Bank of Canada Banque du Canada

Working Paper 2002-11 / Document de travail 2002-11

Risk, Entropy, and the Transformation of Distributions

by

R. Mark Reesor and Don L. McLeish

ISSN 1192-5434

Printed in Canada on recycled paper

Bank of Canada Working Paper 2002-11

April 2002

Risk, Entropy, and the Transformation of Distributions

by

R. Mark Reesor[1] and Don L. McLeish[2]

[1]Financial Markets Department
Bank of Canada
Ottawa, Ontario, Canada K1A 0G9
mreesor@bankofcanada.ca

[2]Department of Statistics and Actuarial Science
University of Waterloo
Waterloo, Ontario, Canada N2L 3G1
dlmcleis@icarus.math.uwaterloo.ca

Contents

Acknowledgements

We would like to thank Adam Kolkiewicz, Mary Hardy, and Harry Panjer of the University of Waterloo for their helpful comments and assistance. We would also like to thank Dave Bolder of the Bank of Canada and Jeremy Graveline of Stanford University. The authors acknowledge financial support from the National Science and Engineering Research Council. As usual, all remaining errors are our responsibility.

Abstract

The exponential family, relative entropy, and distortion are methods of transforming probability distributions. We establish a link between those methods, focusing on the relation between relative entropy and distortion. Relative entropy is commonly used to price risky financial assets in incomplete markets, while distortion is widely used to price insurance risks and in risk management. The link between relative entropy and distortion provides some intuition behind distorted risk measures such as value-at-risk. Furthermore, distorted risk measures that have desirable properties, such as coherence, are easily generated via relative entropy.

JEL classification: C0, C1, D8, G0

Bank classification: Econometric and statistical methods; Financial markets; Market structure and pricing

Résumé

Classification JEL : C0, C1, D8, G0

Classification de la Banque : Méthodes économétriques et statistiques; Marchés financiers; Structure de marché et fixation des prix

1 Introduction

A considerable variety of applications use reweighted probability distributions. These applications come from areas such as statistics, financial economics, insurance, and risk management. Not surprisingly, many methods can be used to transform one probability distribution into another.

Importance sampling (Fishman 1995) is a well-known example from statistics that uses a transformed distribution to improve the efficiency of Monte Carlo simulations. In financial economics, risky assets are typically priced under a risk-neutral measure, which is given by a transformation of the original real-world measure (Björk 1998). Many insurance premiums can be expressed as the expected value of some deformed probability distribution (Wang 1996a,b). In risk management, one is usually interested in some feature (e.g., percentile) of the loss distribution of a portfolio at some fixed date in the future, say 10 days from the present. As a result, real-world predictions of market conditions 10 days hence are required. Using these future market conditions, the portfolio is marked-to-market using a risk-neutral measure to price the individual assets, thus generating the future loss distribution (Duffie and Singleton 2002).

We examine three methods of transforming probability distributions: the exponential family, relative entropy, and distortion. We establish a link between those methods, focusing on the relation between relative entropy and distortion. Relative entropy is commonly used to price risky financial assets in incomplete markets, while distortion is widely used to price insurance risks and in risk management.

These three approaches to modifying a distribution may at first seem to be quite different. According to the entropy optimization postulate, however, distributions reweighted via the exponential family or via distortion can also be obtained as the solution to an entropy optimization problem.

Remark 1 (Entropy optimization postulate) *Every probability distribution, theoretical or observed, is an entropy optimization distribution; i.e., it can be obtained by maximizing an appropriate entropy measure or by minimizing a cross-entropy measure with respect to an appropriate a priori distribution, subject to its satisfying appropriate constraints (Kapur and Kesavan 1992, 297).*

In section 2, we discuss the exponential family, relative entropy, and distortion, noting the obvious connection between relative entropy and the exponential family. Section 3 establishes the connections between minimum relative entropy distributions and distorted probability distributions. In particular, we show the relation between the set of moment constraints imposed on the distribution in the entropy optimization problem and the distortion function. This relation provides a way to construct distortion functions via entropy optimization as well as some intuition about the effect a given distortion function has on the original distribution.

Section 4 introduces the Choquet integral and its properties. We also give results about distortion functions and the corresponding Choquet integral that are relevant in risk measurement. In section 5 we discuss the measurement of financial and insurance risks, noting in particular the concept of a coherent risk measure. Section 6 gives examples of distortion functions and the corresponding moment constraints in the entropy optimization framework, and discusses some properties of the implied distorted risk measures. Section 7 summarizes the findings of our research.

2 The Exponential Family, Relative Entropy, and Distortion

2.1 Exponential families of distributions

One simple method of embedding a given distribution, P, in a family of distributions is provided by the usual notion of an exponential family of distributions. For example, suppose that P is a probability distribution of X (P might be a prior distribution or some loss distribution). Suppose that $G_1(X), \ldots, G_N(X)$ are any N statistics or functions of the random variable X. The exponential family of distributions is a parametric family of distributions, P_λ, described by their Radon-Nykodym derivative with respect to P. Namely,

$$\frac{dP_\lambda}{dP} = \exp\left(\sum_{i=1}^{N} \lambda_i G_i(x) - \psi(\lambda)\right), \tag{1}$$

where $\lambda = (\lambda_1, ..., \lambda_N)$ is the vector of parameters, and ψ is defined by

$$\exp\left(\psi(\lambda)\right) = E_P\left[\exp\left(\sum_{i=1}^{N}\lambda_i G_i(x)\right)\right]. \tag{2}$$

Furthermore, we assume that $E_P\left[\exp\left(\sum_{i=1}^{N}\lambda_i G_i(x)\right)\right] < \infty$ for all λ in some neighbourhood of 0 and $E_P[\cdot]$ denotes the expected value taken with respect to the distribution P. The definition of ψ is such that the resulting measures, P_λ, are probability distributions, all absolutely continuous with respect to P. The statistic $(G_1(X), ..., G_N(X))$ is the canonical sufficient statistic and in some sense represents the natural transformation of the data for identification of parameters in this model. Indeed, it provides the minimum variance unbiased estimator of its mean. Virtually all common statistical distributions (the normal, Poisson, binomial, beta, gamma, and negative binomial, for example) are special cases of the exponential family of distributions. See Lehmann (1983, section 1.4) for details.

2.2 Relative entropy

Relative entropy optimization is another method for deforming distributions that is closely related to the exponential family. Relative entropy provides a notion of distance from one probability distribution to another. For example, suppose that $G_1(X), \ldots, G_N(X)$ are functions of a loss random variable, X, the moments of which are used to describe some aspect of the risk. Specifically, let X be a random variable, $c_1, \ldots, c_N$ be constants, and $\mathcal{M}$ be a set of moment constraints of the form

$$E\left[G_i(X)\right] = c_i, \qquad \text{for} \quad i = 1, ..., N, \tag{3}$$

that we want X to satisfy under the new distribution. Using relative entropy as a measure of distance, the minimum relative entropy distribution is the measure that is closest to the original distribution, satisfies the set of moment constraints, $\mathcal{M}$, and is absolutely continuous with respect to the original distribution. The minimum relative entropy distribution is simply a reweighting of the original measure. Let us formally define relative entropy and the minimum relative entropy distribution.

Definition 1 (Relative entropy) *Given two probability measures, P and $\widetilde{P}$, such that $\widetilde{P} \ll P$, the* relative entropy *of $\widetilde{P}$ with respect to P is defined by*

$$H\left(\widetilde{P}|P\right) = E_{\widetilde{P}}\left[\ln\left(\frac{d\widetilde{P}}{dP}\right)\right] = E_P\left[\frac{d\widetilde{P}}{dP}\ln\left(\frac{d\widetilde{P}}{dP}\right)\right], \tag{4}$$

where $\frac{d\widetilde{P}}{dP}$ is a Radon-Nykodym derivative. If the absolute continuity condition, $\widetilde{P} \ll P$, is not satisfied, define the relative entropy to be infinite.

Definition 2 (Minimum relative entropy distribution) *A* minimum relative entropy distribution *is a distribution, P^*, that solves the convex optimization problem:*

$$\min_{\widetilde{P}} H\left(\widetilde{P}|P\right) \text{ subject to the constraints} \tag{5}$$

$$E_{\widetilde{P}}\left[G_i(X)\right] = c_i, \quad \text{for} \quad i = 1, ..., N,$$

$$\int d\widetilde{P} = 1, \quad \text{and}$$

$$\widetilde{P} \ll P.$$

The above optimization problem is easily solved by Lagrangian methods, yielding an explicit expression for the minimum relative entropy distribution. One of the consequences of this minimization problem is that it results in exactly the exponential family of distributions (see equation (1) and Theorem 1). The functions whose moments are constrained in the minimum relative entropy distribution are exactly the canonical sufficient statistics in the exponential family.

Theorem 1 *If there is a measure satisfying the constraints in (5), then the unique solution, P^*, to the constrained minimum relative entropy optimization problem has the form*

$$\frac{dP^*}{dP} = \exp\left(\sum_{i=1}^{N} \lambda_i G_i(x) - \psi(\lambda)\right), \tag{6}$$

where ψ satisfies (2) and $\lambda = (\lambda_1, ..., \lambda_N)$ is determined uniquely by the moment constraints.

This result provides a simple link between the solution to minimum relative entropy optimization problems and the exponential family of distributions. The many attractive properties of the exponential family of distributions seem to indicate that minimizing relative entropy between a distribution and a target subject to certain moment constraints is natural. As we alter the constants, c_i, in the constraints of the optimization problem (5), we remain within the same class of distributions, with only the parameters, λ, varying.

2.3 Distortion

Suppose that, under the probability measure, P, a random variable, X, has cumulative distribution function (cdf) F defined by $F(x) = P[X \leq x]$. There are several other simple ways to transform a distribution through either the cdf of a random variable or the decumulative distribution function (ddf): $S(x) = P[X \geq x] = 1 - F(x-) = 1 - \lim_{h \downarrow 0} F(x-h)$. For example, if $g(u)$ is a left-continuous distortion function (see Definition 3) and $S(x)$ is a ddf, then $S^*(x) = g(S(x))$ is a ddf corresponding to a distorted probability distribution. Similarly, we can apply a distortion function to the cdf to obtain $F^*(x) = g(F(x))$. Distorted probability distributions are special cases of the more general theory of monotone set functions and non-additive measures. See Denneberg (1994) for a careful treatment of the more general theory. We begin by defining a distortion function and a distorted probability distribution.

Definition 3 (Distortion function) *A* distortion function, g, *is any non-decreasing function on* $[0, 1]$ *such that* $g(0) = 0$ *and* $g(1) = 1$.

We have seen that a distorted probability distribution can be defined in terms of the ddf by

$$S^*(x) = g(S(x)). \tag{7}$$

Depending on whether g is continuous, right continuous, or left continuous, equation (7) may not correctly define a ddf at discontinuity points of $S(x)$. However, since there are at most countably many such points, (7) is assumed to hold only at continuity points and uniquely determines the distribution. *For the rest of this paper, we assume identity in equations such as (7) at all* x *except possibly those corresponding to discontinuities in* $S(x)$. *We also assume throughout this paper that all integrals (e.g., expected values and Choquet integrals) are properly defined.*

We use the ddf rather than the cdf because it is compatible with the Choquet integral (Definition 5). We have shown that there are two possible ways to distort a distribution:

one through the cdf and the other through the ddf. These methods correspond with the replacement of the distortion function with a *dual* distortion function.

Definition 4 (Dual distortion function) *Given a distortion function, g, the* dual distortion function*, $\bar{g}$, is defined by*

$$\bar{g}(u) = 1 - g(1 - u).$$

The dual distortion has the useful property that if S^* is obtained from S using distortion function g, then the corresponding cumulative distribution functions are related via the dual distortion function:

$$S^*(x) = g(S(x)) \text{ if and only if } F^*(x) = \bar{g}(F(x)). \tag{8}$$

This is easily seen, since, if $S^*(x) = g(S(x))$, it follows that

$$\begin{aligned} F^*(x) &= P^*(X \leq x) = 1 - S^*(x-) \\ &= 1 - g(S(x-)) = 1 - g(1 - F(x-)) \\ &= \bar{g}(F(x-)), \end{aligned}$$

and this equals $\bar{g}(F(x))$ except at the (at most countably many) discontinuity points of F.

For differentiable distortion functions, equation (7) provides a way to compute the distorted probability density function (pdf). Namely,

$$f^*(x) = -\frac{d}{dx} S^*(x) = -\frac{d}{dx} g(S(x)) = f(x) g'(S(x)), \tag{9}$$

where f is the original pdf of X. Furthermore, the distorted probability distribution derived using the dual distortion function has pdf

$$f^*(x) = -\frac{d}{dx} \bar{g}\left(S^*(x)\right) = \frac{d}{dx} \bar{g}(F(x)) = f(x) \bar{g}'(F(x)). \tag{10}$$

This is exactly the density obtained by using F instead of S in the distortion function, g.

3 Link Between Relative Entropy and Distortion

Equations (6) and (9) reveal an obvious link between distorted probability distributions and minimum relative entropy distributions. Theorem 2 formalizes this relationship. For the purpose of this theorem, we will refer to a random variable, X, as discrete if it takes values on a countable *ordered* set, $x_1 < x_2 < \ldots$.

Theorem 2 *Let X be a discrete or continuous random variable with ddf $S(x)$.*

(i) If S^* is the minimum relative entropy distribution satisfying a set of moment constraints, $\mathcal{M}$, then $S^*(x) = g_{\mathcal{M}}(S(x))$ for some distortion function, $g_{\mathcal{M}}$.

(ii) If a distorted probability distribution, P^*, is defined by $S^*(x) = g(S(x))$ for a differentiable distortion function, g, and if a Radon-Nykodym derivative

$$\frac{dP^*}{dP}$$

has finite ν moment for some $\nu > 1$; i.e., if

$$E_P\left[\left(\frac{dP^*}{dP}\right)^{\nu}\right] < \infty,$$

then S^* is the minimum relative entropy distribution satisfying a set of moment constraints.

Proof. Suppose that S^* is a minimum relative entropy distribution. Then, by Theorem 1 for some values of $\lambda_1, \ldots, \lambda_N$,

$$\frac{dP^*}{dP} = \exp\left(\sum_{j=1}^{N} \lambda_j G_j(x) - \psi(\lambda)\right). \tag{11}$$

Define a pseudo-inverse on $(0,1)$ by

$$S^-(y) = \inf\{z; S(z) \le y\},$$

and note that

$$S^-(S(x)) = x,$$

except possibly on a set of x values having P-probability zero. Assume that $S(x)$ is continuous and define

$$g_{\mathcal{M}}(u) = \int_0^u \exp\left(\sum_{j=1}^{N} \lambda_j G_j\left(S^-(v)\right) - \psi(\lambda)\right) dv. \tag{12}$$

Obviously, the condition $g_{\mathcal{M}}(0) = 0$ is satisfied and equation (11) ensures that $g_{\mathcal{M}}(1) = 1$. Furthermore, $g_{\mathcal{M}}$ is non-decreasing, and hence is a distortion function. Then,

$$\frac{dP^*}{dP} = g'_{\mathcal{M}}(S(x)) = \exp\left(\sum_{j=1}^{N} \lambda_j G_j(x) - \psi(\lambda)\right) \text{ almost surely } (P),$$

and therefore $S^*(x) = g_{\mathcal{M}}(S(x))$. In the case of a discrete random variable with probability functions $f(x)$ and $f^*(x)$ corresponding to the original and the distorted distributions, let $s_i = S(x_i)$ and note that $1 = s_1 > s_2 > \dots$, and $s_N \to 0$ as $N \to \infty$. Define

$$g(1) = 1,\ g(0) = 0,$$

and

$$g(s) = g(s_i) - (s_i - s)\exp\left(\sum_j \lambda_j G_j(x_i) - \psi(\lambda)\right), \text{ for } s_i > s \geq s_{i+1}.$$

Then, it is easy to check the requirement that

$$\frac{f^*(x_i)}{f(x_i)} = \frac{g(s_{i+1}) - g(s_i)}{s_{i+1} - s_i} = \exp\left(\sum_j \lambda_j G_j(x_i) - \psi(\lambda)\right).$$

For part (ii), assume that $S^*(x) = g(S(x))$. Then, if the distribution is continuous, the pdf satisfies

$$\begin{aligned} f^*(x) &= f(x) g'(S(x)) \\ &= f(x)\exp(\phi(x)), \end{aligned} \tag{13}$$

where we define $\phi(x) = \ln(g'(S(x)))$ if $g'(S(x))$ is positive and $\phi(x) = -\infty$ if $g'(S(x)) = 0$. Now, consider distributions supported on the set $B = \{x; g'(S(x)) > 0\}$. If we solve for the minimum relative entropy distribution subject to the constraint

$$E(\phi(X)) = c,$$

the solution takes the form

$$f^*(x) = f(x)\exp(\lambda\phi(x) - \psi(\lambda)),$$

implicitly defining the function ψ. When $\lambda = 1$, equation (13) implies that $\psi(1) = 0$, and therefore the distorted probability distribution is also a minimum relative entropy distribution for a particular value of c. The discrete case is similar, with $g'(s_i)$ replaced by a forward difference:

$$\frac{g(s_{i+1}) - g(s_i)}{s_{i+1} - s_i}.$$

■

A consequence of Theorem 2 is that distortion functions are implicitly defined through the formulation of a minimum relative entropy optimization problem. The implied distortion in the continuous case is given by

$$g(S(x)) = \int_0^{S(x)} \exp\left(\sum_{j=1}^{N} \lambda_j G_j \left(S^{-}(u)\right) - \psi(\lambda) \right) du, \tag{14}$$

with $\psi(\lambda)$ chosen so that $g(1) = 1$. The conditions are quite mild. For example, if X is continuous, the condition

$$\infty > E_P \left[\left(\frac{dP^*}{dP} \right)^{\nu} \right] = E_P \left[(g'(S(X)))^{\nu} \right] = E_P \left[(g'(U))^{\nu} \right],$$

where $U \sim \text{Uniform}[0,1]$ holds for some $\nu > 1$ provided that g has a bounded first derivative.

The constraints are not uniquely determined by the distortion function in part (ii) of Theorem 2. Indeed, typically there is an infinite set of constraints that lead to the same minimum relative entropy distribution, P^*, because there are an infinite number of ways of interpolating between two distributions within the exponential family of distributions. For example, suppose that P is the uniform distribution on the interval $[0,1]$. Consider minimizing the relative entropy $H(P^*|P)$ subject to the moment constraints:

$$E_{P^*}(X) = 0.7420$$
$$E_{P^*}(X^2) = 0.6084.$$

The solution is a pdf of the form

$$f^*(x) = \exp\left(\lambda_0 + \lambda_1 x + \lambda_2 x^2\right), \quad \text{for} \quad 0 < x < 1,$$

with $\lambda_0 = -\psi(\lambda) = -1.5707$, $\lambda_1 = 1$, and $\lambda_2 = 2$. This same pdf, however, solves a minimum relative entropy problem, with the above moment constraints replaced by

$$E_{P^*}(X + 2X^2) = 1.9588.$$

The method used in the above proof is not always the most convenient. For a member of an exponential family of distributions, the natural sufficient statistic provides the usual moment constraint and the whole family of distributions can be generated by simply varying the constants, c_i, in the constraints $E(G_i(X)) = c_i$. Table 2 gives examples of the families of distortion functions that result.

For the applications to be studied later, it is of interest to know whether a distortion function is concave or convex. Theorem 3 gives a necessary and sufficient condition on the moment constraints to determine the concavity or convexity of the corresponding distortion function.

Theorem 3 *Suppose that a distortion function, g, is defined by the moment constraints in a relative entropy optimization problem. Thus, g is as given in equation (14). Define the functions $h_j = G_j o S^-$, assumed to be differentiable, for $j = 1, \ldots, N$. Then,*

(i) g is concave if and only if $\sum_{i=1}^N \lambda_i h_i'(u) \le 0$ for all $u \in [0,1]$.

(ii) g is convex if and only if $\sum_{i=1}^N \lambda_i h_i'(u) \ge 0$ for all $u \in [0,1]$.

Proof. Differentiating g', we get

$$g''(u) = \exp\left(\sum_{i=1}^{N} \lambda_i h_i(u) - \psi(\lambda)\right) \sum_{i=1}^{N} \lambda_i h_i'(u), \tag{15}$$

and, since g is a distortion function, it is non-decreasing; hence, $g' \ge 0$. Since g is a non-negative twice-differentiable function with $g' \ge 0$, then g is concave $\Leftrightarrow$ $g'' \le 0$ $\Leftrightarrow$ $\sum_{i=1}^N \lambda_i h_i'(u) \le 0$ for all $u \in [0,1]$.

The proof of part (ii) is similar. ■

One can construct more complicated distortion functions by mixing and composing existing ones. These operations preserve the property of concavity or convexity as stated in the following lemma (Wang 1996b).

Lemma 4 *Let g_i, $i = 1, \ldots, n$ be concave (convex) distortion functions.*

(i) For $p_i \ge 0$ with $\sum_{i=1}^n p_i = 1$, the function $g = \sum_{i=1}^n p_i g_i$ is a concave (convex) distortion function.

(ii) The function $g = g_2 \circ g_1$ is a concave (convex) distortion function.

3.1 The constraints

Throughout this section, let X be a random variable, with f, F, and S its pdf, cdf, and ddf, respectively. One can reasonably apply various constraints to naturally occurring random variables in deforming a distribution. Note that, under the true distribution, $S(X)$ and $F(X)$ are both uniformly distributed on $[0, 1]$. It is natural to apply constraints to various functions of these quantities, since they operate on easily understood scales. For example, both $S(X)$ and $F(X)$ are uniformly distributed if X has a continuous distribution and both $\Theta(X) = -\ln(S(X))$ and $\bar{\Theta}(X) = -\ln(F(X))$ have a standard exponential distribution. Because they are monotone functions of X, increasing them increases (or decreases) the values of X.

Perhaps a more natural quantity that can be used to express the severity of a change in the loss distribution or the attitudes to risk is a ratio

$$T_+(X) = \frac{S^*(X)}{S(X)},$$

for some transformed loss distribution, $S^*(X)$. For example, suppose that we constrain

$$E_{P^*}\left[T_+(X)\right] = 10.$$

Then, under the transformed distribution, losses of magnitude greater than X occur with 10 times the frequency that they have under the original distribution. Because $T_+(X)$ represents the ratio of tail probabilities, it is a natural vehicle for constraints.

Similarly, the ratio

$$T_-(X) = \frac{F^*(X)}{F(X)}$$

represents the ratio of left tail probabilities and, if we wish to control them, it is natural to constrain the expected value of $T_-(X)$. Therefore, moments of T_- or T_+ such as T_+^{ν} appear to be reasonable quantities to constrain in a risk-management model. We might use the powers such as $\nu = 1$ or logarithms (the equivalent of $\nu = 0$).

For continuous distributions, the cumulative hazard function $\Theta(x) = -\ln(S(x))$, and, if we apply a constraint to its expected value, $E_{P^*}(\Theta(X)) = E_{P^*}(-\ln(S(X)))$, under the transformed measure, then we are changing the mean of an exponential(1) random variable by altering the expected cumulative hazard experienced over a lifetime. This constraint has another simple interpretation as an expected value under the original distribution. Using integration by parts and assuming a continuous distribution,

$$\begin{aligned} E_{P^*}\left[\ln(T_+(X))\right] + 1 &= E_{P^*}\left[\Theta(X)\right] \\ &= -\int \ln(S(x)) f^*(x) dx \\ &= \int \frac{1}{S(x)} f(x) S^*(x) dx \\ &= E_P\left[\frac{S^*(X)}{S(X)}\right] = E_P\left[T_+(X)\right]. \end{aligned} \tag{16}$$

Thus, constraining the logarithm of the (original) survivor function under the distorted measure is equivalent to constraining the first moment of T_+ under the original measure. Similarly,

$$\begin{aligned} E_{P^*}\left[\bar{\Theta}(X)\right] &= E_P\left[\frac{F^*(X)}{F(X)}\right] = E_P\left[T_-(X)\right], \\ E_P\left[\Theta^*(X)\right] &= E_{P^*}\left[T_+^{-1}(X)\right], \text{ and} \\ E_P\left[\bar{\Theta}^*(X)\right] &= E_{P^*}\left[T_-^{-1}(X)\right]. \end{aligned}$$

In general, constraints on the logarithm of the cdf or the ddf under one measure translate to constraints on the expected cdf ratio or ddf ratio under the other measure.

According to Table 1, simple constraints applied to functions of T_- and T_+ introduce specific terms into the distortion function, g (actually, its derivative), in the relation $S^*(x) = g(S(x))$. The constant K ensures that the condition $g(1) = 1$ is satisfied.

Table 1: Moment constraints corresponding to terms in a distortion function

Constraint on	$g'(u)$ multiplied by $K\times$
$E_{P^*}[\Theta(X)]$, or, $E_{P^*}[\ln(T_+(X))]$, or, $E_P[T_+(X)]$	u^{a-1}
$E_{P^*}[\bar{\Theta}(X)]$, or, $E_{P^*}[\ln(T_-(X))]$, or, $E_P[T_-(X)]$	$(1-u)^{b-1}$
$E_{P^*}[S(X)]$, or, $E_{P^*}[F(X)]$	$e^{-u/c}$
$E_{P^*}[T_+^{-1}(X)]$, or, $E_P[\Theta^*(X)]$, or, $E_P[\ln(T_+(X))]$	$\exp(\lambda \frac{u}{g(u)})$
$E_{P^*}[T_-^{-1}(X)]$, or, $E_P[\bar{\Theta}^*(X)]$, or, $E_P[\ln(T_-(X))]$	$\exp(\lambda \frac{1-u}{1-g(u)})$

4 The Choquet Integral, Distortion Functions, and Risk Measurement

In this section, we define the Choquet integral and list some of its well-known properties. We also provide additional results about the Choquet integral that are useful in the construction of insurance premium principles and risk measures with desirable characteristics. Since many insurance premium principles and risk measures have a Choquet integral representation (Wang 1996a,b, Wang, Young, and Panjer 1997, Wirch and Hardy 1999), these properties may also be used to verify whether a given risk measure or premium principle has a certain characteristic. Desirable characteristics of risk measures and premium principles are briefly mentioned in this section and discussed more fully in section 5. Although this section focuses almost entirely on distortion functions, the reader is reminded of the equivalence of relative entropy and distortion. To motivate the discussion, suppose that X is a random variable and g is a distortion function that leads to a distorted distribution defined by $S^*(x) = g(S(x))$ or, equivalently, $F^*(x) = \bar{g}(F(x))$. Notice that the expected value of the random variable under the distorted distribution is given by

$$\begin{aligned} E_{P^*}[X] &= E_{P^*}\left[X^+\right] - E_{P^*}\left[X^-\right] \qquad (17) \\ &= \int_0^\infty S^*(x)dx - \int_0^\infty F^*(-x)dx \\ &= \int_0^\infty g(S(x))dx - \int_0^\infty \bar{g}(F(-x))dx \\ &= \int_0^\infty g(S(x))dx - \int_0^\infty [1 - g(S(-x))]dx \\ &= \int_0^\infty g(S(x))dx + \int_{-\infty}^0 [g(S(x)) - 1]dx, \end{aligned}$$

where $X^+ = \max(X, 0)$ and $X^- = \max(0, -X)$. This leads to the definition of the Choquet integral.

Definition 5 (Choquet integral) *For any random variable X with ddf $S(x)$, the* Choquet integral *with respect to distortion function g is defined by (Denneberg 1994):*

$$
\begin{aligned}
H_g(X) &= \int_0^{\infty} g(S(x))dx + \int_{-\infty}^{0} [g(S(x)) - 1]dx \\
&= \int_0^{\infty} [1 - \bar{g}(F(x))]dx - \int_{-\infty}^{0} \bar{g}(F(x))dx.
\end{aligned}
$$

According to (17), if X is a random variable, then the Choquet integral is equivalent to the expected value of X under the deformed probability distribution with ddf S^*. That is, for any distortion, g,

$$H_g(X) = E_{P^*}[X]\,. \tag{18}$$

Note that $H_g(\cdot)$ is not to be confused with the notation for relative entropy introduced in section 2.2.

Definition 6 (Comonotonic random variables) *Two random variables, X and Y, are* comonotonic *if there exists another random variable, Z, and increasing real-valued functions, u and v, such that*

$$X = u(Z) \qquad Y = v(Z). \tag{19}$$

A class, $\mathcal{C}$, of random variables is called comonotonic *if and only if each pair of random variables in $\mathcal{C}$ is comonotonic (Denneberg 1994).*

If X and Y represent risks and are comonotonic, this means that one is not a hedge for the other. For example, a stock and a call option written on the stock are comonotonic risks. Theorem 5 lists some well-known properties of the Choquet integral (see Denneberg 1994, chapters 5 and 6).

Theorem 5 (Properties of H_g) *For any distortion function, g, and real-valued random variables, X, Y, the following properties hold:*

(i) Monotonicity
$X \leq Y$ *implies* $H_g(X) \leq H_g(Y)$.

(ii) Positive homogeneity
$H_g(cX) = cH_g(X)$ *for* $c \geq 0$.

(iii) *If $X = c$ for any constant c, then $H_g(X) = c$.*

(iv) *Translation invariant*

$H_g(X + c) = H_g(X) + c$ for any constant c.

(v) *Comonotonic additivity*

If X, Y are comonotonic, then $H_g(X + Y) = H_g(X) + H_g(Y)$.

(vi) *Subadditive for concave g*

If g is concave, then $H_g(X + Y) \leq H_g(X) + H_g(Y)$.

(vii) *Superadditive for convex g*

If g is convex, then $H_g(X + Y) \geq H_g(X) + H_g(Y)$.

(viii) *Asymmetry*

$H_g(-X) = -H_{\bar{g}}(X)$, where $\bar{g}$ is the dual distortion function of g.

Note that the monotonicity property shown above ensures that

$$H_g(X) \leq \max(X) \tag{20}$$

for any distortion, g, and any random variable, X. This corresponds to the principle of non-excessive loading. That is, the price of a risk should not exceed the maximum possible loss. Subadditivity is another nice property for a risk measure; it simply states that there should be no incentive to split risks. The following proposition addresses the issue of subadditivity of the Choquet integral.

Proposition 6 *Let g be a continuous distortion function. Then, the following are equivalent:*

(i) g is concave.

(ii) H_g is subadditive. That is, $H_g(X+Y) \leq H_g(X) + H_g(Y)$ for all random variables X and Y.

(iii) $H_g(X + Y) \leq H_g(X) + H_g(Y)$ for all Bernoulli random variables X and Y.

Proof. Notice that 1 implies 2 by Theorem 5 and for all concave distortion functions. Clearly, 2 implies 3. To show that 3 implies 1, suppose that H_g is subadditive for Bernoulli random variables X and Y, and consider the random variables X and Y with the discrete joint distribution given in the following chart with $0 \leq r \leq s \leq 1$:

$\backslash$	Y	
X	0	1
0	$1-s$	$\frac{s-r}{2}$
1	$\frac{s-r}{2}$	r

The Choquet integrals are

$$H_g(X) = g\left(\frac{s+r}{2}\right),$$
$$H_g(Y) = g\left(\frac{s+r}{2}\right), \quad \text{and}$$
$$H_g(X+Y) = g(s) + g(r).$$

Subtracting, we have that

$$\begin{aligned} H_g(X+Y) - (H_g(X) + H_g(Y)) &= g(s) + g(r) - 2g\left(\frac{s+r}{2}\right) \\ &\leq 0, \end{aligned}$$

where the inequality follows from the assumed subadditivity of H_g. This becomes

$$\frac{g(r)+g(s)}{2} \leq g\left(\frac{r+s}{2}\right). \tag{21}$$

For a continuous function, g, this is sufficient to prove concavity. We will prove this by induction.

In particular, we wish to prove that

$$g(\lambda r + (1-\lambda)s) \geq \lambda g(r) + (1-\lambda)g(s). \tag{22}$$

We know that (22) holds for all $r \leq s$ and for $\lambda = 1/2$. Suppose that it holds for all $r \leq s$ and for all λ of the form $\lambda_{j,n} = j2^{-n}$, where $0 \leq j \leq 2^n$. We will show that it holds for all $r \leq s$ and for all λ of the form $\lambda_{k,n+1} = k/2^{n+1}$. We may assume that k is odd, since otherwise the result is a trivial consequence of the assumption that it holds for all $\lambda_{j,n}$;

therefore, we assume that $k = 2j + 1$, with $j < 2^n$. Define $x_{j,n} = \lambda_{j,n} r + (1 - \lambda_{j,n})s$ and $x_{j+1,n} = \lambda_{j+1,n} r + (1 - \lambda_{j+1,n})s$. Then,

$$\begin{aligned} g\left(\lambda_{k,n+1} r \right. & \left. + (1 - \lambda_{k,n+1})s\ \right) \\ &= g\left(\frac{x_{j,n} + x_{j+1,n}}{2}\right) \\ &\geq \frac{g(x_{j,n}) + g(x_{j+1,n})}{2} \\ &\geq \frac{1}{2}\left[\lambda_{j,n} g(r) + (1 - \lambda_{j,n})g(s) + \lambda_{j+1,n} g(r) + (1 - \lambda_{j+1,n})g(s)\right] \\ &= \lambda_{k,n+1} g(r) + (1 - \lambda_{k,n+1})g(s), \end{aligned}$$

where the first inequality follows from (21), the second inequality follows from the induction hypothesis, and the final equality follows from $\frac{1}{2}(\lambda_{j,n} + \lambda_{j+1,n}) = \lambda_{k,n+1}$. For a general value of λ, it is possible to find values $\lambda_{j_n,n} = j_n 2^{-n}$, which approach λ as $n \to \infty$. The general result follows from the continuity of the function $g(u)$. Wirch and Hardy (2000) give an alternate proof valid under stronger conditions. ■

For insurance and risk-management applications of the Choquet integral, it is useful to know conditions on the distortion function that ensure that the integral is bounded below by the expected value of the random variable (non-negative loading). The condition is quite simple, as the following result shows.

Proposition 7 *For a distortion function, g, the following are equivalent:*

(i) $g(u) \geq u$ *for all* $u \in [0, 1]$.

(ii) $E[X] \leq H_g(X)$ *for all random variables,* X.

(iii) $E[X] \leq H_g(X)$ *for all Bernoulli random variables.*

Proof. First note that, from (17),

$$H_g(X) - E(X) = \int_{-\infty}^{\infty} h(S(x))dx,$$

where $h(u) = g(u) - u$. If $h(u) \geq 0$ for all u, then, clearly, $H_g(X) - E(X) \geq 0$ for all random variables. Clearly, 2 implies 3. Assume that $H_g(X) - E(X) \geq 0$ for all Bernoulli random variables, X. We modify a proof from Wirch and Hardy (2000). Define a Bernoulli random variable by

$$X = \begin{cases} 0 & \text{w.p.} \quad 1 - u_0 \\ 1 & \text{w.p.} \quad u_0 \end{cases} \tag{23}$$

and note that $H_g(X) - E(X) = \int_0^1 (g(u_0) - u_0)dx = g(u_0) - u_0 \geq 0$ for all u_0 implies $g(u) \geq u$ for all $u \in [0, 1]$. ■

For the purposes of constructing risk measures via the Choquet integral and distortion (or relative entropy), the following result says that concavity of the distortion function, g, is sufficient to ensure that H_g is subadditive and bounded below by the mean.

Corollary 8 *If g is a concave distortion function, then H_g is subadditive and $E[X] \leq H_g(X)$.*

Proof. Theorem 5 gives the subadditivity property. By noting that concavity implies $g(u) \geq u$ for all $u \in [0, 1]$, Proposition 7 gives $E[X] \leq H_g(X)$. ■

In Theorem 2 we showed the equivalence of distorted probability distributions and minimum relative entropy distributions. The following corollary ties together (i) Choquet integrals that are subadditive and bounded below by the mean, (ii) concave distortion functions, and (iii) conditions on the moment constraints in the corresponding relative entropy optimization problem. This may be used to test whether a given risk measure with a Choquet integral representation is subadditive and bounded below by the expected value.

Corollary 9 *For all random variables and a twice-differentiable distortion function, g, with*

$$g'(u) = \exp\left(\sum_{i=1}^{N} \lambda_i h_i(u) - \psi(\lambda)\right), \tag{24}$$

where h_i are given functions and $\lambda = (\lambda_1, \ldots, \lambda_N)'$ and ψ satisfy $\frac{\partial \psi}{\partial \lambda_i} = C_i$ for given numbers $C_i, \ i = 1, \ldots, N$, the following are equivalent:

(i) H_g *is subadditive and* $E[X] \leq H_g(X)$.

(ii) g *is a concave distortion function.*

(iii) $\sum_{i=1}^{N} \lambda_i h_i'(u) \leq 0$.

Proof. Proposition 6 shows that $1 \Rightarrow 2$. Corollary 8 shows that $1 \Leftarrow 2$. Theorem 3 shows that $2 \Leftrightarrow 3$. ∎

5 Measuring Risks

Risk measures are used extensively in finance and insurance. Prices of risks, such as an insurance premium, are determined by measuring the risks associated with the product. Exchanges and clearing houses determine margin requirements by measuring the riskiness of an investor's portfolio. Risk measures are also used to set capital requirements to help ensure the solvency of the company. These requirements may be imposed by a regulator or by a company's internal risk-management protocol. Companies with many lines of business use risk measures to rate the performance of each business line through its risk-adjusted return on capital (RAROC). The need for good risk measures is quite apparent.

Recently, the Choquet integral has been identified as an important tool in the measuring and pricing of financial and insurance risks. Chateauneuf, Kast, and Lapied (1996) use it to explain apparent discrepancies in observed market prices, such as violation of put-call parity, owing to friction in the market caused by the bid/ask spread. The Choquet integral with the normal distortion function (see section 6.3) has been proposed to price both financial and insurance risks (Wang 2000). Many recent papers in the insurance literature illustrate the use of the Choquet integral as a premium principle. Some examples are Wang (1995, 1996a,b), Wang and Young (1998), Kamps (1998), and Wang and Dhaene (1998).

In fact, by taking an axiomatic approach to insurance prices, Wang, Young, and Panjer (1997) show that any market premium functional that satisfies the prescribed axioms has a Choquet integral representation. Artzner et al. (1999) use a similar axiomatic approach for general risk measures. Premium principles can be thought of more generally as risk measures and therefore the generic term "risk measure" includes them in what follows. We outline the axioms of Artzner et al. below and show that the Choquet integral is useful not only in constructing good risk measures but also in testing the properties of a given risk measure (Wirch 1999). The link between relative entropy and distortion gives further insight into risk measures that have a Choquet integral representation.

5.1 Coherent risk measures

The definitions and axioms presented here are slightly modified versions of those given by Artzner et al. (1999).

Definition 7 (Risk) *A* risk*, X, is a random variable representing the future net loss of an investor at some particular time in the future. $X < 0$ represents a gain and $X \geq 0$ represents a loss.*

Definition 8 (Risk measure) *A* risk measure*, ρ, is a functional that maps X to the real line, $\rho : X \mapsto \Re$.*

For simplicity, we assume that the risk-free rate of interest is zero. Artzner et al. (1999) identify the following axioms as desirable characteristics of a risk measure:

(i) Monotonicity
If $P(X \leq Y) = 1$, then $\rho(X) \leq \rho(Y)$.

(ii) Positive homogeneity
For all $\lambda \geq 0$, $\rho(\lambda X) = \lambda \rho(X)$.

(iii) Translation invariance
For all risks and real numbers, α, $\rho(X + \alpha) = \rho(X) + \alpha$.

(iv) Subadditivity
For all risks X, Y, $\rho(X + Y) \leq \rho(X) + \rho(Y)$.

(v) Relevance
For all non-negative risks with $P(X > 0) > 0$, $\rho(X) > 0$.

The monotonicity axiom says that if risk Y is always greater than risk X, then the risk measure for Y should also be greater than the risk measure for X. The axiom ensures that the property of non-excessive loading is satisfied, ruling out the standard deviation principle, for example. Positive homogeneity reflects that the size, λ, of a position taken on risk X increases the risk measure associated with X by a factor of λ. Translation invariance means that adding (subtracting) the sure amount, α, to risk X increases (decreases) the risk measure by α. That is, if α units of the risk-free asset are removed from a portfolio, then the risk measure should increase by the same amount. Subadditivity can be restated as "a merger does not create extra risk," or "there is no incentive to split risks." The relevance axiom is a necessary, but not sufficient, condition to prevent concentration of risks to remain undetected.

Definition 9 (Coherent risk measure) *A risk measure is* coherent *if it satisfies the four axioms of monotonicity, positive homogeneity, translation invariance, and subadditivity.*

From Proposition 4.1 in Artzner et al. (1999), we see that if ρ is a coherent risk measure, then it is bounded below by the mean net loss; that is,

$$\rho(X) \geq E[X] \tag{25}$$

for all risks, X. We call this the non-negative loading property, consistent with insurance premium terminology.

From Theorem 5 we see that, for any distortion function, the Choquet integral satisfies the axioms of monotonicity, positive homogeneity, and translation invariance. The other results in section 4 address necessary and sufficient conditions on the distortion function for the Choquet integral to be subadditive and satisfy the non-negative loading property. As such, the Choquet integral is a natural tool for constructing coherent risk measures by specifying an appropriate distortion function. Furthermore, if a given risk measure has a Choquet integral representation, the results from section 4 allow one to verify whether the risk measure is coherent.

Definition 10 (Distorted risk measure) *A risk measure is a* distorted risk measure *if it has a Choquet integral representation. The notation ρ_g is used to denote the distorted risk measure with distortion function g.*

Commonly used risk measures such as value-at-risk (VaR) and tail-VaR have a Choquet integral representation; hence, they are examples of distorted risk measures (Wirch and Hardy 1999). Corollary 8 provides sufficient conditions on the distortion function to ensure that a distorted risk measure is coherent. This is useful for constructing coherent risk measures. We restate the result in terms of the coherence of the distorted risk measure.

Corollary 10 *If g is a concave distortion function, then the distorted risk measure, $\rho_g(X)$, is coherent.*

Typically, risk measures are used to compute capital requirements or reserves to protect a company against ruin. A risk measure that considers only the loss part of the distribution is conservative, as potential losses are not offset by potential gains. As stated in Artzner et al. (1999, remark 4.5), actuaries have been considering only the loss part of the distribution since the 1800s. The condition $g(u) \geq \min\left(1, \frac{u}{S(0)}\right)$ for all $u \in [0,1]$ is sufficient to ensure that the relevance axiom is satisfied, as only the loss part of the distribution is used to calculate the distorted risk measure. It is quite strong, however, and can probably be relaxed somewhat for particular distributions. Finally, note that distorted risk measures always satisfy the relevance axiom for non-negative random variables.

As previously stated, the properties of the Choquet integral can be used to test the coherence of a given distorted risk measure. The following corollary provides a test for certain distortion functions and also reminds us of the link to relative entropy (compare with Corollary 9).

Corollary 11 *For all random variables and a twice-differentiable distortion function, g, with*

$$g'(u) = \exp\left(\sum_{i=1}^{N} \lambda_i h_i(u) - \psi(\lambda)\right), \tag{26}$$

where h_i are given functions and $\lambda = (\lambda_1, \ldots, \lambda_N)'$ and ψ satisfy $\frac{\partial \psi}{\partial \lambda_i} = C_i$ for given numbers C_i, $i = 1, \ldots, N$, the following are equivalent:

(i) the distorted risk measure $\rho_g(X)$ is coherent.

(ii) g is a concave distortion function.

(iii) $\sum_{i=1}^{N} \lambda_i h_i'(u) \leq 0$.

5.1.1 *Risk preferences through distortion*

For random variables X representing potential loss, one could define a distorted risk measure depending only on the positive part of the loss by $\rho_g(X) = H_g(X^+)$. One can think of the distortion function, or equivalently the moment constraints, as reflecting attitudes towards risk. Distorted risk measures can therefore be tailored to specific risk attitudes by specifying the distortion function and its parameters.

For non-negative random variables and differentiable distortion functions, distorted risk measures can be regarded as expected utility, for some implied utility function, u (Wirch and Hardy 1999). In particular, if we define the utility function $u(y) = -yg'(S(-y))$, then

$$E(u(-X^+)) = -\int_0^\infty (-x)g'(S(x))f(x)dx \tag{27}$$

$$= \int_0^\infty xg'(S(x))f(x)dx = -H_g(X^+). \tag{28}$$

In other words, the utility is the negative of the risk. This underlines the fact that distortion functions or, equivalently, moment constraints in a relative entropy optimization problem reflect risk preferences. Apparently, the implied utility function, u, depends on the distribution, S, as well as the distortion function, g. In fact, the density $g'(S(x))$ describes how much the "risk-neutral" utility, $u(x) = x$, is modulated by the distortion. As one might expect, the degree to which a given loss is anticipated (and reflected in risk-averse prices) in the market depends not only on the size of the loss but also on the historical frequency, $S(x)$, with which losses of this magnitude are observed.

Distorted risk measures also play a large role in two other economic theories: one by Yaari (1987), and the other by Schmeidler (1989).

6 Examples

In this section, we provide some examples of specific distortions obtained from the simple rules on the moment constraints in the relative entropy framework outlined in section 3.1. The results are summarized in Table 2. Again, note that the constant K ensures that the condition $g(1) = 1$ is satisfied. We discuss special cases of these examples and their corresponding distorted risk measures. For the purposes of exposition, throughout this section assume that X is a continuous random variable with f, F, and S being its pdf, cdf, and ddf, respectively.

Table 2: Moment constraints and resulting distortion functions

Constraints	$\frac{f^*(x)}{f(x)} = g'(S(x))$ where $g'(u) =$	Distortion function	Fixed parameters
$E_{P^*}\left(\bar{\Theta}(X)\right) = c_1$ $E_{P^*}\left(\Theta(X)\right) = c_2$ $E_{P^*}\left(S(X)\right) = c_3$	$Ku^{a-1}(1-u)^{b-1}\exp(-u/c)$, $u \in [0,1]$	Gamma-beta	
$P_{P^*}\left(F^{-1}(1-\alpha_{i-1}) < X \leq F^{-1}(1-\alpha_i)\right) = c_i$, for $i = 1, \ldots, p$, and $\sum_{i=1}^{p} c_i = 1$	k_i, for $\alpha_{i-1} < u \leq \alpha_i$, for $i = 1, 2, \ldots, p$	Piecewise linear	$\alpha_1, \ldots, \alpha_p$
$E_{P^*}\left(\Phi^{-1}(S(X))\right) = c$	$\frac{\phi\left(\Phi^{-1}(u)-\mu\right)}{\phi\left(\Phi^{-1}(u)\right)}$, $u \in [0,1]$	Normal	
$E_{P^*}\left(h_i(X)\right) = c_i$, $i = 1, 2, ..., p$	$K\exp\left\{\sum_{i=1}^{p} \lambda_i h_i\left(S^-(u)\right)\right\}$ $u \in [0,1]$	Generalized Esscher	
$E_{P^*}(\Theta(X)) = c_1$ $E_{P^*}[\ln(d + kS(X))] = c_2$	$Ku^{a-1}(d+ku)^{b-1}$, $u \in [0,1]$	F	d, k

6.1 Gamma-beta distortion

The gamma-beta distortion is a very general distortion, because it can be generated via relative entropy by constraining the expected log ratios of left and right tail probabilities ($\ln(T_-(X))$ and $\ln(T_+(X))$, respectively), along with the expected value of the ddf (equivalently, the cdf). It is defined as

$$g(u) = \int_0^u K t^{a-1}(1-t)^{b-1}\exp(-t/c)dt,$$

where

$$K^{-1} = \int_0^1 t^{a-1}(1-t)^{b-1}\exp(-t/c)dt,$$

and the parameters a, b, and c are all positive. Note that K^{-1} can be obtained from the moment generating function of a beta random variable. The conditions $a \leq 1,\ b \geq 1$, and $c > 0$ are sufficient to ensure that the corresponding distorted risk measure is coherent (Corollary 11). The generality of this distortion function is evident by the fact that it contains a number of other distortion functions as special cases.

6.1.1 *Beta distortion*

With $c = \infty$, we obtain the beta distortion function as a special case of the gamma-beta. This distortion is the incomplete beta function (or the cdf of a $\beta(a, b)$ random variable), which is defined as

$$g(u) = \int_0^u \frac{1}{\beta(a,b)} t^{a-1}(1-t)^{b-1}dt, \tag{29}$$

where

$$K^{-1} = \beta(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)} = \int_0^1 t^{a-1}(1-t)^{b-1}dt, \tag{30}$$

and $\Gamma(a) = \int_0^\infty t^{a-1}e^{-t}dt$ is the gamma function.

The link between minimum relative entropy distributions and distorted probability distributions can be used to determine the parameters (a, b) in terms of (c_1, c_2). Specifically, it can be shown that

$$c_1 = \frac{\Gamma'(a+b)}{\Gamma(a+b)} - \frac{\Gamma'(b)}{\Gamma(b)} \qquad \text{and} \qquad c_2 = \frac{\Gamma'(a+b)}{\Gamma(a+b)} - \frac{\Gamma'(a)}{\Gamma(a)}.$$

Corollary 11 implies that the beta-distorted risk measure, ρ_g, is coherent if and only if $a \leq 1$ and $b \geq 1$ (these are necessary and sufficient conditions for the concavity of g). The parameters a and b control the effect of upper and lower tails, respectively. The application of the beta transform to risk measurement was first proposed by Wirch (1999). The beta distortion is a two-parameter distortion that includes the proportional hazards and dual power distortions as special cases.

6.1.2 *Proportional hazards distortion*

The proportional hazards transform is obtained by constraining the expected integrated hazard rate $\Theta(X)$, or $\ln(T_+(X))$, under the deformed distribution. This distortion is a special case of the gamma-beta distortion simply obtained by setting $b = 1$ and $c = \infty$. Namely,

$$g(u) = u^a,$$

where $K = a$. As above, Theorems 1 and 2 together give us the relation

$$c_2 = -a.$$

Theorem 3 implies that g is concave if and only if $a \leq 1$, and convex if and only if $a \geq 1$. Hence, the distorted risk measure is coherent if and only if $a \leq 1$, which inflates the right tail of the distribution of X. This distorted risk measure has been extensively studied in insurance applications (Wang 1995, 1996a,b, and Wang, Young, and Panjer 1997). It can easily be shown that the distorted probability distribution has a hazard rate of $a\theta(t)$, hence the name proportional hazards (PH) distortion.

6.1.3 *Dual power distortion*

This transform is the dual to the PH distortion (hence the name) and is obtained by constraining $\bar{\Theta}(X)$, or $\ln(T_-(X))$, under the distorted distribution. It is another special case of the gamma-beta distortion with $a = 1$ and $c = \infty$. Also, using the link between relative entropy and distortion, it is easy to show that

$$c_1 = -b. \tag{31}$$

Furthermore, ρ_g is coherent if and only if $b \geq 1$, where the parameter b has the effect of deflating the lower tail of the distribution (Wang 1996a,b).

6.1.4 *Gamma distortion*

This is the gamma cumulative density function conditioned to lie in the interval $[0, 1]$; that is,

$$g(u) = \int_0^u K t^{a-1} \exp(-t/c) dt,$$

giving

$$K^{-1} = \int_0^1 t^{a-1} \exp(-t/c) dt = c^a \gamma(1/c; a),$$

where γ is the incomplete gamma function. This is yet another special case of the gamma-beta distortion with $b = 1$. Thus, $a \leq 1$ implies that g is concave, and therefore that ρ_g is a coherent risk measure. This distortion includes the exponential distortion (with $a = 1$) and the previously discussed PH distortion ($g'(u) = au^{a-1}$ for $c = \infty$) as special cases.

6.1.5 *Exponential distortion*

The exponential distortion is simply the cdf of an exponential random variable constrained to the unit interval. The distortion is generated by restricting the expected right-hand tail, $S(X)$, under the reweighted distribution. Here we have

$$g(u) = \frac{1 - \mathrm{e}^{-u/c}}{1 - \mathrm{e}^{-1/c}},$$

with Theorems 1 and 2 implying

$$c_3 = c + \frac{\mathrm{e}^{-1/c}}{c(1 - \mathrm{e}^{-1/c})}.$$

In this case, the exponential distortion function is always concave and is twice-differentiable; therefore, ρ_g is coherent by Corollary 11.

6.2 Piecewise linear distortion

The single most important example of piecewise linear distortions is one in which we alter the confidence level associated with a VaR. For example, suppose that, under the original distribution, $\text{VaR}_\alpha = \$100$, where $1-\alpha$ is the confidence level associated with the VaR. In other words, if X represents the loss in a specific time horizon (e.g., one day), then $P(X \geq 100) = \alpha$. Suppose that, under the transformed distribution, this is $\text{VaR}_{k_1 \times \alpha}$, so that we apply a constraint $P^*(X > 100) = k_1\alpha$ and, of course, $P^*(X \leq 100) = 1 - k_1\alpha$. According to Table 2, the corresponding distortion function must satisfy $g'(u) = k_1$ or k_2 as $u \leq \alpha$ or $u > \alpha$. Hence, the transformed distribution is obtained from a piecewise linear distortion function, namely,

$$S^*(x) = \begin{cases} k_1 S(x) & \text{if } S(x) \leq \alpha, \\ 1 + k_2(S(x) - 1) & \text{if } S(x) > \alpha, \end{cases}$$

where $k_2 = \frac{1-k_1\alpha}{1-\alpha}$.

Consider one important special case, in which $k_1 = 1/\alpha$ and so the transformed measure has all of its mass in the right tail, $X \geq 100$. In this case, the distortion function is

$$g(u) = \min\left(\frac{u}{\alpha}, 1\right) \quad \text{for} \quad 0 < \alpha < 1, u \in [0,1]. \tag{32}$$

The corresponding risk measure is $H_g(X) = E(X|X > 100)$ and is known as the conditional tail expectation (CTE) or tail-VaR, defined as $E[X|X > F^{-1}(1-\alpha)]$. Interestingly, one obtains the risk-measure CTE by minimizing the relative entropy between the distorted and original distributions subject to the constraint that the entire distribution is supported in the tail $\{x; x > \text{VaR}_\alpha\}$. It is easy to show that CTE is coherent, as its defining distortion function is concave. Furthermore, it is well-known that VaR is not coherent, as it does not satisfy the subadditivity property (Artzner et al. 1999).

6.3 Normal distortion

Through the inverse probability transform, the random variable, X, is mapped to a normal random variable whose mean is then shifted by, μ. This distortion is a special case of the exponential family of distortions discussed in Reesor (2001, section 4.5). Specifically, we have

$$g(u) = \Phi\left[\Phi^{-1}(u) - \mu\right],$$

where Φ is the standard normal cdf. The link between relative entropy and distortion gives $c = \mu$. From the symmetric property of the normal distribution, the dual distortion is

$$\bar{g}(u) = 1 - g(1-u) = \Phi\left[\Phi^{-1}(u) + \mu\right],$$

which is just the inverse of g. The normal distortion is concave if and only if $\mu \leq 0$; therefore, ρ_g is coherent if and only if this condition holds (Corollary 11).

The use of the normal distortion to price financial and insurance risks through the Choquet integral was proposed by Wang (2000). He shows that this transform has some desirable properties. In particular, this distortion is able to reproduce and generalize the capital asset pricing model (CAPM), to reproduce the Black-Scholes formula, and provides a symmetric treatment of assets and liabilities (owing to the symmetric nature of the normal density function). This distortion has also been used in the structural approach to credit-risk modelling as a mapping between actual and risk-neutral default probabilities (Duffie and Singleton 2002).

6.4 Generalized Esscher distortion

In a relative entropy optimization problem, it is sometimes more natural to think about imposing moment constraints on the random variable, X, than on functions of the cdf and ddf. We can write any function h_i as

$$h_i(x) = h_i\left(S^{-}(S(x))\right).$$

This explicitly shows that the moment constraints in this example can be written as

$$E_{\tilde{f}}\left[\bar{h}_i(S(X))\right] = c_i,$$

where $\bar{h}_i = h_i \circ S^{-}$. The derivative of the distortion function is

$$g'(u) = \exp\left\{\sum_{i=1}^{p} \lambda_i h_i\left(S^{-}(u)\right) - \psi(\lambda)\right\},$$

which, upon substituting $u = S(x)$, becomes

$$g'(S(x)) = \exp\left\{\sum_{i=1}^{p} \lambda_i h_i(x) - \psi(\lambda)\right\}.$$

When $p = 1$, this is the generalized Esscher transform given in Kamps (1998). That is, the distribution obtained through a generalized Esscher transform is exactly the minimum relative entropy distribution obtained by constraining the expected value of $h_1(X)$. If $h_1(u) = u$, then the deformed distribution is the one obtained by the usual Esscher transform.

From an asset pricing perspective, arbitrage considerations typically prescribe the mean of the underlying asset, which gives $h_1(u) = u$ and $p = 1$. In addition to the constraint on the mean, suppose we wish to impose a calibrating volatility on the underlying asset. This is accomplished with $h_1(u) = u$, $h_2(u) = u^2$, and $p = 2$. These are exactly the constraints used in the option pricing models developed in Reesor (2001, chapter 1).

In this example, we will not discuss coherence of the related distorted risk measure, as the distortion function depends very much on the undistorted distribution of X. Using only one moment constraint ($p = 1$), the distorted risk measure, ρ_g, takes the form of some common premium principles. Namely, ρ_g is

(i) the net premium principle when $h_1(u) = 1$,

(ii) a modified variance principle when $h_1(u) = \ln u$, and

(iii) the Esscher premium principle when $h_1(u) = u$ (Kamps 1998).

Kamps points out that when h_1 corresponds to a ddf, the resulting ρ_g has a certain renewal-theoretic interpretation.

7 Conclusion

This paper has established a relationship between relative entropy and distortion, two commonly used methods for reweighting probability distributions. We have shown that moment constraints of functions of the cumulative and decumulative distribution functions introduce specific terms into the implied distortion function. We have provided some results of the relationship between the features of a distortion function and the properties of the corresponding Choquet integral.

The Choquet integral is a natural tool for risk measurement. Risk preferences are contained in the distortion function or in the equivalent moment constraints in the relative entropy framework. The link to relative entropy provides additional intuition behind the many premium principles and risk measures that have a Choquet integral representation. Our results permit verification of whether a risk measure with a Choquet integral representation is coherent. In addition, relative entropy provides an easy way of constructing new coherent risk measures by prescribing new sets of moment constraints. These ideas are evident in the examples that were discussed in section 6.

The link between relative entropy and distortion opens many possible avenues for future research. For example, it may prove interesting to apply concepts from extreme value theory in the construction of risk measures. A set of constraints may be imposed such that the deformed distribution has the same mean excess function as the generalized Pareto distribution (GPD). The GPD is an integral tool in extreme value theory, because it appears as the limit distribution of scaled excesses over high thresholds (Embrechts, Klüppelberg, and Mikosch 1997).

The Choquet integral and distortion are central to a dual theory of utility by Yaari (1987) and another by Schmeidler (1989). The established link provides further intuition to this dual theory. Furthermore, in the usual expected utility theory, a utility function contains attitudes towards risk, and maximizing expected utility in this framework is also equivalent to risk minimization. Using the relation between relative entropy and distortion developed in this paper, it will be interesting to establish further unifying results connecting risk minimization, utility theory, and the dual utility theory.

The connection between relative entropy and distortion gives a new perspective to the growing body of literature that applies the Choquet integral to problems in finance, risk management, and insurance.

References

Artzner, P., F. Delbaen, J.-M. Eber, and D. Heath (1999): "Coherent Measures of Risk," *Mathematical Finance*, 9(3), 203–228.

Björk, T. (1998): *Arbitrage Theory in Continuous Time*. Oxford University Press Inc., New York.

Chateauneuf, A., R. Kast, and A. Lapied (1996): "Choquet Pricing for Financial Markets with Frictions," *Mathematical Finance*, 6(3), 323–330.

Denneberg, D. (1994): *Non-Additive Measure and Integral*. Kluwer Academic Publishers, Dordrecht, The Netherlands.

Duffie, D., and K. J. Singleton (2002): *Credit Risk for Financial Instutions: Management and Pricing*. Princeton University Press, forthcoming.

Embrechts, P., C. Klüppelberg, and T. Mikosch (1997): *Modelling Extremal Events for Insurance and Finance*. Springer-Verlag, Milan.

Fishman, G. S. (1995): *Monte Carlo: Concepts, Algorithms, and Applications*. Springer-Verlag, 175 Fifth Avenue, New York, NY, Springer Series in Operations Research.

Kamps, U. (1998): "On a Class of Premium Principles Including the Esscher Principle," *Scandinavian Actuarial Journal*, 1, 75–80.

Kapur, J., and H. Kesavan (1992): *Entropy Optimization Principles with Applications*. Academic Press, Inc., San Diego, CA.

Lehmann, E. (1983): *Theory of Point Estimation*. John Wiley & Sons, New York.

Reesor, R. M. (2001): "Relative Entropy, Distortion, the Bootstrap and Risk," Ph.D. thesis, Department of Statistics and Actuarial Science, University of Waterloo, Waterloo, Ontario, Canada.

Schmeidler, C. (1989): "Subjective Probability and Expected Utility Without Additivity," *Econometrica*, 57, 571–587.

Wang, S. (1995): "Insurance Pricing and Increased Limits Ratemaking by Proportional Hazards Transforms," *Insurance: Mathematics and Economics*, 17, 43–54.

——— (1996a): "Ordering of Risks Under PH-Transforms," *Insurance: Mathematics and Economics*, 18, 109–114.

——— (1996b): "Premium Calculation by Transforming the Layer Premium Density," *ASTIN Bulletin*, 26(1), 71–92.

——— (2000): "A Class of Distortion Operators for Pricing Financial and Insurance Risks," *Journal of Risk and Insurance*, 67(1), 15–36.

Wang, S., and J. Dhaene (1998): "Comonotonicity, Correlation Order and Premium Principles," *Insurance: Mathematics and Economics*, 22, 235–242.

Wang, S. S., and V. R. Young (1998): "Risk-Adjusted Credibility Premiums Using Distorted Probabilities," *Scandinavian Actuarial Journal*, 2, 143–165.

Wang, S. S., V. R. Young, and H. H. Panjer (1997): "Axiomatic Characterization of Insurance Prices," *Insurance: Mathematics and Economics*, 21, 173–183.

Wirch, J. L. (1999): "Coherent Beta Risk Measures for Capital Requirements," Ph.D. thesis, University of Waterloo, Waterloo, Ontario.

Wirch, J. L., and M. R. Hardy (1999): "A Synthesis of Risk Measures for Capital Adequacy," *Insurance: Mathematics and Economics*, 25, 337–347.

Wirch, J. L., and M. R. Hardy (2000): "Distortion Risk Measures: Coherence and Stochastic Dominance," Working paper presented to 2000 AFIR Colloquium, Tromso, Norway.

Yaari, M. E. (1987): "The Dual Theory of Choice Under Risk," *Econometrica*, 55(1), 95–115.

Bank of Canada Working Papers
Documents de travail de la Banque du Canada

Working papers are generally published in the language of the author, with an abstract in both official languages. ***Les documents de travail sont publiés généralement dans la langue utilisée par les auteurs; ils sont cependant précédés d'un résumé bilingue*.**

2002

2002-10	La fiabilité des estimations de l'écart de production au Canada	J.-P. Cayen et S. van Norden
2002-9	The Microstructure of Multiple-Dealer Equity and Government Securities Markets: How They Differ	T. Gravelle
2002-8	Restructuring in the Canadian Economy: A Survey of Firms	C. Kwan
2002-7	Contribution of ICT Use to Output and Labour-Productivity Growth in Canada	H. Khan and M. Santos
2002-6	Currency Fluctuations, Liability Dollarization, and the Choice of Exchange Rate Regimes in Emerging Markets	P.N. Osakwe
2002-5	The Effects of Bank Consolidation on Risk Capital Allocation and Market Liquidity	C. D'Souza and A. Lai
2002-4	Does Micro Evidence Support the Wage Phillips Curve in Canada?	J. Farès
2002-3	An Introduction to Wavelets for Economists	C. Schleicher
2002-2	Asset Allocation Using Extreme Value Theory	Y. Bensalah
2002-1	Taylor Rules in the Quarterly Projection Model	J. Armour, B. Fung, and D. Maclean

2001

2001-27	The Monetary Transmission Mechanism at the Sectoral Level	J. Farès and G. Srour
2001-26	An Estimated Canadian DSGE Model with Nominal and Real Rigidities	A. Dib
2001-25	New Phillips Curve with Alternative Marginal Cost Measures for Canada, the United States, and the Euro Area	E. Gagnon and H. Khan
2001-24	Price-Level versus Inflation Targeting in a Small Open Economy	G. Srour
2001-23	Modelling Mortgage Rate Changes with a Smooth Transition Error-Correction Model	Y. Liu
2001-22	On Inflation and the Persistence of Shocks to Output	M. Kichian and R. Luger

Copies and a complete list of working papers are available from:
Pour obtenir des exemplaires et une liste complète des documents de travail, prière de s'adresser à :

Publications Distribution, Bank of Canada
234 Wellington Street, Ottawa, Ontario K1A 0G9
E-mail: publications@bankofcanada.ca
Web site: http://www.bankofcanada.ca

Diffusion des publications, Banque du Canada
234, rue Wellington, Ottawa (Ontario) K1A 0G9
Adresse électronique : publications@banqueducanada.ca
Site Web : http://www.banqueducanada.ca