The Thailand Econometrics Society, Vol. 2, No. 2 (January 2010), 8 - 19

How to Relate Spectral Risk Measures and Utilities

Songsak Sriboonchitta[a], Hung T. Nguyen[a,b], and Vladik Kreinovich[a,c]

[a]Faculty of Economics, Chiang Mai University, Chiang Mai, Thailand

[b] Department of Mathematical Sciences, New Mexico State University, Las Cruces, USA

[c] Department of Computer Science, University of Texas at El Paso, El Paso, USA

ARTICLE INFO

Keywords:

JEL classification codes:

ABSTRACT

Traditional decision theory describes human behavior and human preferences in terms of utility functions. In the last decades, it was shown that in many economic situations, a reasonable description of the actual decisions can be found if we use a different approach - of spectral risk measures. In each of these approaches, we first need to empirically find the corresponding function: utility function in the traditional approach and the weighting function for spectral risk measures. Since both approaches provide a reasonable description of the same actual behavior (in particular, of the same actual economic behavior), it is desirable to be able, given utility function, to find an appropriate weighting function (and vice versa). Some empirical rules for such transition have been proposed; these rules are purely heuristic and approximate, they are not theoretically justified. In the present paper, we recall how both the utility and the risk measure approaches can be reformulated in statistical terms, and use these reformulations to provide a statistically justified transition between utility and weighting functions.

1. Formulation of the Problem

Decision theory: main objectives. One of the main objectives of decision theory is to formally *describe* how people make decisions, what is preferable and what is not, so as to be able to *help* decision makers by prompting them decisions which should be beneficial for them.

There exist several approaches to describing such a decision.

Traditional decision theory approach: a brief reminder. Traditional decision theory describes human behavior and human preferences in terms of utility functions (Fishburn (1969,1988); Keeney and Raiffa (1976); Luce and Raiffa (1989); and Raiffa (1970)).

In the utility theory approach, we first select a *utility function* that assigns, to each alternative x, a "utility" value $u(x)$ describing how valuable this outcomes is for the decision maker. For example, in the economic applications, we assign the utility value $u(x)$ to each possible monetary outcome.

The quality of each action - that leads to different outcomes with different probabilities - is characterized by the expected values of the corresponding utility. We therefore select an action which leads to the largest value of expected utility.

Spectral risk measures approach: a brief reminder. In the last decades, it was

* Corresponding author.
E-mail addresses: songsak@econ.cmu.ac.th, hunguyen@nmsu.edu, vladik@utep.edu

shown that in many economic situations, a reasonable description of the actual decisions can be found if we use a different approach - of spectral risk measures (Acerbi (2002,2004); Cotter and Dowd (2006); Sriboonchita, et al. (2009); Dowd, et al. (2008); Overbeck (2002); and Szegö (2002)).

In this approach, we select a "weighting" function $\varphi(p)$ which assigns a weight to all possible probability values $p \in [0,1]$, and we then characterize the quality of an action by the value

$$\int_0^1 \varphi(p) \cdot F^{-1}(p) dp,$$

where $F^{-1}(p)$ denotes a function which is an inverse to the cumulative distribution function $F(x)$ - the probability that the action's outcome will be $\leq x$.

Relation between utility and spectral risk measures: an open problem. We have described two reasonable approaches to describing human decisions: utility theory and spectral risk measures. In each of these approaches we first need to empirically find the corresponding function:

- utility function in the traditional approach and
- the weighting function for spectral risk measures.

Since both approaches provide a reasonable description of the same actual behavior (in particular, of the same actual economic behavior), it is desirable to be able, given utility function, to find an appropriate weighting function and vice versa.

Some empirical rules for such transition have been proposed; see, e.g., (Dowd, et al. (2008)). For example, they suggest to associate:

- to the exponential utility function $u(x)=1\text{-e}^{-k \cdot x}$, the exponential weighting function $\varphi(p)=\frac{k}{1\text{-e}^{-k}} \cdot \exp(-k \cdot (1-p))$;
- to the power utility function $u(x)=x^{1-\gamma}$, the power weighting function $\varphi(p)=\gamma \cdot (1-p)^{\gamma-1}$

However, as the authors of these papers themselves observe, these rules are purely heuristic, approximate, and not theoretically justified. A detailed analysis performed in Dowd, et al. (2008) shows that the proposed match is not perfect: e.g., for power utility functions, the related power weighting function exhibits a bizarre behavior under which decisions are drastically different from the decisions related to the original utility functions.

It is therefore desirable to provide a theoretically justified relation. Such a relation is provided in the present paper.

2. Decision Approaches Reformulated in Statistical Terms

Our idea. To solve the problem of comparing different approaches to decision making, we do the following:

- First, we recall how both the utility and the risk measure approaches can be reformulated in statistical terms.
- Then, we use these reformulations to provide a statistically justified transition between utility and weighting functions.

Let us start by recalling how both approaches can be naturally reformulated in statistical terms.

Utility theory - presented from the statistical viewpoint. The traditional decision theory (see, e.g., Fishburn (1969,1988); Keeney and Raiffa (1976); Luce and Raiffa (1989); and Raiffa (1970)) is based on the notion of *utility*. The traditional utility can be described in simple probabilistic terms. Namely, let us select two alternatives: a very unfavorable alternative A_0 and a very favorable alternative A_1. With this choice, most real-life alternatives lie in between $A0$ and $A1$. A natural scale for such alternatives emerges when we consider, for all possible values p from the interval "lotteries" $A(p)$ in which we get A_1 with probability p and A_0 with the remaining probability $1 - p$.

When $p = 0$, the corresponding lottery $A(0)$ is simply equivalent to the unfavorable outcomes A_0. When p increases, the probability of a favorable outcome increases and thus, the lottery itself becomes more favorable. When the probability p reaches its largest possible value $p = 1$, the corresponding lottery $A(1)$ is equivalent to the very favorable outcome A_1.

Let A be an arbitrary alternative between A_0 and A_1, i.e., an alternative which is better than A_0 ($A_0 < A$) and worse than A_1 ($A < A_1$). When p goes from 0 to 1, the lottery $A(p)$ continuously changes from the very unfavorable alternative $A0$ to the very favorable alternative A_1. Thus, it is reasonable to expect that there exists a value p for which the alternative A is equivalent (of the same quality) as the lottery $A(p)$. This probability p is called the *utility* of the alternative A.

Expected utility. As we have mentioned, one of the main objectives of the utility theory is to help a user select the best action. It is rarely possible to predict the exact results of each action. At best, we can predict the *probabilities* of different consequences of each action.

Suppose that we have an action with possible consequences $C_1,\ldots,C_n$, we know the utility $u_i = u(C_i)$ of each of these consequences, and we know the probabilities $p_1,\ldots,p_n$ of these consequences, $p_1 + \ldots + p_n = 1$. How can we then describe the benefit of this action?

The action means that we get each C_i with probability p_i. By definition of utility, each alternative C_i is equivalent to a lottery in which we get A_1 with probability u_i. Thus, the action is equivalent to a "compound" lottery in which, with probability u_i, we select a new lottery in which the very favorable outcome $A1$ occurs with probability p_i. The total probability of A_1 in such a compound lottery can be then determined by the formula of complete probability: it is equal to

$$\bar{u} \stackrel{\text{def}}{=} p_1 \cdot u_1 + \ldots + p_n \cdot u_n. \quad (1)$$

Thus, the original action is equivalent to the lottery in which we get A_1 with probability u (and A_0 with the remaining probability $1 - \bar{u}$). By definition of utility, this means that the utility of the action is equal to the expression (1).

From the statistical viewpoint, the expression (1) is simply the expected value of the utility u. Thus, the utility of the action is equal to the expected value of the utilities of its consequences.

Re-scaling utility. The above definition of utility depends on the selection of two alternatives A_0 and A_1. What will happen if we select two different alternatives, e.g., alternatives $\tilde{A}_0$ and $\tilde{A}_1$ for which $\tilde{A}_0 < A_0$ and $A_1 < \tilde{A}_1$? How is the utility $\hat{u}(A)$ based on the new selection related to the utility $u(A)$ based on the original selection? In this case, since both A_0 and A_1 are in between $\tilde{A}_0$ and $\tilde{A}_1$, for some probabilities $\tilde{p}_0$ and $\tilde{p}_1$,

- the alternative A_0 is equivalent to a lottery $\tilde{A}(\tilde{p}_0)$ in which we get $\tilde{A}_1$ with probability $\tilde{p}_0$ and $\tilde{A}_0$ with the remaining probability $1 - \tilde{p}_0$, and
- *the alternative A_1 is equivalent to a lottery $\tilde{A}(\tilde{p}_1)$ in which we get $\tilde{A}_1$ with* probability $\tilde{p}_1$ and $\tilde{A}_1$ with the remaining probability $1 - \tilde{p}_1$.

Each alternative A is equivalent to a lottery $A(u(A))$ in which we get A_1 with probability $u(A)$ and A_0 with the remaining probability 1 - $u(A)$. Replacing each of the alternatives A_0 and A_1 with the corresponding lottery $\tilde{A}(\tilde{p}_0)$ or $\tilde{A}(\tilde{p}_1)$, we thus get a new composite lottery in which:

- with probability $u(A)$, we launch a lottery in which we get $\tilde{A}_1$ with probability $\tilde{p}_1$, and
- with probability 1 - $u(A)$, we launch a lottery in which we get $\tilde{A}_1$ with probability $\tilde{p}_0$.

The total probability of getting $\tilde{A}_1$ in this compound lottery is equal to

$$\tilde{u} \stackrel{\text{def}}{=} \tilde{p}_1 \cdot u(A) + \tilde{p}_0 \cdot (1 - u(A)). \qquad (2)$$

Thus, the alternative A is equivalent to a lottery $\tilde{A}(\tilde{u})$ in which we get the new favorable alternative $\tilde{A}_1$ with probability $\tilde{u}$ and the new unfavorable alternative with probability 1-$\tilde{u}$. By definition of utility, this means that in the new scale, the utility $\tilde{u}(A)$ of the alternative A is equal to $\tilde{u}$. Formula (2) can be rewritten as a linear transformation:

$$\tilde{u}(A) = a \cdot u(A) + b, ;$$

where $a \stackrel{\text{def}}{=} \tilde{p}_1 - \tilde{p}_0$ and $b = \tilde{p}_0$. Thus, in general, the change in a scale corresponds to a linear re-scaling of utility.

In other words, the numerical values of utility are determined modulo an arbitrary linear transformation.

Spectral risk measures - presented from the statistical viewpoint. Spectral risk theory provides an alternative description of human preferences. This description is based on the idea of risk aversion; see, e.g., Acerbi (2002,2004); Cotter and Dowd (2006); Sriboonchita, et al. (2009); Dowd, et al. (2008); Overbeck (2002); and Szegö (2002).

Let us start with an extreme idealized case. For example, what does it mean that a person is fully intolerable to risk? Intuitively, this means that if you propose this person some favorable alternative with a certain probability p, this person would never prefer it. In other words, to this person, the quality of an action is determined by what we can *guarantee*, i.e., by the worst possible consequence - because more favorable alternatives come with risk and thus, do not count.

Of course, in reality, such an idealized behavior does not occur. Every person has a certain tolerance for risk, i.e., a probability p of failure which this person can still tolerate. In this case, we can dismiss the worst alternatives as long as their total probability does not exceed p. In mathematical terms, this means that as a numerical criterion of an action, we take the value $F^{-1}(p)$ for which the probability of benefits being smaller than $F^{-1}(p)$ is equal to p. This value - inverse to the cumulative distribution function $F(x)$ - is called the *p-th quantile* of the corresponding probability distribution.

The quantiles describe decisions of individual person. However, important decisions are rarely made by individuals taking only their preferences into account. Most important decisions take into account preferences of several persons. Each of these persons may have their own risk tolerance value p. For each of them, the benefit of each action is proportional to the corresponding quantile - i.e., in simplified terms, each of these persons is willing to buy his or her participation of this action for the amount $F^{-1}(p)$. If we denote by $\varphi(p)$ the proportion of persons with risk tolerance p, then the total amount that all the participants are willing to pay to participate in this action can be described as the average value

$$\int \varphi(p) \cdot F^{-1}(p) dp \quad (3)$$

This expression (3) is called a *spectral risk measure*, and the corresponding function $\varphi(p)$ is called a *weighting function*.

3. Towards Comparing the Two Approaches: Let Us Reformulate Both Approaches for the Practical Case of a Sample

From the general idea (arbitrary distribution) to a practical implementation (sample). As a result of each action, we have different monetary amounts with different probabilities.

Both the utility and the spectral risk measure approaches allow arbitrary probability distributions. In practice, we usually do not know corresponding probability distribution, we usually only have a *sample* $x_1,...,x_n$ of the corresponding monetary amounts.

It is natural to build a histogram based on these values, i.e., equivalently, to build an "empirical" distribution in which we have each of the *n* values with equal probability 1/*n*. It is well known that when the sample size increases, this empirical distribution converges to the actual one.

How will both approaches look like for this empirical distribution?

Utility approach on the example of a sample. For a utility function $u(x)$, the utility of each alternative x_i is equal to $u(x_i)$, and the probability of each alternative is equal to 1/*n*. Thus, the expected value of the utility is equal to

$$\bar{u} = \frac{1}{n} \cdot u(x_1) + ... + \frac{1}{n} \cdot u(x_n).$$

Utility approach reformulated in terms of an equivalent monetary value. It is difficult to directly compare the utility value with the value provided by the spectral risk measures. Indeed:

- the utility approach provides an equivalent *utility* value, while
- the risk measures approach provides an equivalent monetary value.

To make this comparison possible, let us reformulate the utility approach in such a way that it will also lead to a monetary value.

In other words, instead of describing value of an action to a person as the utility value, we want to describe the value of an action as the amount of money x that this person is willing to pay to participate in this action.

Once a person paid the amount of money x, in each alternative *i*, the person gains the value $x_i - x$. The expected utility is this equal to

$$\frac{1}{n} \cdot u(x_1 - x) + ... + \frac{1}{n} \cdot u(x_n - x).$$

Under the appropriate value x, this expected utility is equal to the utility of gaining nothing, i.e., to *u*(0):

$$\frac{1}{n} \cdot u(x_1 - x) + ... + \frac{1}{n} \cdot u(x_n - x).$$

We have mentioned that a utility function is defined modulo an arbitrary linear transformation. Thus, we can always "normalize" the utility function to get *u*(0) = 0. After this normalization, the above formula takes a simplified form

$$\frac{1}{n} \cdot u(x_1 - x) + ... + \frac{1}{n} \cdot u(x_n - x) = 0,$$

i.e., multiplying both sides by *n*, the form

$$u(x_1 - x) + ... + u(x_n - x) = 0 \quad (4)$$

Spectral risk measure on the example of a sample. For a sample distribution, once we order the sample values $x_1,...,x_n$ into an increasing sequence

$$x_{(1)} \le x_{(2)} \le ... \le x_{(n)},$$

the value $x_{(1)}$ is the (1/*n*)-th quantile, the value $x_{(2)}$ is the (2/*n*)-th quantile, . . . , $x_{(i)}$ is the (*i/n*)-th quantile, etc.

Thus, the formula (3) becomes proportional to

$$x = \frac{1}{n}\cdot\varphi\left(\frac{1}{n}\right)\cdot x_{(1)} + \frac{1}{n}\cdot\varphi\left(\frac{2}{n}\right)\cdot x_{(2)} + ... + \frac{1}{n}\cdot\varphi\left(\frac{n-1}{n}\right)\cdot x_{(n-1)} + \frac{1}{n}\cdot\varphi(1)\cdot x_{(n)}$$

$$= \frac{1}{n}\cdot\sum_{i=1}^{n}\varphi\left(\frac{i}{n}\right)\cdot x_{(i)}. \quad (5)$$

Resulting reformulation of the problem. In these sample terms, the original problem about the relation between the utility function and the weighting function takes the following form:

- given a function $u(x)$, find the function $\varphi(p)$ for which the estimates (5) are close to estimates obtained from the equation (4), and
- given a function $\varphi(p)$, find the function $u(x)$ for which the estimates obtained from the equation (4) are close to estimates (5).

4. A Similar Problem Is Already Solved In Robust Statistics

Robust statistics: reminder. In this section, we will recall that a similar mathematical problem is already solved in *robust statistics* - an area of statistics in which we need to make statistical estimates under partial information about the probability distribution.

In robust statistics (see, e.g., Huber (2004)), there are several different types of techniques for estimating a shift-type parameter *a* based on a sample $x_1,...,x_n$.

M-methods: reminder. The most widely used methods are *M-methods*, methods which are similar to the maximum likelihood approach from the traditional (non-robust) statistics. In the maximum likelihood approach, if we know that the probability density function has the form $f_0(x_i - a)$ for some unknown value *a*, and that the values $x_1,...,x_n$ are independent, then the likelihood to get the sample $x_1,...,x_n$ is equal to the product

$$\prod_{i=1}^{n} f_0(x_i - a).$$

In the Maximum Likelihood approach, we select the value $a = a_M$ for which this likelihood is the largest possible:

$$\prod_{i=1}^{n} f_0(x_i - a) \to \max_a .$$

It is well known that for standard distributions like normal, the problem becomes computationally easier if we replace the original problem of maximizing the product with the equivalent problem of maximizing the logarithm of this product:

$$\ln\left(\prod_{i=1}^{n} f_0(x_i - a)\right) \to \max_a,$$

and take into account that the logarithm of the product is equal to the sum of the logarithms:

$$\sum_{i=1}^{n} \ln(f_0(x_i - a)) \to \max_a .$$

To find this maximum, we can differentiate the objective function by *a* and equate the resulting derivative to 0. For each i, due to the chain rule, the derivative of the logarithm has the form

$$\frac{d}{da}(\ln(f_0(x_i - a))) = -\frac{f_0'(x_i - a)}{f_0(x_i - a)},$$

where $f_0'(x)$ denotes the derivative of $f_0(x)$. In other words, we get the following equation for determining the Maximum Likelihood estimate a_M:

$$U(x_1 - a_M) + \ldots + U(x_n - a_M) = 0, \quad (6)$$

where we denoted

$$U(x) = -(\ln(f_0))' = -\frac{f_0'(x)}{f_0(x)}. \quad (7)$$

Comment. This formula is, in effect, identical to our formula (4).

M-methods: robust case. In the Maximum Likelihood approach, we know the probability density function $f_0(x)$. In the robust approach, we apply a similar method with some function $U(x)$.

Each of these robust M-methods coincides with the Maximum Likelihood method for an appropriate probability density function. Once we know the function $U(x)$, we can find this probability density function as follow. First, we can find $-\ln(f_0(x))$ as the integral of $U(x)$:

$$-\ln(f_0(x)) = \int_c^x U(t)dt$$

for an appropriate lower bound c, hence

$$f_0(x) = \exp\left(-\int_c^x U(t)dt\right)$$

L-estimates. Another important class of robust estimates are *L-estimates*, i.e., estimates of the type

$$a_L = \frac{1}{n} \cdot \sum_{i=1}^n m\left(\frac{i}{n}\right) \cdot x_{(i)}, \quad (8)$$

for some function $m(x)$ for which $\int_0^1 m(t)dt = 1$..

Comment. This formula is, in effect, identical to our formula (5).

A problem which is solved in robust statistics. The question solved in robust statistics is: what is the natural correspondence between M-estimates and L-estimates?

Correspondence between M- and L-estimates: case of traditional statistics. To explain the meaning of this correspondence, let us first consider the case when we know the exact shape $f_0(x)$ of the probability density function, and we know that the actual probability density function has the form $f_0(x-a)$ for some (unknown) parameter a. In this case,

- for each function $U(x)$, we can use the solution of the corresponding equation (6) as an M-method estimate $a_M(U)$ for the parameter a;
- for each function $m(p)$, we can use the estimate (8) as an L-method estimate $a_L(m)$ for the parameter a.

The quality of each estimate can be estimated as the mean square of the difference between the estimate and the actual value a, i.e.,

- for M-estimates, as $q_M(U) = E[(a_M(U) - a)^2]$; and
- for L-estimates, as $q_L(m) = E[(a_L(m) - a)^2]$.

For a given probability density function $f_0(x)$:

- we can find the optimal function $U(x)$, i.e., the function $U(x)$ for which the value $q_M(U) = E[(a_M(U) - a)^2]$ is the smallest possible, and

- we can find the optimal function $m(p)$, i.e., the function $m(p)$ for which the value $q_L(m) = E[(a_L(p) - a)^2]$ is the smallest possible.

Specifically, when we know the exact shape $f_0(x)$ of the probability distribution functions, then the optimal M-estimate has the form (7), i.e., $U(x) = -(\ln(f_0))'$..

The optimal L-estimate can be found as follows (see, e.g., Collins and Wiens (1989); and Huber(2004)):

- first, we compute the cumulative distribution function $F_0(x)$ as
$$F_0(x)=\int_{-\infty}^{x} f_0(t)dt;$$
- then, we find the auxiliary function $M(p)$ as
$$M(F_0(x))= -(\ln(f_0(x))'';$$
- after that, we normalize the auxiliary function $M(p)$ to get
$$m(p) = \frac{M(p)}{\int_0^1 M(q)dq}..:$$

These formulas can be further simplified. For example, since $-(\ln(f_0))' = U(x)$, we have $-(\ln(f_0(x)))'' = U'(x)$. So, the formula for $M(F_0(x))$ can be rewritten as $M(F_0(x)) = U'(x)$.

The correspondence between the functions $U(x)$ and $m(p)$ can now be described as follows.

Let us first assume that we know the function $U(x)$, then, to find the corresponding function $m(p)$, we do the following:

- first, we find a probability density function $f_0(x)$ for which $U(x)$ leads to the optimal M-estimate;
- then, we use this probability density function $f_0(x)$ to find the function $m(p)$ which leads to the optimal L-estimate for this $f_0(x)$.

Similarly, if we know the function $m(p)$, then, to find the corresponding function $U(x)$, we do the following:

- first, we find a probability density function $f_0(x)$ for which $m(p)$ leads to the optimal L-estimate;
- then, we use this probability density function $f_0(x)$ to find the function $U(x)$ which leads to the optimal M-estimate for this $f_0(x)$.

Correspondence between M- and L-estimates: explicit description. Once we know $U(x)$, we can find the corresponding function $m(p)$ as follows:

- first, we compute the function $f_0(x) = \exp\left(-\int_c^x U(t)dt\right)$;
- then, we compute $F_0(x)=\int_{-\infty}^{x} f_0(t)dt$;
- after that, we find the function $M(p)$ from the formula $M(F_0(x))=U'(x)$, i.e., as $M(p)=U'(F_0^{-1}(p))$, where $F_0^{-1}(p)$ denotes an inverse function(i.e., a function for which $F_0^{-1}(p) = x$ if and only if $f_0(x)=p$);
- finally, we compute $I \stackrel{\text{def}}{=} \int_0^1 M(q)dq$, and take $m(p) = \frac{M(p)}{I}$

Comment. It turns out that for the resulting functions $U(x)$ and $m(p)$, the quality values $q_M(U) = E[(a_M(U) - a)^2]$ and $q_L(m) = E[(a_L(p) - a)^2]$ are asymptotically equal when the sample size n tends to infinity:

$$\frac{q_M(U)}{q_L(m)} = \frac{E[(a_M(U)-a)^2]}{E[(a_L(p)-a)^2]} \to 1 \text{ as } n \to +\infty$$

Correspondence between M- and L-estimates: robust case. In the robust case, when we do not know the exact shape of a probability density function, we only know the *class* F_0 of possible shapes, and we know that the actual probability density function has the form $f_0(x\text{-}a)$, where $f_0(x)$ is one of the shapes from the class F_0, and a is an (unknown) parameter. In this case too, we can consider M-estimates $a_M(U)$ (described by the formula (6)) and L-estimates $a_L(m)$ (described by the formula (8)).

In the robust case, since the distribution is not known exactly, for different distributions $f_0(x)$ from the class F_0, we get different accuracies

$E_{f_0}[(a_M(U)-a)^2]$ and $E_{f_0}[(a_L(m)-a)^2]$.

As a natural measure of quality of a given estimate, we can take the *worst-case* accuracy

$$q_M(U) = \sup_{f_0 \in F} E_{f_0}[(a_M(U)-a)^2];$$

$$q_L(m) = \sup_{f_0 \in F} E_{f_0}[(a_L(m)-a)^2].$$

As shown in Huber (2004), for many reasonable classes F_0 of distributions,

- we can find the optimal (*minimax*) function $U(x)$, i.e., the function $U(x)$ for which the value $q_M(U)$ is the smallest possible, and
- we can find the optimal (*minimax*) function $m(p)$, i.e., the function $m(p)$ for which the value $q_L(m)$ is the smallest possible.

These optimal M-estimates and L-estimates can be obtained as follows Collins and Wiens (1989); and Huber (2004):

- first, in the class F_0, we find the probability distribution $f_0(x)$ for which the Fisher information
$$I(f_0) = \int \left(\frac{f_0'(x)}{f_0(x)} \right)^2 \cdot f_0(x)dx$$
is the smallest possible;
- then, we find M-estimate and L-estimate which are optimal for this distribution $f_0(x)$.

The correspondence between the functions $U(x)$ and $m(p)$ can then be described as follows.

Let us first assume that we know the function $U(x)$, then, to find the corresponding function $m(p)$, we do the following:

- first, we find a class F_0 of probability density functions for which $U(x)$ leads to the optimal M-estimate;
- then, we use this class F_0 to find the function $m(p)$ which leads to the optimal L-estimate for this class F_0.

Similarly, if we know the function $m(p)$, then, to find the corresponding function $U(x)$, we do the following:

- first, we find a class F_0 of probability density functions for which $m(p)$ leads to the optimal L-estimate;
- then, we use this class F_0 to find the function $U(x)$ which leads to the optimal M-estimate for this class F_0.

It turns out that for the resulting functions $U(x)$ and $m(p)$, the quality values $q_M(U)$ and $q_L(m)$ are also asymptotically equal when the sample size n tends to infinity:

$$\frac{q_M(U)}{q_L(m)} \to 1 \text{ as } n \to +\infty .$$

Correspondence between M- and L-estimates: explicit description. We have mentioned that the robust M- and L-estimates coincide with M- and L-estimates for an appropriate probability density function $f_0(x)$. Thus, the robust-case correspondence between M- and L-estimates can be described by exactly the same formulas as for the traditional statistical case.

Examples. Several examples are given in Collins and Wiens (1989); and Huber (2004).

For example, when $U(x)=x$, this procedure leads to $m(t) = 1$, i.e., to an average of all possible values $x(i)$. Indeed, in this case,

$$\int_c^x U(t)dt = \frac{1}{2}\cdot x^2,$$

so $f_0(x)=\exp\left(-\frac{1}{2}\cdot x^2\right)$ is proportional to the probability density of the normal distribution. Hence, $F_0(x)=\int_{-\infty}^x f_0(t)dt;$ is the cumulative distribution function of a normal distribution. Here, $U'(x)=x$, so $M(p)=U'(F_0^{-1}(p))=1$. The integral of $M(p) = 1$ over the interval [0, 1] is 1, so $m(p) = M(p) = 1$.

Another example: when $U(x) = \max[-c_0, \min(c_0, x)]$, i.e., when

- $U(x) = -c_0$ for all $x \le -c_0$,
- $U(x) = x$ for all $x \in [-c_0, c_0]$, and
- $U(x) = c_0$ for all $x \ge c_0$

then, for an appropriate value α_0, we have $m(p) = \frac{1}{1-2\alpha_0}$ for all p from the interval $[\alpha_0, 1-\alpha_0]$.

5. Relation Between Utility and Spectral Risk Measures: Our Main Idea

Let us apply the solution from robust statistics to the economic situation. We have seen that, mathematically,

- M-estimates correspond to utility estimates, and
- L-estimates correspond to spectral risk estimates.

We can therefore use the solution provided by robust statistics to find the desired correspondence between the utility function and the spectral risk measures.

Resulting solution. Specifically, once we know the utility function $u(x)$ for which $u(0) = 0$, we can find the corresponding weighting function $\varphi(p)$ as follows:

- first, we compute an auxiliary function $f_0(x) = \exp\left(-\int_c^x u(t)dt\right)$;
- then, we compute the second auxiliary function $F_0(x)=\int_{-\infty}^x f_0(t)dt;$;
- after that, we find the third auxiliary function $M(p)$ from the formula $M(F_0(x)) = u'(x)$; i.e., as $M(p) = u'(F_0^{-1}(p))$, where $F_0^{-1}(p)$ denotes an inverse function;
- finally, we compute $I \stackrel{\text{def}}{=} \int_0^1 M(q)dq$ and take $\varphi(p) = \frac{M(p)}{I}$.

Comment. The above procedure describes how, knowing the utility function $U(x)$, we can find the corresponding weighting function $\varphi(p)$. What if we know the weighting function $\varphi(p)$ and we want to find the corresponding utility function $U(x)$? To find $U(x)$, we can use the above formula $M(F_0(x)) = -(\ln(f_0(x)))''$, where $f_0(x) = F_0''(x)$, and $M(p) = I \cdot m(p)$ for $I = \int_0^1 M(q)dq$.

Thus, given $U(x)$, we can find *m*(*p*) as follows:

- first, we find the auxiliary function *F*0(*x*) and the auxiliary value *I* by solving the equation $I \cdot m(F_0(x)) = -(\ln(F_0'(x)))''$;
- then, we find $f_0(x) = F_0'(x)$ and $U(x) = -\frac{f_0'(x)}{f_0(x)}$.

Economic interpretation. The above examples from robust statistics, when interpreted in economic terms, show the following:

- if the utility is simply proportional to the monetary value, i.e., if the decision maker is completely risk-neutral, then in the corresponding all possible values *p* are equally probable;
- if the utility is bounded by some value *c*, i.e., if very strong gains and very severe losses are ignored by the decision maker, then very small ($p < \alpha_0$) and very high ($p > 1-\alpha_0$) values of α can also be ignored - because they only affect a decision when combined with very large gains or losses.

These examples show that - at least in the simplest cases when the above procedure leads to an explicit formula - the above mathematical procedure makes economic sense.

Comment. For the exponential utility function $u(x) = 1 - e^{-k \cdot x}$ and for the power utility function $u(x) = x^{1-\gamma}$, the above algorithm *does not* lead to the simple weighting functions proposed in Dowd, et al. (2008).

For example, for $u(x) = x^{1-\gamma}$, we get

$$f_0(x) = \exp(-\int_c^x t^{1-\gamma} dt) = A \cdot \exp(-\text{const} \cdot x^{2-\gamma})$$

Thus,

$$F_0(x) = \int^x f_0(x) = A \cdot \int^x \exp(-\text{const} \cdot x^{2-\gamma}),$$

and the equation for *M*(*p*) takes the form

$$M\left(A \cdot \int^x \exp(-\text{const} \cdot x^{2-\gamma})\right) = (1-\gamma) \cdot x^{-\gamma}$$

Similar, for the exponential utility function, we get complex implicit expressions for the weighting functions - expressions which, because of their complexity, are not easy to analyze. We hope that that these complex expressions will lead to a more reasonable economic behavior, behavior which is closer to the behavior corresponding to the original utility functions.

Acknowledgments. The authors are greatly thankful to the Faculty of Economics of Chiang Mai University for the financial support.
This work was also supported in part by the National Science Foundation grants HRD-0734825 and DUE-0926721, by Grant 1 T36 GM078000-01 from the National Institutes of Health, and by Grant 5015 from the Science and Technology Centre in Ukraine (STCU), funded by European Union.

References

Acerbi, C., 2002, Spectral measures of risk: a coherent representation of subjective risk aversion. Journal of Banking and Finance 26, 1505-1518.

Acerbi, C., 2004, Coherent representations of subjective risk aversion. In: G.

Szegö (ed.), Risk Measures for the 21st Century, Wiley, New York, 147-207.
Collins, J., and Wiens, D., 1989, Mimimax properties of M-, R-, and L-estimators of location in Lévy neighborhoods. The Annals of Statistics 17(1), 327{336.
Cotter, J., and Dowd, K., 2006, Extreme spectral risk measures: an application to futures clearinghouse margin requirements. Journal of Banking and Finance 30, 3469-3485.
Dowd, K., Cotter, J., Sorwar, G., 2008, Spectral risk measures: properties and limitations. Journal of Financial Services Research 34, 61-75.
Fishburn, P.C., 1969, Utility Theory for Decision Making. John Wiley & Sons Inc., New York.
Fishburn, P.C., 1988, Nonlinear Preference and Utility Theory. The John Hopkins Press, Baltimore, MD.
Huber, P. J., 2004, Robust Statistics, Wiley, Hoboken, New Jersey.
Keeney, R. L., and Raiffa, H., 1976, Decisions with Multiple Objectives. John Wiley and Sons, New York.
Luce, R. D., and Raiffa, H., 1989, Games and Decisions: Introduction and Critical Survey. Dover, New York.
Overbeck, L., 2002, Spectral capital allocation. In: A. Das (ed.), Capital Allocation, Risk Books, London.
Raiffa, H., 1970, Decision Analysis. Addison-Wesley, Reading, Massachusetts.
Sriboonchita, S., Wong, W. K., Dhompongsa, S., and Nguyen, H. T., 2009, Stochastic Dominance and Applications to Finance, Risk and Economics. Chapman & Hall/CRC, Boca Raton, Florida.
Szegö, G., 2002, Measures of risk. Journal of Banking and Finance 26, 1253-1272.