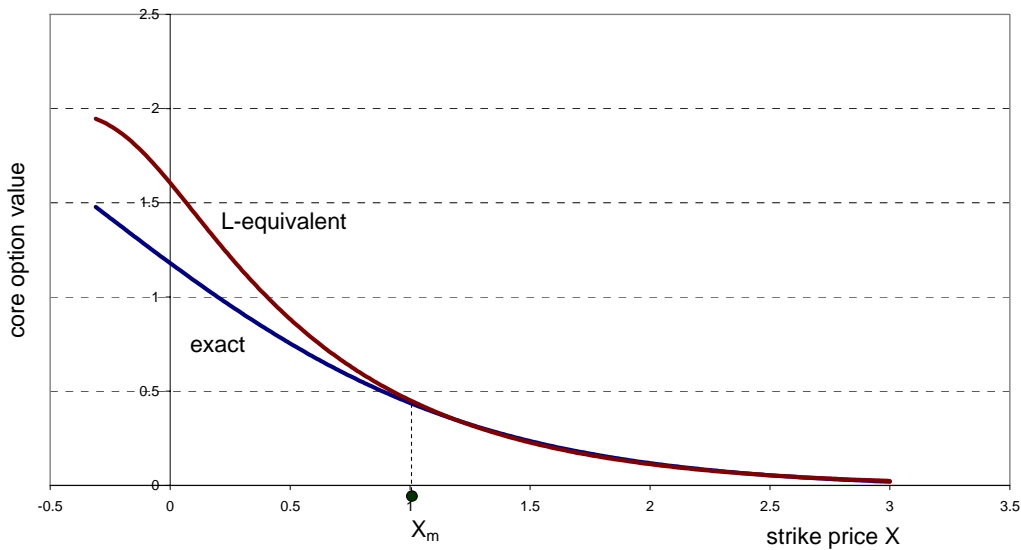


Figure 6: Exact versus L_q -equivalent option price: Gumbel distribution

5. Discovery processes and value revelation

Discussion in this section turns to a more dynamic context in which information is revealed through time. A simple framework is a two period discovery process, represented by a mixed distribution. At time zero, the investor in a venture capital does not know which of two prospects or states will be true. If bad, the project has little value; if good, it may have considerable value or again may not, so that a greater spread exists in the good state. Analogies exist with value at risk on the downside. Thus a bank asset and liability manager may think in terms of normal business continuing, or on the other hand, the existence of some unknown event that could cause a radical change in bank capital, perhaps extending to a disaster scenario as an extreme value. The framework is used first to re-examine locational and directional entropy in terms of bivariate entropy, then to show how the latter evolves and financial value along with it.

Mixture distributions provide a convenient framework for thinking about discovery processes. Suppose that the distribution of the variable x of primary interest depends upon that of another, α . Thus there might be two possible states of the world or latent properties of the venture, ‘good’ and ‘bad’ say, and the realised value of x would depend upon which turned out to hold. In this case α would be a categorical variable with two possible states g, b . In

state g , the relevant joint density element is $f(x, g) = p(g)f(x | g)$; similarly for state b . The marginal distribution is given schematically by

$$f(x) = f(x, g) + f(x, b) = p(g)f(x | g) + p(b)f(x | b); p(g) + p(b) = 1,$$

so that $p(g), p(b)$ play the role of mixing parameters. Mixtures have been suggested as a systematic way to model asymmetric distributions; see McLachlan and Peel (2000).

The mixing model can be applied to analyse the impact of information upon financial value. One usually thinks of entropy as diminishing⁴ with information. In the extreme case, a final value is revealed and options that have been written at what is ex post now in the money, as a favourable outcome, will have generated a positive return on their initial value. In the stages approach to venture capital, the final value may not yet be revealed, but at the next stage the distribution may have shifted to the right and/or become more concentrated. The options will likely be worth more, particularly if they have been written at a point to the left of the median of the ex post distribution.

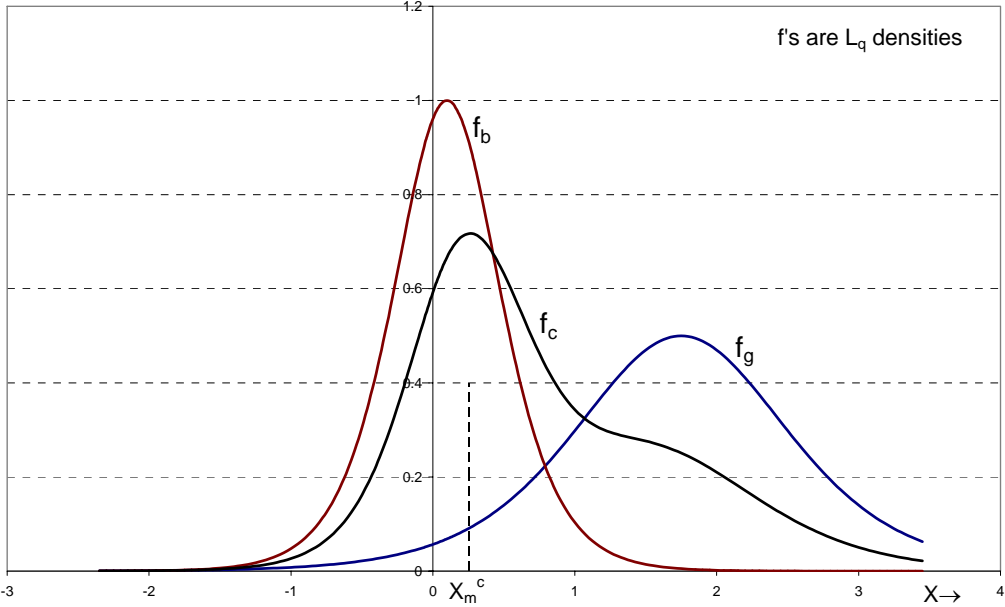


Figure 7: Component and combined L_q densities

However, option value can be generated also where locational entropy has increased following the revelation of information. If bad outcome has at the outset been viewed as more probable, this narrows the ex ante distribution relative to what happens if the good state is revealed. Figure 7 illustrates a case where the ex post distribution has *more* total entropy than the ex ante distribution. The ex ante distribution mixes logistic good (g) and bad (b) outcomes to end up with a combined (c) distribution that is not itself logistic, but has a longer right hand tail. The L_q version of the respective densities has been depicted in each case, as a

better guide to option value. For the parent logistics (b and g) the natural and L_q densities coincide, but the L_q density for the mixed (c) is now shifted to the right relative to the natural measure, which is why the natural combined median, marked in as X_m^c , looks wrong in the picture. Any option that has been initially priced off the ex ante density at a strike price a little to the right of X_m^c will have generated considerable increase in value if the new information indicates that the good state holds.

The gains in directional entropy in the good state are illustrated explicitly in figure 8, based on parameters as in table 1. There are two core possibilities as seen at time zero, a good state or the world, and a bad state, with the probability of the good state seen as 0.25, which becomes the mixing parameter. The distributions in each case are logistic; the bad state has a lower median μ and also a lower dispersion parameter β . The good state is intended to capture a more blue sky outcome, with a higher median and dispersion, and a higher total entropy, than for the bad state.

Base distributions	
g = good; b = bad.	
μ_g	0.8
β_g	0.55
μ_b	0.1
β_b	0.25
Mixing probabilities	
θ (good)	0.25
Total entropies (κ)	
g	1.402163
b	0.613706
Mixed (c)	0.9465

Table 1: Data for figure 8

The initial distribution mixes the good and bad outcomes with probabilities 0.25, 0.75 to create a mixture marginal distribution for x ($c =$ combined). This has total entropy intermediate between the two. The upper directional entropy is denoted by $\kappa_{U'}^c(X)$. Suppose now that time has moved on, and it becomes apparent that it is the good state that will hold. The upper directional entropy has changed to $\kappa_{U'}^g(X)$. As one might expect, the curve has shifted to the right. Note however that the gain is not uniform – it has a maximum around $X = 0.5$. This approximately coincides with a point of inflection in the graph of gain against strike price for options written on the ex ante (combined) distribution.

That information could increase uncertainty is counterintuitive at first sight. In fact, this may be just what happens for many ‘blue sky’ ventures. Investors think that should the project be successful, the prospective gains might have an extended range of possible values, ranging from the moderate to very large. So the ‘success’ contingent distribution has high total entropy. But in the early stages, the probability distribution is dominated by a prospect of failure (‘No Go’). This constricts the probability distribution. If a ‘Go’ signal becomes revealed, then the distribution broadens out into the blue sky prospect.

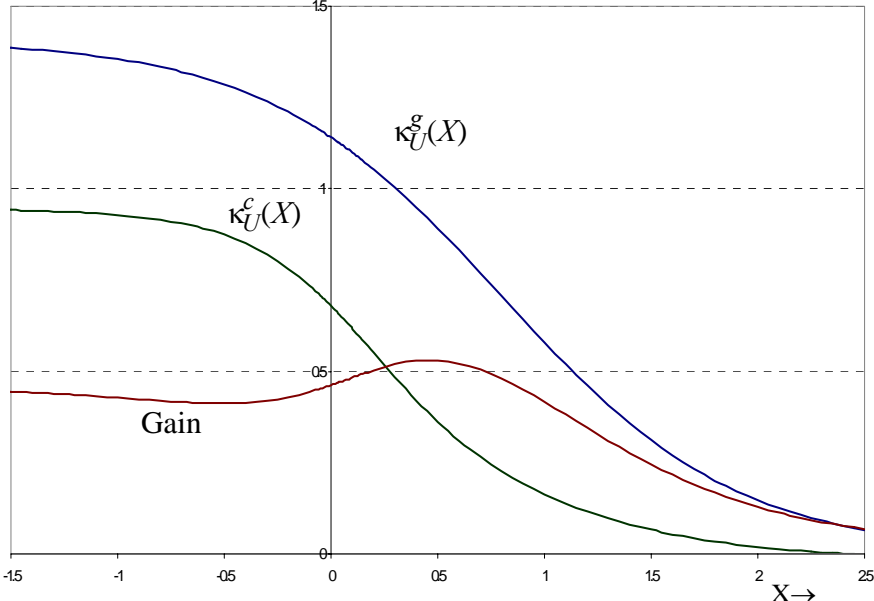


Figure 8: Directional entropy gain function

Finally, the mixed distribution framework as above is not the only possible model of distributional mixtures. The Appendix develops a model based on the logistic distribution in which the log odds of the initial distribution are a mixture of good and bad state odds, instead of the densities themselves. If the new information favours a good outcome (‘Go’), then the effective dispersion parameter β rises, as does directional entropy and option value.

6. Concluding remarks

The directional entropy concepts developed in the present paper provide uncertainty metrics suitable for the longer tail distributions that have become a focus of concern for risk management and similar contexts. The associated framework of measure change to an equivalent local logistic enables the user to assess the adequacy of tail probabilities such as value at risk, as against more structured alternatives such as expected shortfall or higher order

tail moments. The magnitude of the VaR rescaling factor tells just how sad the tail could potentially be. Transforming to L_q measure also enables one to see at a glance the option value potential or to understand how it might be generated. The following remarks elaborate on some of the practicalities.

(a) 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. With 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 3.2, 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 3.2, 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.

(b) Concepts developed in the present paper do not appear to be any more demanding in their empirical implication than standard value at risk, as a leading application. As noted in sections 3.2-3, transformation to equivalent logistic measure is accomplished via the lower or upper conditional mean functions $\bar{\xi}_L(X), \bar{\xi}_U(X)$. 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 outside (e.g. 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 significant correction needs to be made to value at risk probabilities.

(c) Applications to risk management and to investment have homologies arising from the interpretation of loss measures like expected shortfall 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 (lemma 2, section 3.4). Thus for a general distribution, tail risk metrics based on expected shortfall reduce to a value at risk, once

conversion is made to the L_q -equivalent logistic. It is in this sense that corrections to the value at risk succeed in capturing expected shortfall elements.

Appendix: Alternative mixtures, based on log odds (section 5)

Another possibility is to assume that it is the log odds that are a mixture, and not the densities. Assume both good and bad state distributions are logistic. Thus the investor could assess the prior distribution as having log odds given by

$$\lambda_c(X) = \theta_g \lambda_g(X) + \theta_b \lambda_b(X); \quad \theta_g, \theta_b \in (0,1); \quad \theta_g + \theta_b = 1,$$

where the subscript c denotes the combination and θ_g, θ_b are probabilities. The combined prior or ex ante distribution continues to be logistic, but with dispersion parameter given by the harmonic means of those for the two regimes:

$$\frac{1}{\beta_c} = \theta_g \frac{1}{\beta_g} + \theta_b \frac{1}{\beta_b} \quad ;$$

$$\mu_c = \left(\frac{\theta_g / \beta_g}{\theta_g / \beta_g + \theta_b / \beta_b} \right) \mu_g + \left(\frac{\theta_b / \beta_b}{\theta_g / \beta_g + \theta_b / \beta_b} \right) \mu_b .$$

Once the new information is at hand, the distribution collapses to one or other of the two regimes. Assume it is the good state: thus $\beta_c \rightarrow \beta_g; \quad \mu_c \rightarrow \mu_g$. This has the effect of raising β (increasing entropy) and also the median μ , shifting the distribution to the right. Both are potent sources of option value. In particular, an option written ‘out of the money’ with respect to the initial combined median will show very high value growth.

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, 1261-1273.
- 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.
- Harrison, M. and D. Kreps (1979), Martingales and arbitrage in multiperiod securities markets. *Journal of Economic Theory*, 20, 381-408.
- Kahneman, D. and A. Tversky (1979), Prospect theory: An analysis of decision making under risk, *Econometrica*, 47, 263-292.
- 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.
- Uryasev, S. (2000), Conditional value-at-risk: Optimizing algorithms and applications, *Financial Engineering News*, 14, 1-5.
- Yamai, Y. and T. Yoshiba (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 function are said to generate ‘mixtures’ of distributions.

² More precisely, for any given marker value X , we find a logistic distribution F^l say, such that $F^l(X; \mu, \beta) =$

$\kappa_L(X)$. We set β to give the required total entropy κ , i.e. $\beta = e^{\kappa-2}$ so the condition becomes $F^l(X; \mu, e^{\kappa-2}) =$

$\kappa_L(X)$. Solving gives $\mu = \mu(X)$. So the local logistic is $F^l(x; \mu(X), e^{\kappa-2})$. Then $F_q(X) = F^l(X; \mu(X), e^{\kappa-2})$.

³ For instance in the logistic option formula (17a), simply replace β by $\beta/(1+r)$, where r is the chosen discount rate for an option with maturity one period, if that is the holding horizon.

⁴ In contrast with the second law of thermodynamics in physics, which says the entropy of a closed system always increases through time.