B.A.J. **9**, IV, 959-991 (2003)

RISK MEASURES AND THEORIES OF CHOICE

By A. Tsanakas and E. Desli

ABSTRACT

We discuss classes of risk measures in terms both of their axiomatic definitions and of the economic theories of choice that they can be derived from. More specifically, expected utility theory gives rise to the exponential premium principle, proposed by Gerber (1974), Dhaene *et al.* (2003), whereas Yaari's (1987) dual theory of choice under risk can be viewed as the source of the distortion premium principle (Denneberg, 1990; Wang, 1996). We argue that the properties of the exponential and distortion premium principles are complementary, without either of the two performing completely satisfactorily as a risk measure. Using generalised expected utility theory (Quiggin, 1993), we derive a new risk measure, which we call the distortion-exponential principle. This risk measure satisfies the axioms of convex measures of risk, proposed by Föllmer & Shied (2002a,b), and its properties lie between those of the exponential and distortion principles, which can be obtained as special cases.

KEYWORDS

Risk Measures; Premium Calculation Principles; Coherent Measures of Risk; Distortion Premium Principle; Exponential Premium Principle; Expected Utility; Dual Theory of Choice Under Risk; Generalised Expected Utility

CONTACT ADDRESS

Andreas Tsanakas, Market Risk and Reserving Unit, Lloyd's of London, One Lime Street, London EC3M 7HA, U.K. Tel: +44(0)20-7327-5268; Fax: +44(0)20-7327-5658; E-mail: andreas.tsanakas@lloyds.com.

1. Introduction

Recent years have witnessed the emergence in the financial literature of a sophisticated theory of risk measures, as a means for determining capital requirements for the holders of risky portfolios. A landmark in this development has been the axiomatic definition of coherent risk measures by Arztner *et al.* (1999), which has by now achieved the status of a classic. However, the functional forms and fundamental properties of risk measures have been extensively studied in the actuarial literature for more than 30 years, in the guise of premium calculation principles (e.g. Bühlmann, 1970; Goovaerts *et al.*, 1984). A specific class of coherent risk measures, termed distortion premium principles, were introduced in an insurance pricing context by Denneberg (1990) and Wang (1996).

There are two alternative ways of defining (classes of) risk measures. One is to start with a set of desirable properties that risk measures should satisfy, and then determine mathematically their corresponding functional representations. Artzner *et al*. (1999) propagated such an axiomatic approach. The second way, often encountered in the actuarial literature, is to determine the functional form of risk measures via economic indifference arguments. The zero-utility premium principle was defined by Bühlmann (1970) by considering an insurer whose preferences are characterised by expected utility (von Neumann & Morgenstern, 1947). Furthermore it can be shown that the distortion risk measures proposed by Denneberg (1990) and Wang (1996) can also be obtained via indifference arguments, based on Yaari's (1987) dual theory of choice, which was proposed as a complementary approach to expected utility.

Our aim in this paper is to associate these two approaches of defining risk measures by discussing the sets of properties that risk measures, derived via alternative economic theories of choice, are characterised by, and vice versa. We believe that the simultaneous study of sets of properties and corresponding economic theories is illuminating in both directions. On the one hand, it sheds light on the implicit economic assumptions behind the choice of a specific risk measure, while, on the other, it studies the explicit consequences of decision making with a specific theory of choice. As an example of such an interaction between axiomatic definition and economic theory, we note the recent discussion of the appropriateness of coherent measures of risk for setting capital requirements. Föllmer & Schied (2002a) observed that the positive homogeneity and subadditivity properties of coherent risk measures make them insensitive to liquidity risk. Meanwhile Dhaene *et al*. (2003) phrased a similar criticism, pointing out the incompatibility of coherent risk measures with expected utility theory. They proposed instead the use of the exponential premium principle, which is derived from an exponential utility function, and has an additional justification originating in actuarial ruin theory.

Our view is that both the (coherent) distortion risk measures, which are compatible with the dual theory of choice, and risk measures derived from expected utility, such as the exponential premium principle, are not completely satisfactory. Our criticism lies both in terms of their corresponding sets of properties and of the accuracy with which they describe economic agents' preferences. We propose a class of risk measures based on generalised expected utility theory (Quiggin, 1993), which is a combination of expected utility and the dual theory of choice, and provides a better characterisation of individual preferences. Using an exponential utility function and an arbitrary probability distortion function, we define a risk measure which we call the distortion-exponential premium principle. The properties of this risk measure lie in between those of the distortion and the exponential premium principles. We provide numerical illustrations of the proposed risk measure's properties, which show that, for relatively small portfolios of risks, it behaves approximately like a coherent risk measure, whereas for larger portfolios, for which liquidity and risk aggregation becomes an issue, it inherits properties of the exponential premium principle.

The structure of the paper is as follows. In Section 2 we give a brief exposition of alternative characterisations of economic agents' preferences, namely expected utility theory, the dual theory of choice under risk, and generalised expected utility. In Section 3 we define risk measures and discuss alternative sets of properties that they should obey. Subsequently, in Section 4 we proceed with the definition of risk measures via indifference arguments for the three theories of choice that were presented earlier, and discuss their respective properties. In Section 5 we define the distortion-exponential principle and compare its basic properties to those of distortion risk measures and the exponential premium principle. Finally, we present our conclusions in Section 6.

2. Choice under Risk

2.1 *Preference Operators and Functionals*

Suppose that an economic agent (e.g. an insurance company) is exposed to a risk $X \in \mathcal{X}$, where X is a random variable representing (the payoff from) a portfolio and $\mathcal{X}$ is the set of available portfolios that the agent may hold. The agent will prefer some elements of $\mathcal{X}$ to others. We characterise the agent's preferences via the binary operator '$\succeq$', where $X \succeq Y$ for $X, Y \in \mathcal{X}$ is understood as meaning 'portfolio X is at least as preferable as portfolio Y'. Given that the operator '$\succeq$' has some reasonable properties (see Appendix A for details), we can use it to order all portfolios with random payoffs in $\mathcal{X}$ with respect to how desirable they are to the agent.

In order to be able to carry out such an ordering of risky portfolios in practice, it is useful to associate the preference relation '$\succeq$' with a real-valued function V, defined on $\mathcal{X}$ via:

$$X \succeq Y \Leftrightarrow V(X) \geq V(Y) \text{ for } X, Y \in \mathcal{X}. \tag{1}$$

V is called the preference functional associated with '$\succeq$'. The exact form of V will depend on the assumptions that we make for the general properties of the preference operator '$\succeq$'; different assumptions lead to different representations in terms of V. In this section we will discuss three cases: expected utility theory, based on the work of von Neumann & Morgenstern (1947), the dual theory of choice developed by Yaari (1987), and the generalised or rank-dependent utility framework, which is a synthesis of the two previous approaches (Schmeidler, 1989; Quiggin, 1993). Conditions under which each theory emerges are discussed in more detail in Appendix A, while a table summarising alternative characterisations of preferences can be found in Appendix B.

2.2 *Risk Aversion and Second Stochastic Dominance*

Economic theories of choice provide formal ways in which to characterise agents' behaviour, which often is risk averse. Risk aversion is a standard assumption in the literature, while being also widely observed in the practice of economic decision making. One prominent example is the buying of insurance by an individual at a higher cost than the expected loss that the individual faces. Formally, an agent is risk averse if he always prefers a deterministic portfolio R to a random one X, with expected value $E[X] = R$ (if the reverse is true, the agent is called risk seeking):

$$E[X] \succeq X \Leftrightarrow V(E[X]) \geq V(X) \ \ \forall X \in \mathcal{X}. \tag{2}$$

A more sophisticated characterisation of risk aversion is via the concept of second order stochastic dominance. Stochastic dominance provides a way of ordering risky portfolios in terms of their risk characteristics. Let $X, Y \in \mathcal{X}$ be two portfolios that we wish to compare and S_X, S_Y their respective survival functions. We say that X second stochastically dominates Y and write $X \succeq_{2\text{nd}} Y$, if:

$$\int_x^\infty S_Y(t)dt \leq \int_x^\infty S_X(t)dt \text{ for all } x \in \mathbb{R} \tag{3}$$

with the inequality being strict for some $x \in \mathbb{R}$. The above equations have the interpretation that, if portfolio X second stochastically dominates Y, the expected profit from X in excess of any level x is higher than the expected profit from Y.

It is desirable that a preference relation is consistent with second stochastic dominance. We say that a preference operator '$\succeq$' preserves second stochastic dominance whenever:

$$X \succeq_{2\text{nd}} Y \Rightarrow V(X) \geq V(Y). \tag{4}$$

2.3 *Expected Utility Theory*

The first approach to modelling preferences on a collection of risky portfolios that we discuss here is expected utility theory. This theory was introduced in the classic book by von Neumann & Morgenstern (1947), and developed further by several authors. The intellectual roots of utility theory are however earlier, and date back to the 18th century and the work of the mathematician Daniel Bernoulli. Expected utility theory, despite its drawbacks that will be discussed later, is the framework most widely used by economists for modelling the preferences of economic agents.

We define expected utility as a preference functional $U : \mathcal{X} \mapsto \mathbb{R}$ of the form:

$$U(X) = E[u(X)] = \int_{-\infty}^{+\infty} u(x)dF_X(x) \quad (5)$$

where $E[\cdot]$ is the mathematical expectation operator, F_X is the cumulative probability distribution of X, and u is a real valued increasing function, called a utility function. The utility function u can be interpreted as a non-linear transformation of the agent's wealth, that produces a mechanism for assigning a different weight to each outcome of X. The fact that we require the utility function to be increasing is simply a consequence of the fact that any reasonable agent would prefer more rather than less money (this simple, but important, property is usually referred to as 'increasing preferences').

Given that an agent's preferences are consistent with expected utility theory, a utility function u will exist, such that the agent's preferences are characterised by the expected utility functional $U(X) = E[u(X)]$, in the sense of equation (5). Furthermore, this utility function will be unique up to an affine transformation, i.e. if the utility function u characterises the agent's preferences, so does the function $au(\cdot) + b$, where $a > 0$.

Utility functions have proved very useful for modelling the risk aversion of economic agents. Characterisation of risk aversion via utility functions turns out to be very simple. In the framework of expected utility, risk aversion is equivalent to requiring:

$$E[u(X)] \leq u(E[X]). \quad (6)$$

It follows from Jensen's inequality that concavity of the utility function u is a sufficient condition of risk aversion:

$$u \text{ is a concave function} \Rightarrow E[u(X)] \leq u(E[X]) \quad \forall X \in \mathcal{X}. \quad (7)$$

The concavity of the utility function represents 'diminishing marginal returns', in the sense that, as the risk averse agent's wealth increases, the additional benefit is smaller than the one obtained by the previous unit of wealth. In the context of choosing between risky positions, this property can be interpreted as follows. A risk averse agent with initial wealth w is offered a fair bet: a coin is tossed, and if the outcome is heads the agent receives an amount a; while if the outcome is tails the agent has to pay a. The expected value of the bet is zero, so if the agent is risk averse he will not accept the bet. In terms of a concave utility function, an increase of wealth by a produces a lower increase in the agent's utility than the decrease caused by losing a, and the bet is thus unfavourable.

Concavity of the utility function is a necessary and sufficient condition for the preservation of second stochastic dominance, a stronger assumption than just risk aversion. It can be shown that X second stochastically dominates Y if, and only if, it is $E[u(X)] \geq E[u(Y)]$ for all concave utility functions (Quiggin, 1993).

It is apparent that risk aversion is closely related to the concavity of the utility function. It would be desirable to quantify this relationship and define a corresponding measure of risk aversion. Assuming that the utility function is twice differentiable, a sensible measure of risk aversion is the Arrow-Pratt absolute risk aversion coefficient (risk aversion coefficient for short):

$$ra(w) = -\frac{u''(w)}{u'(w)}. \tag{8}$$

Since the utility function is increasing by definition, it is obvious that the risk aversion is positive if and only if the utility function is concave, equivalently the agent is risk averse. The division by the first derivative of u is carried out in order to ensure that the risk aversion is invariant under affine transformations of the utility function. Note that, given a risk aversion coefficient $ra(w)$, and subject to a normalisation (e.g. $u(0) = 0, u'(0) = 1$), the corresponding utility function can be uniquely determined by solving the differential equation (8).

In some cases an agent might actually be risk seeking. In such a case he will always prefer a risky position with mean R to a safe one equal to R, the corresponding utility function will be convex, and the risk aversion coefficient will be negative.

It can be seen that risk aversion is, in general, a function of the agent's wealth. Thus risk aversion can be decreasing in wealth, as for a logarithmic utility function:

$$u(w) = \ln w \Rightarrow ra(w) = \frac{1}{w} \tag{9}$$

or increasing in wealth, as for a quadratic utility:

$$u(w) = w - \frac{1}{2c}w^2 \ (c > 0, w \leq c) \Rightarrow ra(w) = \frac{1}{c - w}. \tag{10}$$

A very important case is the one of exponential utility, which exhibits constant risk aversion:

$$u(w) = \frac{1}{a}(1 - e^{-aw}) \Rightarrow ra(w) = a. \tag{11}$$

Exponential utility functions are widely used in applications, the main reason being that the assumption of constant risk aversion often yields explicit solutions to otherwise intractable mathematical problems. The independence of risk aversion from wealth is particularly appealing, as in many cases the initial wealth of an economic agent is not known.

2.4 *The Dual Theory of Choice under Risk*

Even though the axioms on which utility theory is built are generally accepted as being reasonable, violations of utility theory often take place in practice. These violations manifest themselves in the famous Elsberg and Allais paradoxes (e.g. Quiggin, 1993). A fairly common violation of expected utility occurs when an individual simultaneously buys insurance (risk averse behaviour) and plays the lottery (which constitutes risk seeking behaviour, since the odds are always against the player). In general, there is empirical evidence that agents tend in their economic decision making to overstate the probabilities of extreme events, both of adverse (such as having a serious accident) and of favourable ones (such as winning the lottery) (Quiggin, 1993).

Attempts to address these problems of utility theory have been made by several authors, including Quiggin (1982) and Machina (1982). In terms of the axiomatic foundations of alternative theories of choice, the point of contention has, in general, been the 'independence axiom' of expected utility theory, which has been substituted by alternative formulations or dropped altogether (Puppe, 1991; Quiggin, 1993), see Appendix A. Here we briefly discuss Yaari's (1987) 'dual theory of choice under risk', which was developed as a complementary approach to expected utility theory. Yaari's theory resolves some of the paradoxes created by expected utility, while creating others of its own.

Yaari (1987) proposes the following preference functional:

$$H(X) = \int_{-\infty}^{0} (h(S_X(x)) - 1)dx + \int_{0}^{\infty} h(S_X(x))dx \tag{12}$$

where $S_X(x) = 1 - F_X(x) = \mathbb{P}(X > x)$ is the decumulative probability distribution function (survival function) of X, while $h : [0, 1] \mapsto [0, 1]$ is an increasing function satisfying $h(0) = 0$ and $h(1) = 1$, which we will call a *probability distortion function*. Without loss of generality, let us, for simplicity, assume that $X \geq 0$, such that:

$$H(X) = \int_{0}^{\infty} h(S_X(x))dx. \tag{13}$$

Observe that the mean of X can be written as:

https://doi.org/10.1017/S1357321700004414 Published online by Cambridge University Press

$$E[X] = \int_0^\infty S_X(x)dx. \tag{14}$$

Also observe that the function $h(S_X(x))$ is itself a decumulative distribution, though different from the original one. Thus, from equation (13) it can be seen that the effect of the probability distortion h is to modify ('distort') the probability distribution of the random portfolio. The preference functional is the expected value of the portfolio with respect to the modified distribution function. In expected utility theory the preference functional is given as the expected value of a non-linear transformation of wealth; in the dual theory it is the expected value of wealth under a non-linear transformation of the probability distribution.

Similarly to the case of expected utility theory, risk aversion can be simply characterised in terms of the distortion function. It follows from equations (14) and (13) that an agent is risk averse if, and only if, his distortion function is always below the line $h(s) = s$:

$$H(X) \leq E[X] = H(E[X]) \; \forall X \in \mathcal{X} \Leftrightarrow h(s) \leq s \quad \forall s \in [0, 1]. \tag{15}$$

However, for the agent's preferences to preserve second stochastic dominance, a more stringent condition on the distortion function is required (Quiggin, 1993):

$$\begin{gathered}\text{The operator } H() \text{ preserves second stochastic dominance on } \mathcal{X} \\ \Leftrightarrow h \text{ is a convex function.}\end{gathered} \tag{16}$$

Thus, in the dual theory of choice, risk aversion is closely related to the convexity of the distortion function. To better understand the effect of the distortion function, rewrite equation (13) as:

$$H(X) = \int_0^\infty x h'(S_X(x)) f_X(x)dx = E[X h'(S_X(X))] \tag{17}$$

where f_X is the probability density of X. Because h is convex, its first derivative h' is an increasing function, and thus $h'(S_X(x))$ is decreasing in x, i.e. $h'(S_X(x))$ becomes small for large values of the random variable X. We can thus regard $h'(S_X(x))$ as a weighting mechanism which discounts the probability of desirable events (high X), while loading the probability of adverse events (low X). In expected utility theory risk aversion is induced by exaggerating the effects of adverse events; in the dual theory this is done by exaggerating their probability.

To give a measure of risk aversion in the dual theory of choice, we define the *uncertainty aversion coefficient*, which is a dual to the Arrow-Pratt risk aversion coefficient of utility theory:

$$ua(w) = \frac{h''(s)}{h'(s)}. \tag{18}$$

Examples of probability distortion functions are the dual power distortion (which has increasing uncertainty aversion):

$$h(s) = 1 - (1-s)^q \quad 0 < q \leq 1 \Rightarrow ua(s) = (1-q)(1-s)^{-1} \tag{19}$$

and the exponential distortion function (which has constant uncertainty aversion):

$$h(s) = \frac{e^{\beta s} - 1}{e^{\beta} - 1} \quad \beta > 0 \Rightarrow ua(s) = \beta. \tag{20}$$

2.5 *Generalised Expected Utility*

The von Neumann-Morgenstern utility theory and Yaari's (1987) dual theory of choice are complementary approaches. In fact, Yaari presented his theory less as a complete characterisation of agents' preferences, but more as an illustrative example of preference functionals and orderings alternative to the ones produced by expected utility.

We can actually produce a comprehensive description of an economic agent's preferences by combining the two above-mentioned approaches. (A rigorous axiomatic characterisation of such preferences is given by Schmeidler (1989), while the book by Quiggin (1993) is also dedicated to the subject). Thus, we equip our agent with both a utility u and a probability distortion h. This yields the generalised expected utility operator:

$$V_{u,h}(X) = \int_{-\infty}^{0} (h(S_{u(X)}(x)) - 1)dx + \int_{0}^{\infty} h(S_{u(X)}(x))dx. \tag{21}$$

The preference functional $V_{u,h}$ represents expected utility, calculated under a distorted probability distribution.

In the context of generalised utility theory, second stochastic dominance again gives us a clue about the shape of the utility and distortion functions. The preference functional $V_{u,h}$ preserves second stochastic dominance if, and only if, the utility function u is concave *and* the distortion function h is convex (Quiggin, 1993). These two conditions will be standing assumptions for the rest of this paper.

3. Risk Measures and their Properties

3.1 *Definition of Risk Measures*

Risk measures are defined as functions that take random variables, which represent the terminal value of assets and/or liabilities at a fixed later date, as arguments. Risk measures return a real number which encapsules the risk assessment of the portfolio, whose random value is used as an argument. The output of the risk measure represents the amount of additional capital that the holder of the portfolio must add to his position in order that a regulator, or a rating agency, or indeed the agent himself deems the aggregate position acceptable (this definition is due to Artzner *et al*. (1999)). Formally, a risk measure ρ is defined as a real valued functional on a collection of portfolios $\mathcal{X}$. Hence the holder of the risky investment X has to keep an amount $\rho(X)$ safely invested (for simplicity, in our paper this means with zero interest). The higher the risk assessment (as produced by the risk measure) of the position X, the more will be the capital that the agent has to reserve.

3.2 *Risk Measures and Premium Principles*

Risk measures are in many respects akin to actuarial premium calculations principles. For an insurance carrier exposed to a liability X, a premium calculation principle Π gives the minimum amount $\Pi(X)$ that the insurer must raise from the insured in order that it is in his interest to proceed with the contract. Thus, a premium calculation principle can be directly interpreted as a risk measure. A notational issue emerges here, because in the actuarial literature insurance liabilities are usually denoted by positive random variables, while in the more general risk management literature liabilities are denoted by negative variables. In order not to contradict either convention, we will use the notation $\Pi(X) = \rho(-X)$ for a premium calculation principle and the associated risk measure. The relation between risk measures and premium calculation principles is further discussed in Section 4.1.

3.3 *Properties of Risk Measures*

Before discussing specific classes of risk measures emerging from the three theories of choice examined in the previous sections, we give a brief exposition of possible (sets of) desirable properties that risk measures and the associated premium principles should satisfy.

A first natural requirement that we can impose on a risk measure is that a portfolio whose value at the end of one period is always higher than that of another should also induce lower capital requirements. This property, *monotonicity*, can be summarised as:

Monotonicity: If $X \leq Y$ almost surely, then $\rho(X) \geq \rho(Y)$.

https://doi.org/10.1017/S1357321700004414 Published online by Cambridge University Press

(Note that, in terms of premium calculation principles, monotonicity is expressed as $\Pi(X) \leq \Pi(Y)$ for $X \leq Y$.)

Recall that a risk measure has been defined as a function giving the amount of safely invested capital that the holder of a risky position has to add to his position in order to make it acceptable. Then it is obvious that, should the investor add a cash amount to his position, the risk of the aggregate should be reduced by an equal amount. This consideration induces the *translation invariance* property of risk measures:

Translation invariance: If $a \in \mathbb{R}$ then $\rho(X + a) = \rho(X) - a$.

(In terms of premium calculation principles, translation invariance is expressed as $\Pi(X - a) = \Pi(X) - a$.)

The behaviour of a risk measure with respect to the aggregation of risky positions, manifested by the award of diversification discounts or the imposition of penalties, is crucial. The question can be posed as follows: given two portfolios X and Y and their joint probability distribution, how does the risk of the aggregate $\rho(X + Y)$ relate to the risks of the individual positions $\rho(X)$ and $\rho(Y)$?

The answer to this question might be related to the way in which X and Y stochastically depend on each other. Let us now present some well-known notions (e.g. Joe, 1997) of dependence that we will refer to in this section. A useful characterisation of positive dependence is *positive quadrant dependence* (PQD). Two risks X, Y are PQD whenever:

$$\mathbb{P}(X \leq x \text{ and } Y \leq y) \geq \mathbb{P}(X \leq x)\mathbb{P}(Y \leq y) \quad \forall x, y \in \mathbb{R}. \tag{22}$$

Essentially two risks being PQD means that their probability of assuming low (or high) values simultaneously is higher than it would be, were they independent. The negative analogue of PQD is *negative quadrant dependence* (NQD):

$$\mathbb{P}(X \leq x \text{ and } Y \leq y) \leq \mathbb{P}(X \leq x)\mathbb{P}(Y \leq y) \quad \forall x, y \in \mathbb{R}. \tag{23}$$

Finally, *comonotonicity* is defined as the strongest possible form of positive dependence. Two random variables X, Y are called comonotonic if there is a random variable U and non-decreasing functions g, h such that:

$$X = g(U),\ Y = h(U). \tag{24}$$

We now return to the issue of defining suitable aggregation properties for risk measures. Few would disagree with the claim that, if the positions X and Y are negatively related, then to some extent the one risk is a hedge for the other, and if X and Y are pooled, this fact should be acknowledged with a reduction in risk capital. Thus, a risk measure should be *subadditive for NQD risks:*

Subadditivity for NQD risks: If X, Y are NQD, then $\rho(X+Y) \leq \rho(X) + \rho(Y)$.

Opinions are divided as to what should happen when positive dependence occurs. On the one hand one could argue (as Dhaene *et al*. (2003) do extensively) that aggregating positively dependent risks actually increases the riskiness of the portfolio and that this should induce higher capital requirements. Thus, in addition to subadditivity for NQD risks we could require:

Superadditivity for PQD risks: If X, Y are PQD, then $\rho(X+Y) \geq \rho(X) + \rho(Y)$

and

Additivity for independent risks: If X, Y are independent, then $\rho(X+Y) = \rho(X) + \rho(Y)$.

A different school of thought, represented by Wang *et al*. (1997) and Artzner *et al*. (1999), holds that risk measures should be subadditive for any dependence structure:

Subadditivity: $\rho(X+Y) \leq \rho(X) + \rho(Y)$ for any $X, Y \in \mathcal{X}$.

The argument for subadditivity is that even positively dependent risks provide some diversification, so there is no reason that capital requirements should be increased. From a regulatory perspective, Artzner *et al*. (1999) argue that there should be no incentives for investors to split their portfolios, which would be the result of superadditive risk measurement.

In a subadditive risk measure, positive dependence can be acknowledged by requiring that, for comonotonic risks, it becomes additive:

Additivity for comonotonic risks: If X, Y are comonotonic, then $\rho(X+Y) = \rho(X) + \rho(Y)$.

That is equivalent to saying that, since perfect positive dependence means that pooling X and Y does not produce any diversification, there is no reason for the risk measure to be subadditive.

Artzner *et al*. (1999) proposed the further requirement of *positive homogeneity* on a risk measure:

Positive homogeneity: If $a \in \mathbb{R}_+$ then $\rho(aX) = a\rho(X)$.

The interpretation that they gave for this property was the fairly uncontroversial statement that a change in units of measurement (e.g. currency) should not result in a change in capital requirements. However, positive homogeneity can also be understood in a different way: that changes in the *size* of an agent's portfolio, given that its composition is unchanged, should only affect a proportional change in capital requirements. Such an assertion has not been without its critics, see for example Föllmer & Schied (2002a) and Dhaene *et al*. (2003). The criticism can be summarised in the observation that positive homogeneity does not account for liquidity risk. Positive homogeneity does not acknowledge that very large portfolios of risks might produce very high losses that, in turn, can make it difficult for the holder of the portfolio to raise enough cash in order to meet his obligations. Note that positive homogeneity is consistent with comonotonic additivity, while being contradictory to superadditivity for PQD risks.

A relaxation of positive homogeneity and subadditivity has been proposed by Föllmer & Schied (2002a), who replaced both properties with the requirement for *convexity*:

Convexity: $\rho(\lambda X + (1-\lambda)Y) \leq \lambda\rho(X) + (1-\lambda)\rho(Y)$ for all $X \; Y \in \mathcal{X}$, $\lambda \in [0, 1]$.

Convexity has the interpretation of reduction in risk brought about by diversification, while not implying the insensitivity to risk aggregation and liquidity risk brought about by the combination of subadditivity and positive homogeneity.

We conclude with some terminology. Risk measures that satisfy monotonicity, translation invariance, subadditivity and positive homogeneity are called *coherent* (Artzner *et al*., 1999). If additivity for comonotonic risks is added to these properties, we end up with a specific sub-class of coherent risk measures called *distortion premium principles* (Wang, 1996; Wang *et al*., 1997). On the other hand, risk measures satisfying monotonicity, translation invariance and convexity are called *convex* (Föllmer & Schied, 2002a). Finally, risk measures that are additive for independent risks are called *additive* (Gerber, 1974).

4. Risk Measures Induced by Theories of Choice

4.1 *Risk Measures and Indifference Arguments*

Risk measures provide, in some sense, an ordering of portfolios with random payoffs in terms of their riskiness. In this section we will be concerned with the way in which such an ordering can be associated with the preference orderings stemming from theories of choice that were discussed previously. In other words, given a risk measure ρ, we are interested in preference functionals V such that:

$$\rho(X) \leq \rho(Y) \Leftrightarrow V(X) \geq V(Y) \quad \forall X, Y \in \mathcal{X}. \tag{25}$$

A relation such as the above emerges when producing risk measures from preference functionals, via indifference arguments. Such a process can be explained simply through the example of pricing an insurance contract. In an insurance context, an interpretation of risk measures is as premium calculation mechanisms. The premium $\Pi(X)$ that an insurer asks for insuring a liability $-X$ should be equal to the capital $\rho(-X)$ that the insurer has then to reserve in order to remain solvent. (This is, of course, a simplified version of reality. We are not concerned here with the more realistic case where the insurer invests the premium in, say, stocks, thus achieving a higher return, but being exposed to new sources of risk.) Consider now an insurer with initial surplus (cash) w, who insures a liability $-X$. The insurer's preferences are described by a preference functional $V : \mathcal{X} \mapsto \mathbb{R}$, such as the ones encountered in Section 2. The premium $\Pi(X)$ that the insurer will charge can be determined via an indifference argument, i.e. it should be the sum for which it is indifferent to the insurer whether to go ahead with the contract or not. Such indifference can be formalised by requiring that evaluations of the preference functional before and after the contract yield the same result:

$$V(w) = V(w - X + \Pi(X)). \tag{26}$$

The premium $\Pi(X)$ will then be obtained as a solution to the above equation. As discussed above, a risk measure can then be defined as:

$$\rho(X) = \Pi(-X) \quad \forall X \in \mathcal{X}. \tag{27}$$

From now on, whenever the relationship (27) holds, we will call Π the premium calculation principle associated with the risk measure ρ and vice versa. The minus sign essentially comes from the fact that in premium calculation positive outcomes are generally considered to be losses, while in the more general framework of risk measurement positive values of the portfolio denote gains.

In Sections 4.2, 4.3 and 4.4 we show how classes of premium principles (and the associated risk measures), satisfying alternative sets of properties, can be derived from the economic theories of choice discussed earlier. The comparison between these premium principles is summarised in Appendix B.

4.2 *Risk Measures Derived from Expected Utility*

4.2.1 *Zero-utility premium calculation*

In this section we consider agents whose preferences conform to the

https://doi.org/10.1017/S1357321700004414 Published online by Cambridge University Press

theory of expected utility discussed in Section 2.3. Consider an insurer with a utility function u and initial surplus w, who prices a liability $-X$. The indifference argument (26) now becomes:

$$u(w) = E[u(w - X + \Pi_u(X))]. \tag{28}$$

If w is set to zero, the above equation becomes:

$$u(0) = E[u(-X + \Pi_u(X))]. \tag{29}$$

The premium calculation principle $\Pi_u(\bullet)$ that emerges as the solution of (29) is called the *principle of zero utility* (Bühlmann, 1970). Note that the zero-utility principle is monotonic and translation invariant, as a direct result of equation (29). Moreover, all zero-utility principles associated with concave utility functions are consistent with the ordering of risks according to second stochastic dominance, as discussed in Section 2.3.

Equation (29) has, in general, no analytical solution. However, in the special case of a quadratic utility function (10), a solution does exist:

$$\Pi_{u_q}(X) = E[X] + c - \sqrt{c^2 - \sigma^2(X)} \tag{30}$$

which, for small values of the variance $\sigma^2(X) \ll c$, can be approximated by (Bühlmann, 1970):

$$\Pi_{u_q}(X) \simeq E[X] + \frac{\sigma^2}{2c}. \tag{31}$$

This premium calculation principle is termed *the variance principle*.

4.2.2 *The exponential premium principle*

A very interesting premium principle emerges when the utility function is of the exponential type (11). Then equation (28) yields the following solution (the exponential utility function's constant risk aversion makes the initial surplus irrelevant, and we do not need to set it to zero):

$$\Pi_{u_{exp}}(X) = \frac{1}{a} \ln E[e^{aX}] \qquad a \in (0, \infty) \tag{32}$$

(where it is understood that for $a = 0$ the premium principle reduces to the net premium, $\Pi_{u_{exp}|a=0}(X) = E[X]$). This premium calculation mechanism is called the *exponential premium principle*, and has been proposed by several authors, see for example Gerber (1974), Bühlmann (1985). The risk measure associated with the exponential premium principle is:

$$\rho_{u_{exp}}(X) = \frac{1}{a} \ln E[e^{-aX}]. \qquad (33)$$

Some important properties of the exponential premium principle and the associated risk measure are (Gerber, 1974; Dhaene *et al.*, 2003):
— monotonicity;
— translation invariance;
— subadditivity for negatively related (NQD) risks;
— additive for independent risks; and
— superadditivity for positively related (PQD) risks.

The last two properties imply that the aggregation of positively dependent investment positions increases the risk of the portfolio superlinearly. Note that this implies that the premium principle and associated risk measure are not positively homogenous. Thus, the exponential premium principle is sensitive to the liquidity risk and aggregation issues discussed in Section 3.3. These properties are a direct result of the association with expected utility theory, since an increase in the scale of losses would lead to a steeper than linear reduction in utility. We also note that the exponential premium principle (with the net premium as a special case) is the only zero-utility premium principle that is additive for independent risks (Gerber, 1974).

We consider, however, the properties of the exponential principle as being excessively strict, as far as relatively small portfolios are concerned. The superadditivity for positively related risks is present even for very small portfolios, where liquidity risk and aggregation are not of primary importance. Furthermore, for small portfolios the exponential principle will approximately equal the net premium, and thus no safety loading will be produced.

4.2.3 *Exponential premium principle and ruin theory*

An alternative interpretation of the exponential premium principle can be given in terms of ruin theory, see for example Bühlmann (1985). Consider an insurer with an initial surplus S_0. Each year t the insurer is exposed to stochastic insurance losses X_t and has a premium income c_t. Thus, the insurer's surplus process is:

$$S_t = S_{t-1} + c_t - X_t \qquad t = 1, 2, \dots . \qquad (34)$$

Assume now that the annual total claims $X_t, t = 1, 2, \dots,$ are identically and independently distributed. Denote as X a random variable with the same distribution as X_t. Let the premium c_t be given by premium principle Π, i.e. $c_t = \Pi(X_t)$. Since the liabilities X_t have same distribution each year, each year's premium will also be the same, so that $c_t = c = \Pi(X)$.

Ruin is defined as S_t becoming negative at some time $t > 0$. Denote the probability of ruin as ψ. The question posed now is: "Given the premium calculation principle Π, what is the probability of ruin?" A similar question is: "Can we control the probability of ruin by choosing an appropriate risk measure?" Assuming that X has exponentially bounded tails, the probability of ruin ψ is bounded by:

$$\psi \leq e^{-\lambda S_0} \tag{35}$$

where λ is called the 'adjustment coefficient' and is the solution of the equation:

$$e^{\lambda c} = E[e^{\lambda X}] \tag{36}$$

which in turn yields:

$$c = \frac{1}{\lambda} \ln E[e^{\lambda X}]. \tag{37}$$

Thus, calculating the premium by the exponential principle with a risk aversion coefficient of λ introduces an upper bound of $e^{-\lambda S_0}$ on the probability of ruin. It can be seen that the higher λ is, the lower the probability of ruin. This relation can be interpreted as a specific upper bound on the probability of ruin, implying a risk aversion parameter for the insurer and vice versa. It is also important that the probability of ruin depends on the initial surplus; the lower that is, the more likely ruin becomes.

4.3 *Risk Measures Derived from the Dual Theory of Choice*

4.3.1 *The distortion premium principle*

Here we repeat the indifference argument of Section 4.1 for the case that preferences are modelled by the dual theory of choice under risk that was discussed in Section 2.4. Thus, risk aversion is induced by an increasing convex probability distortion function h, and the resulting preference functional is of the form (12). The indifference argument (26) becomes:

$$H_h(w) = H_h(w - X + \Pi_h(X)). \tag{38}$$

This equation can be explicitly solved for the premium Π_h:

$$\Pi_h(X) = -H_h(-X) = -\left\{ \int_{-\infty}^{0} (h(S_{-X}(x)) - 1) dx + \int_{0}^{\infty} h(S_{-X}(x)) dx \right\} \tag{39}$$

and the corresponding risk measure is:

https://doi.org/10.1017/S1357321700004414 Published online by Cambridge University Press

$$\rho_h(X) = -H_h(X) = -\left\{ \int_{-\infty}^{0} (h(S_X(x)) - 1)dx + \int_{0}^{\infty} h(S_X(x))dx \right\}. \qquad (40)$$

We can rewrite equation (39) in the simpler form:

$$\Pi_h(X) = \int_{-\infty}^{0} (g(S_X(x)) - 1)dx + \int_{0}^{\infty} g(S_X(x))dx \qquad (41)$$

where $g(s) = 1 - h(1 - s)$ is called the *conjugate distortion function* of h. When the liability X is strictly positive we end up with only the second term in the above equation, which was introduced in the context of insurance pricing by Denneberg (1990) and Wang (1996), and is often known as the *distortion premium principle*.

It can be shown (Denneberg, 1990; Wang *et al*., 1997) that, whenever h is convex (equivalently g is concave), the distortion premium principle satisfies the following five key properties:

— monotonicity;
— translation invariance;
— positive homogeneity;
— subadditivity; and
— additivity for comonotonic risks.

The first four of these properties ensure that the distortion principle is a coherent risk measure in the sense of Artzner *et al.* (1999), while comonotonic additivity is a desirable attribute for a subadditive risk measure, as already discussed in Section 3.3. Furthermore, it can be shown that every premium principle (risk measure) that satisfies the five properties above can be represented as a distortion premium principle, subject to a technical condition (Wang *et al*., 1997).

Note that, because of its association with the dual theory of choice, the distortion premium principle is also consistent with second stochastic dominance. The association with the dual theory is also the reason behind the property of positive homogeneity. It was seen in Section 2.4 that, under the dual theory, the agents are risk neutral with respect to the absolute size of potential losses or gains, since the distortion function affects only the probabilities of adverse events. Thus, positive homogeneity is a consequence of the linearity in wealth that the dual theory implies. This makes the distortion principle insensitive to liquidity risk and risk aggregation.

4.3.2 *Probability distortion and sets of measures*

We conclude with another representation of distortion premium principles, via sets of probability measures. A probability measure $\mathbb{Q} : \mathcal{F} \mapsto \mathbb{R}$ is a real-valued set function defined on the collection (σ-algebra) of events

https://doi.org/10.1017/S1357321700004414 Published online by Cambridge University Press

$\mathcal{F}$ that assigns a probability to each of those events. Denote as $\mathbb{P}$ the 'real-world' or 'actuarial' probability measure, and as $E_{\mathbb{Q}}[X]$ the expected value of X with respect to the measure $\mathbb{Q}$. Then the distortion premium principle (39) and the corresponding risk measure can be written as:

$$\Pi_h(X) = \sup_{\mathbb{Q}\in\mathcal{Q}_g} E_{\mathbb{Q}}[X] = -\inf_{\mathbb{Q}\in\mathcal{Q}_h} E_{\mathbb{Q}}[-X] \tag{42}$$

and

$$\rho_h(X) = \sup_{\mathbb{Q}\in\mathcal{Q}_g} E_{\mathbb{Q}}[-X] = -\inf_{\mathbb{Q}\in\mathcal{Q}_h} E_{\mathbb{Q}}[X] \tag{43}$$

where the sets of probability measures $\mathcal{Q}_g$, $\mathcal{Q}_h$ are defined as:

$$\begin{aligned} \mathcal{Q}_g &= \{\mathbb{Q} : \mathbb{Q}(A) \leq g(\mathbb{P}(A)), \quad \forall A \in \mathcal{F}\} \\ \mathcal{Q}_h &= \{\mathbb{Q} : \mathbb{Q}(A) \geq h(\mathbb{P}(A)), \quad \forall A \in \mathcal{F}\}. \end{aligned} \tag{44}$$

The representation (42) can be interpreted as the expected loss $E_{\mathbb{Q}}[X]$ of X under the 'worst possible' measure in $\mathcal{Q}_g$. The measures $\mathbb{Q}$ are sometimes called 'generalised scenarios' (Artzner *et al.*, 1999). Thus, the distortion premium principle can be re-interpreted as a 'worst-scenario' approach to calculating expected loss over the appropriate set of risk adjusted test probabilities. On the other hand, we can consider the presence of a set of probability measures as a way of quantifying uncertainty with respect to the distribution function of the underlying risk (Schmeidler, 1989). The extent of such uncertainty is determined by the concavity of g (convexity of h) through which the set of measures $\mathcal{Q}_g$ ($\mathcal{Q}_h$) is defined.

4.4 *Generalised Expected Utility and Convex Measures of Risk*

It is an obvious step now to synthesise the two approaches for producing risk measures that were shown in the previous two sections. We assume that an agent's preferences are summarised by both a concave utility u and a convex distortion h. Thus, the indifference argument (26) takes the form:

$$V_{u,h}(w) = V_{u,h}(w - X + \Pi_{u,h}(X)). \tag{45}$$

As the premium principle $\Pi_{u,h}$ emerges from generalised expected utility, it inherits properties both from expected utility and the dual theory of choice. It is thus easily understood that, as discussed by Luan (2001), in the case of a linear utility function the distortion principle is obtained as a special case, while, in the case of a linear probability distortion (and $w = 0$) $\Pi_{u,h}$ reduces to the zero-utility principle. In general, the properties of $\Pi_{u,h}$ will lie somewhere in between those of these two extreme cases. However, its precise behaviour is difficult to pin down. It is, in general, difficult to determine at which point the utility or the distortion function will have the most influence (Quiggin, 1993). One important property is that it is consistent with second order stochastic dominance whenever u is concave and h is convex (Quiggin, 1993). A very useful characterisation has been given by Föllmer & Schied (2002b), who have shown, in a slightly different context, that risk measures emerging from indifference arguments such as (45) are, in fact, convex measures of risk. Thus, the premium principle $\Pi_{u,h}$ and the associated risk measure $\rho_{u,h}(X) = \Pi_{u,h}(-X)$ are monotonic, translation invariant and convex.

5. A Distortion-Exponential Premium Principle

In this section we propose a premium principle (and the corresponding risk measure) which is obtained as a special case of the risk measures considered in Section 4.4 when an exponential utility function is used. This premium principle is mathematically tractable, while having quite interesting properties, and we call it the *distortion-exponential principle*.

5.1 *Definition*

Reconsider equation (45), with an exponential utility function u_{exp} and an arbitrary convex distortion function h. As seen in Section 4.2, calculations involving exponential utilities are much easier, and tend to yield explicit solutions. Then the premium principle $\Pi_{u_{exp},h}(X)$, which we may call the *distortion-exponential principle*, becomes:

$$\Pi_{u_{exp},h}(X) = \frac{1}{a}\ln\left\{\int_{-\infty}^{0}(g(S_{e^{aX}}(x)) - 1)dx + \int_{0}^{\infty} g(S_{e^{aX}}(x))dx\right\} \tag{46}$$

where $g(s) = 1 - h(1 - s)$ is the (concave) conjugate distortion function. The associated risk measure can be defined by $\Pi_{u_{exp},h}(-X)$.

We note that the distortion-exponential principle can be concisely represented via the set of probability measures $\mathcal{Q}_g$ (44):

$$\Pi_{u_{exp},h}(X) = \frac{1}{a}\ln\left\{\sup_{\mathbb{Q}\in\mathcal{Q}_g} E_{\mathbb{Q}}[e^{aX}]\right\}. \tag{47}$$

Now we can rewrite the above equation as:

$$\Pi_{u_{exp},h}(X) = \sup_{\mathbb{Q}\in\mathcal{Q}_g}\left\{\frac{1}{a}\ln E_{\mathbb{Q}}[e^{aX}]\right\} = \sup_{\mathbb{Q}\in\mathcal{Q}_g}\Pi_{u_{exp}}(X). \tag{48}$$

Thus, the distortion-exponential principle is interpreted as an exponential premium principle, evaluated with respect to the worst-case probability measure in $\mathcal{Q}_g$. If we interpret the set of scenarios $\mathcal{Q}_g$ as quantifying the uncertainty regarding the probability distribution of a portfolio X, the above representation can be viewed as being derived from the ruin problem discussed in Section 4.2. Thus, the distortion-exponential principle provides a mechanism for controlling the probability of ruin, in the presence of distributional uncertainty, modelled via distorted probabilities.

5.2 *Properties*

As the distortion-exponential principle emerges from generalised expected utility, it inherits properties both from the exponential premium principle and risk measures based on distorted probabilities. Thus, the properties of this composite risk measure will lie somewhere between these two extremes of economic behaviour. As discussed in the previous section, the distortion-exponential principle and the associated risk measure satisfy:

— monotonicity;
— translation invariance; and
— convexity.

It is easily seen that the exponential and distortion principles, obtained in the previous two sections, are special cases of this more general premium principle. Thus, if we let the utility function become linear (i.e. let the risk aversion coefficient tend to zero), then convexity breaks down into positive homogeneity and subadditivity, and we obtain a coherent risk measure. On the other hand, if the distortion function is linear, convexity reaches its other extreme and can be substituted by additivity for independent risks, subadditivity for NQD risks and superadditivity for PQD ones.

5.3 *Comparison of the Distortion-Exponential to the Distortion and Exponential Principles*

In the case that neither the utility nor the distortion function is linear, whether the properties of the proposed risk measure are closer to the exponential or the distortion principle will, to some extent, depend on the underlying risks that are examined. Consider an insurance liability with loss ratio $X \geq 0$, and an insurance company holding λ units of the liability, i.e. being exposed to risk λX. Note that the effect of the exponential function in (46) will then depend on the product $a\lambda$. When λ is small in relation to a, then the aggregate effect is that the exponential term will be very similar to a linear one. Thus, we can claim that, for small portfolios the distortion-exponential principle becomes asymptotically equivalent to the corresponding distortion principle, and inherits its properties. Thus, for small portfolios the distortion-exponential principle will be positively homogenous and subadditive. However, for higher values of λ the exponential term grows, and we would expect the effect of the utility function to become prevalent.

5.3.1 *Sensitivity to portfolio size*

The main criticism against positive homogeneity is that it disregards liquidity risk. However, for small portfolios liquidity risk can be negligible, hence it is reasonable to require that for such portfolios the risk measure becomes approximately invariant to the scale of potential losses. The influence of the exponential utility function on the risk assessment of portfolios with variable sizes is illustrated in Figures 1 and 2. We assume X is Gamma distributed with mean equal to 1 and variance equal to 0.2. First we consider the case of a linear distortion function and an exponential utility with risk aversion $ra = 0.5 \cdot 10^{-6}$. The corresponding risk assessment of λX for $0 < \lambda \leq 5 \cdot 10^{6}$, given by the exponential premium principle, is shown in Figure 1, along with the expected loss (net premium) of λX. It can be seen that, for small values of λ, the two lines are almost indistinguishable, while

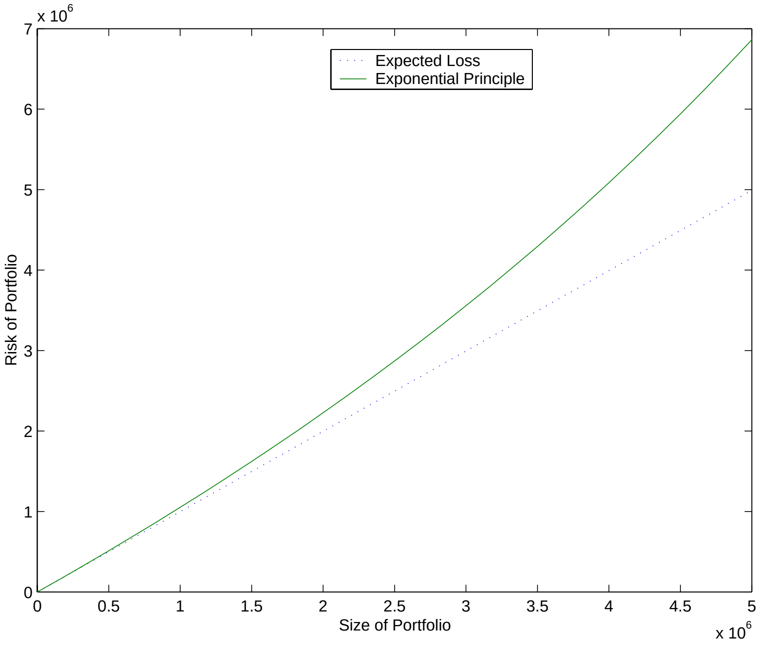

Figure 1. Effect of portfolio size on risk assessment with the exponential premium principle

https://doi.org/10.1017/S1357321700004414 Published online by Cambridge University Press

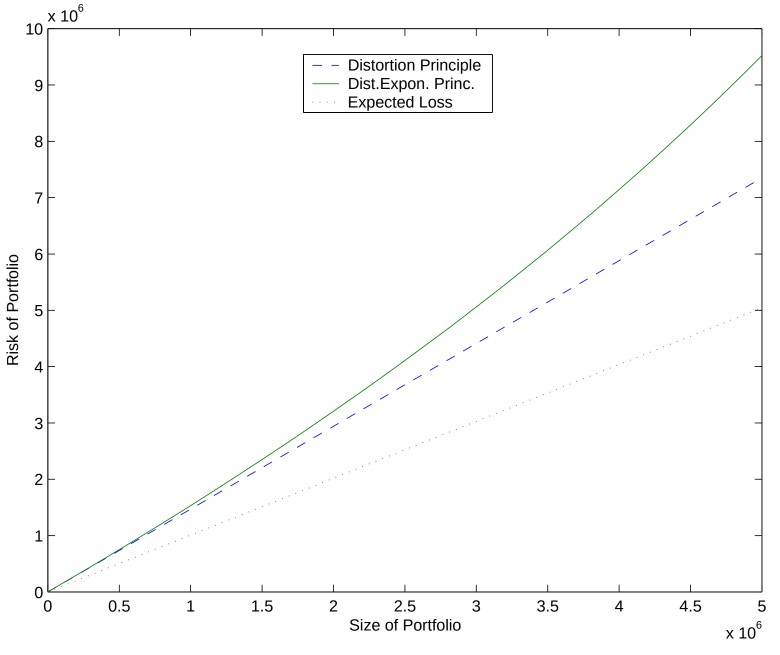

Figure 2. Comparison of the effect of portfolio size on risk assessment, for the distortion and distortion-exponential premium principles

when λ becomes comparable to $1/a$ the line representing risk assessment with the exponential premium principle starts to diverge. Thus, risk aversion only starts to 'kick-in' when the size of the portfolio is such that a potential loss becomes a significant threat to the liability holder's solvency. For small values of λ the line corresponding to the exponential premium principle is approximately straight, which can be interpreted as (approximate) positive homogeneity (scale invariance) at that level. However, the fact that, for small λ, the exponential principle gives the same values as the net premium is something that one could disagree with, as it implies that, for small portfolios, no safety loading is necessary.

This weakness is rectified if we use, instead of the exponential premium, the distortion-exponential principle, where an exponential distortion function, with uncertainty aversion $ua = 4$, is used. In Figure 2 the behaviour of the distortion-exponential principle with respect to portfolio size is depicted, along with risk assessments produced with the corresponding distortion principle and expected loss. It can be seen that, for small portfolios, the risk given by distortion-exponential principle is

https://doi.org/10.1017/S1357321700004414 Published online by Cambridge University Press

asymptotically the same as the one calculated with the distortion principle, thus it is again approximately positively homogenous. On the other hand, positive homogeneity does not imply here the absence of a safety loading, as is seen by the different slopes of the straight lines corresponding to the distortion principle and the expected loss. A safety loading, even for small portfolios, is the result of the uncertainty aversion introduced by the distortion function, while, when the exponential utility's non-linear effect becomes significant, an additional loading is present, resulting from the sensitivity to very high potential losses. Thus, with the distortion-exponential principle we get the best of both worlds: positive homogeneity for small portfolios; a proportional safety loading for all portfolio sizes; plus a non-linear loading for large portfolios that pose an increased liquidity risk.

5.3.2 *Sensitivity to risk aggregation*

We now turn our attention to the issue of risk aggregation. As discussed previously, the distortion principle is insensitive to risk aggregation issues, as it is always subadditive. On the other hand, it can be argued that the exponential premium principle is over-sensitive to risk aggregation, since it is superadditive for all PQD risk, even for small ones, and thus makes no acknowledgement of the possible diversification implicit in risk pooling (in fact using the exponential principle would create an incentive to split even the smallest portfolio of positively related risks). Consider two random liabilities X and Y. Given a premium principle Π, we define the relative saving from pooling $\gamma(X, Y)$ as the proportional reduction in premium (risk capital) when X and Y are pooled:

$$\gamma(X, Y) = \frac{\Pi(X) + \Pi(Y) - \Pi(X + Y)}{\Pi(X) + \Pi(Y)}. \tag{49}$$

If X and Y are positively related (PQD to be more precise), for a distortion principle $\gamma(X, Y)$ will always be positive, due to subadditivity, and for an exponential principle it will be always negative, due to superadditivity. As discussed before, neither of those two extremes is satisfactory, since the former disregards risk aggregation while the latter does not acknowledge diversification.

Again, the distortion-exponential principle occupies the middle ground, and provides an answer that is, in our view, satisfactory. Let X and Y be Gamma distributed, both with mean 0.5 and variance 0.05, and positively dependent with correlation 0.5 (in this example we use a Gaussian copula (Joe, 1997) to model dependence). In Figure 3 we plot $\gamma(\lambda X, \lambda Y)$ as a function of λ for the distortion principle, the exponential principle and the distortion-exponential principle. It can be seen that, as expected, the

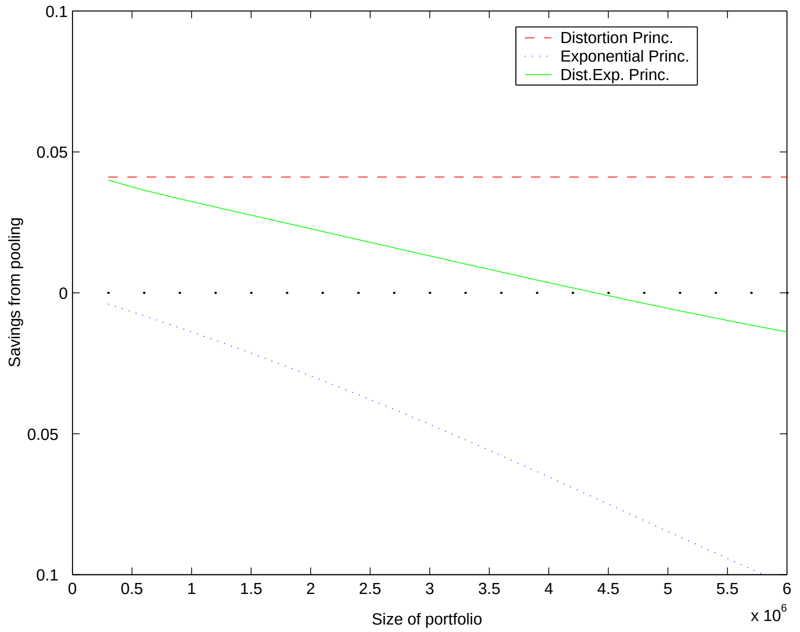

Figure 3. Comparison of savings from pooling positively correlated liabilities, with the distortion, exponential and distortion-exponential premium principles

savings from pooling are always positive for the distortion principle and always negative for the exponential principle. For the distortion-exponential principle the savings are positive for small λ and negative for large ones; the point where the line crosses the x-axis is the one when the dangers from risk aggregation become more significant than the benefits from risk pooling.

Overall, the properties of the risk measure that we introduced in this paper lie between those of the exponential and distortion premium principles. For small portfolios it behaves like a coherent risk measure, while for larger ones, where liquidity risk and risk aggregation become significant, the exponential function progressively introduces superlinearity and superadditivity in the risk assessment. The definition of what a large portfolio is will depend on the specific situation and preferences of its holder, and can be controlled by modifying the risk aversion parameter of the exponential function.

https://doi.org/10.1017/S1357321700004414 Published online by Cambridge University Press

6. Conclusions

In this paper we have discussed alternative theories of choice and the properties of risk measures that are associated with those theories. Our purpose has been to show that these two complementary perspectives of the problem of risk measurement illuminate each other. Thus, risk measures resulting from expected utility theory are sensitive to the relative scale of potential losses, which is a direct result of the non-linearity of the utility function. We have focused on the case of exponential utility functions, as the resulting risk measures, called exponential premium principles, are characterised by a transparent set of properties, while also having an interesting interpretation in terms of the classical ruin problem. The exponential premium principle has been shown to be particularly sensitive to risk aggregation. On the other hand, coherent risk measures derived from the dual theory of choice, called distortion principles, have been shown to be insensitive to risk aggregation, as they are subadditive for all risks. However, for small portfolios of risks, where liquidity and aggregation issues are less prevalent, distortion principles both apply a proportional safety loading and acknowledge diversification, which the exponential principle fails to do.

Thus, we see that the two complementary theories of choice, expected utility and the dual theory, induce classes of risk measures that have complementary sets of properties, none of which are completely satisfactory. These arguments imply that a desirable set of properties of a risk measure would lie somewhere in between those of coherent risk measures, on the one side, and the exponential principle, on the other. Our discussion thus contributes to the criticism of coherent risk measures initiated by Föllmer & Schied (2002a, 2002b) and Dhaene *et al*. (2003). Generalised utility theory, which combines expected utility and the dual theory, gives rise to a different class of convex measures of risk. Using again an exponential utility function, we obtain a risk measure (which we term the distortion-exponential principle) with properties that we believe are satisfactory. This risk measure behaves approximately as a coherent risk measure for smaller portfolios of risks, while for larger portfolios the risk aversion induced by the utility function becomes prevalent, and the sensitivity to liquidity and risk aggregation issues gradually increases. That this increase is gradual is, in our view, a very positive feature, as the alternative would be to define a cut-off point in portfolio size, whose relative arbitrariness could create anomalies in risk management policy.

We observe that the theory of choice which arguably gives the most complete characterisation of agent preferences, generalised expected utility, also induces risk measures with what we believe to be the best properties. This indicates that the association of risk measures with economic theories has been a productive approach, and constitutes more than a theoretical exercise. We must, however, note that, even though generalised expected utility improves on the von Neumann-Morgenstern utility theory, it is also sometimes violated in practice (Fennema & Wakker, 1996; Georgantzis *et al*., 2002). A question then arises as to whether, in the presence of these violations, our analysis retains its validity. Economic theories of choice strive to give mathematical descriptions of rationality, but what happens when individual decision-makers do not conform with prescribed notions of what a rational act is? We believe that such contradictions do not invalidate the usefulness of the theoretical tools that we used. At the level of policy making, one is actually in the position consciously to decide as to what a rational decision is; the operation of risk measures is, by definition, normative rather than descriptive. Regulation tries to strike a balance between how market players should act and how they are actually likely to act, and we believe that risk measures derived from generalised expected utility get this balance about right. On the other hand, decisions violating economic theories of choice do not necessarily imply irrationality, but are often due to other factors which cannot be easily quantified. It is not the function of a risk measure to negate these factors; an economic decision should not depend exclusively on the value returned by a mathematical model. However, the presence of a risk measure provides decision-makers with one additional conceptual tool, serving a specific purpose. It is then a matter of best practice to choose this tool carefully and justify that choice via economic arguments.

The preceding discussion does not suggest how the risk measures that we propose should be calibrated. Calibration would, of course, depend on the context in which the risk measure will be used. If it is used as a regulatory requirement imposed on the market, the uncertainty aversion of the distortion function could be indicated by a regulator. Recall that the distortion function can be interpreted as a mechanism for producing probabilistic scenarios. This bears some similarity with the Value-at-Risk methodology currently employed in banking regulation. On the other hand, if the risk measure is used internally by a financial entity (bank, insurance company, corporation), the decision would depend on the risk manager's view. Regarding the calibration of the (exponential) utility function, the classical insurer's ruin problem, discussed earlier in the paper, without being necessarily valid in a general risk management context, provides some insight. It was seen that the risk aversion parameter depends on the initial surplus of the insurer, as well as on the maximum acceptable probability of ruin. As our motivation for employing an exponential utility function was related to the issue of how to treat large portfolios, it is reasonable to expect the definition of what amounts to a large portfolio to be dependent on the financial size of its holder. The implication is that the calibration of the risk measure is a company-specific issue. We do not consider this as a weakness of the proposed methodology. Different market players have different characteristics and priorities, and the mathematical tools that they use should be flexible enough to cater for this simple fact. Furthermore, the imposition by a regulator of a single risk measure on the whole market has been criticised in the literature (Danielson *et al*., 2001) as a source of systemic risk, as the use by all market agents of the same risk measure would result in similar risk management strategies. This would, in turn, increase the homogeneity of market decision-makers and increase the probability of a systemic crisis.

Acknowledgment

This paper forms part of the research project 'Coherent Risk Measures: Context and Applications', which is supported by a 2002-2003 Faculty and Institute of Actuaries Research Grant. The support of the Faculty and Institute of Actuaries is gratefully acknowledged by the authors.

The views expressed in this paper are solely those of the authors, and do not necessarily reflect the views of Lloyd's of London.

References

Allais, M. (1953). Le comportement de l'homme rationnel devant le risque: critique des postulates et axioms de l'ecole Americaine. *Econometrica*, **21**, 503-546.

Artzner, P., Delbaen, F., Eber, J-M. & Heath, D. (1999). Coherent measures of risk. *Mathematical Finance*, **9** (3), 203-228.

Bühlmann, H. (1970). *Mathematical methods in risk theory*. Springer, Berlin.

Bühlmann, H. (1985). Premium calculation from top down. *ASTIN Bulletin*, **15** (2), 89-101.

Danielson, J., Embrechts, P., Goodhart, C., Keating, C., Muennich, F., Renault, O. & Shin, H.S. (2001). An academic response to Basel II. Financial Markets Group. Working paper, London School of Economics, http://risk.lse.ac.uk

Denneberg, D. (1990). Distorted probabilities and insurance premiums. *Methods of Operations Research*, **63**, 3-5.

Denneberg, D. (1994). *Non-additive measure and integral*. Kluwer Academic Publishers, Dordrecht.

Dhaene, J., Denuit, M., Goovaerts, M.J., Kaas, R. & Vyncke, D. (2002). The concept of comonotonicity in actuarial science and finance: theory. *Insurance: Mathematics and Economics*, **31** (1), 3-33.

Dhaene, J., Goovaerts, M. & Kaas, R. (2003). Economic capital allocation derived from risk measures. *North American Actuarial Journal*, **7** (2).

Fennema, H.P. & Wakker, P.P. (1996). A test of rank dependent utility in the context of ambiguity. *Journal of Risk and Uncertainty*, **13** (1), 19-36.

Föllmer, H. & Schied, A. (2002a). Convex measures of risk and trading constraints. *Finance and Stochastics*, **6** (4), 429-447.

Föllmer, H. & Schied, A. (2002b). Robust preferences and convex risk measures. *Advances in Finance and Stochastics*, Springer (to appear).

Georgantzis, N., Genius, M., Garca-Gallego, A. & Sabater-Grande, G. (2002). Do higher risk premia induce riskier options? An experimental test. Sixth International Congress on Insurance Mathematics and Economics, 15-17 July 2002.

Gerber, H.U. (1974). On additive premium calculation principles. *ASTIN Bulletin*, **7** (3), 215-222.

Gerber, H.U. & Pafumi, G. (1998). Utility functions: from risk theory to finance. *North American Actuarial Journal*, **2** (3), 74-100.

Goovaerts, M.J., de Vylder, F. & Haezendonck, J. (1984). *Insurance premiums*. Elsevier Science.

Joe, H. (1997). *Multivariate models and dependence concepts*. Chapman & Hall, London.

Luan, C. (2001). Insurance premium calculations with anticipated utility theory. *ASTIN Bulletin*, **31** (1), 23-35.

Machina, M. (1982). 'Expected utility' analysis without the independence axiom. *Econometrica*, **50**, 277-323.

Puppe, C. (1991). *Distorted probabilities and choice under risk*. Springer, Berlin.

Quiggin, J. (1982). A theory of anticipated utility. *Journal of Economic Behavior and Organization*, **3**, 323-343.

Quiggin, J. (1993). *Generalized expected utility theory: the rank-dependent model*. Kluwer Academic Publishers, Boston.

Schmeidler, D. (1989). Subjective probability and expected utility without additivity. *Econometrica*, **57** (3), 571-587.

von Neumann, J.R. & Morgenstern, O. (1947). *Theory of games and economic behaviour*. Princeton University Press.

Wang, S. (1996). Premium calculation by transforming the premium layer density. *ASTIN Bulletin*, **26** (1), 71-92.

Wang, S.S., Young, V.R. & Panjer, H.H. (1997). Axiomatic characterization of insurance prices. *Insurance: Mathematics and Economics*, **21** (2), 173-183.

Yaari, M. (1987). The dual theory of choice under risk. *Econometrica*, **55** (1), 95-115.

https://doi.org/10.1017/S1357321700004414 Published online by Cambridge University Press

APPENDIX A

THEORIES OF CHOICE UNDER RISK: AXIOMATIC CHARACTERISATIONS

For each theory of choice discussed in the paper, the corresponding preference functional can be derived from a set of axioms. In this section we briefly present such axiomatisations, without going into great mathematical detail and the precise conditions under which the representation theorems hold. It is significant that the sets of axioms characterising alternative theories of choice differ with respect to the way in which they deal with aggregation of risks. It is thus no coincidence that the risk measures induced by these theories also have different properties relating to the risk assessment of sums of risks (i.e. superadditivity, subadditivity, convexity). Note that alternative axiomatisations are possible, especially in the case of generalised expected utility. Here we follow the treatment by Puppe (1991). We also find the discussions by Yaari (1987) and Schmeidler (1989) very useful.

Let $\mathcal{X}$ be a set of random variables on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. The decumulative distribution function $S_X(x)$ of $X \in \mathcal{X}$ is defined by $S_X(x) = \mathbb{P}(X > x)$. Denote as $\mathcal{D}$ the set of decumulative distributions of elements in $\mathcal{X}$. Let all elements in $\mathcal{X}$ take values in a compact interval $M \subset \mathbb{R}$, so that M is the support of distributions in $\mathcal{D}$. We also assume that $\mathcal{D}$ is endowed with the topology of weak convergence.

Assume a binary preference operator '$\succeq$' defined on $\mathcal{D}$, with '$\succ$' being its asymmetric part (strict preference) and '$\sim$' its symmetric part (indifference). The relation '$\succeq$' is called complete if $\forall F, G \in \mathcal{D}$, $F \succeq G$ or $G \succeq F$ holds. The relation '$\succeq$' is called transitive if $F \succeq G$ and $G \succeq H$ imply $F \succeq H$, $\forall F, G, H \in \mathcal{D}$. '$\succeq$' is called a weak order if it is complete and transitive.

A real valued functional $V : \mathcal{D} \mapsto \mathbb{R}$ is a representation of '$\succeq$' if for all $F, G \in \mathcal{D}$:

$$F \succeq G \Leftrightarrow V(F) \geq V(G).$$

Consider the following axioms on '$\succeq$':

— *Weak order*. The relation '$\succeq$' is a weak order, i.e. preferences are complete and transitive.

— *Continuity*. $\forall F \in \mathcal{D}$ the sets $\{G \in \mathcal{D} : G \succeq F\}$ and $\{G \in \mathcal{D} : F \succeq G\}$ are closed in the topology of weak convergence.

— *Monotonicity*. $\forall F, G \in \mathcal{D}$, $F \succ G$ whenever F dominates G in first order stochastic dominance, i.e. $F(x) \geq G(x) \forall x \in \mathbb{R}$ (and the inequality is strict for at least one x).

— *Independence*. $\forall F, G, H \in \mathcal{D}$ and $a \in [0, 1]$, $F \succeq G$ implies $aF + (1 - a)H \succeq aG + (1 - a)H$.

Then the expected utility theorem can be stated as follows:

Theorem 1 (Expected Utility). Let '$\succeq$' be a binary operator on $\mathcal{D}$. If and only if '$\succeq$' satisfies the weak order, continuity, monotonicity and independence axioms, there exists a strictly increasing real valued function u such that '$\succeq$' can be represented by the expected utility functional:

$$V(X) = \int_M u(x)d(1 - S_X(x)) = E[u(X)].$$

Moreover, the utility function u is unique up to affine transformations.

Note that the independence axiom is responsible for the linearity in probabilities of the expected utility functional. Furthermore, independence has been found to be frequently violated in practice. Yaari's (1987) dual theory of choice, emerges by modifying the independence axiom. Let $S_X \oplus S_Y$ denote the decumulative distribution function of $X + Y$. Consider now the following axiom:

Dual Independence. $\forall F, G, H \in \mathcal{D}$ and $a \in [0, 1]$, $F \succeq G$ implies $aF \oplus (1 - a)H \succeq aG \oplus (1 - a)H$.

The following result holds:

Theorem 2 (Yaari). Let '$\succeq$' be a binary operator on $\mathcal{D}$. If and only if '$\succeq$' satisfies the weak order, continuity, monotonicity and dual independence axioms, there exists a unique continuous and strictly increasing function $h : [0, 1] \mapsto [0, 1]$ such that '$\succeq$' can be represented by the functional:

$$V(X) = \int_{M\cap\mathbb{R}_-} (h(S_X(x)) - 1)dx + \int_{M\cap\mathbb{R}_+} h(S_X(x))dx.$$

Dual independence is responsible for the linearity in payoffs observed in the dual theory of choice. In order to obtain a generalised utility functional depending on both a utility and a distortion function, the independence axiom has to be weakened and not just modified. There are several ways of doing this and some additional conditions are needed so we will not state a theorem for that case. Sets of alternative axioms are discussed in Puppe (1991), while a very interesting approach is that of Schmeidler (1989). He substituted independence with:

Comonotonic Independence. For all pairwise comonotonic $F, G, H \in \mathcal{D}$ and $a \in (0, 1)$, $F \succeq G$ implies $aF + (1 - a)H \succeq aG + (1 - a)H$.

https://doi.org/10.1017/S1357321700004414 Published online by Cambridge University Press

The resulting preference functional is an expected utility under a non-additive probability measure (Denneberg, 1994), which, under some technical conditions, can be represented as a distorted probability.

https://doi.org/10.1017/S1357321700004414 Published online by Cambridge University Press

APPENDIX B

SUMMARY OF THEORIES OF CHOICE AND RESULTING RISK MEASURES

Economic theory	Expected utility	Dual theory of choice	Generalised expected utility
Non-linear distortion of:	wealth ($u(x)$)	probability ($h(\mathbb{P}) = 1 - g(1-\mathbb{P})$)	wealth *and* probability
Preference functional	$U(X) = \int_{-\infty}^{+\infty} u(x)dF_X(x)$	$H(X) = \int_{-\infty}^{0}(h(S_X(x)) - 1)dx$ $+ \int_0^{\infty} h(S_X(x))dx$	$V_{u,h}(X) = \int_{-\infty}^{0}(h(S_{u(X)(x)}) - 1)dx$ $+ \int_0^{\infty} h(S_{u(X)(x)})dx$
Separability axiom	Independence	Dual independence	Comonotonic independence
Premium principle (using exponential utility)	Exponential $\Pi_{u_{exp}}(X) = \frac{1}{a}\ln E[e^{aX}]$	Distortion $\Pi_h(X) = \int_{-\infty}^{0}(g(S_X(x)) - 1)dx$ $+ \int_0^{\infty} g(S_X(x))dx$	Distortion-exponential $\Pi_{u_{exp},h}(X) = \frac{1}{a}\ln\left\{\int_{-\infty}^{0}(g(S_{e^{aX}}(x)) - 1)dx\right.$ $\left.+ \int_0^{\infty} g(S_{e^{aX}}(x))dx\right\}$
Premium as function of portfolio size	superlinear	linear	linear for small portfolios superlinear for large portfolios
Risk loading	zero for small portfolios and increases with size	positive and constant	positive for small portfolios and increases with size
Sensitivity to aggregation	subadditive for NQD additive for independent superadditive for PQD risks	subadditive for all risks additive for comonotonic risks	convex (subadditive for small risks, superadditive for large PQD risks)