Economic Decisions under Uncertainty

by Hans-Werner Sinn

North Holland: Amsterdam, New York and Oxford 1983

Chapter 2: Rational Behavior under Risk

Chapter Two

Rational Behavior under Risk

In chapter one, general decision problems under uncertainty were reduced to the case of pure risk. The next question is how the economic decision maker will evaluate objective risks, that is, what are the properties of the preference functional $R(.)$ in the case of objective probabilities. This chapter attempts to give a partial answer which determines some basic rules for rational behavior under risk. The following chapter is devoted to the task of formulating a supplementary hypothesis concerning man's preferences.

In order to give the problem more structure than was necessary in chapter one, it is assumed that the result vector of an economic action is represented by an equivalent objective probability distribution of end-of-period wealth[1], V. The decision rule under risk or uncertainty, in a shortened form, is then

$$\max R(V),$$

or in words: choose that action out of the set of possible alternatives which brings about a probability distribution of end-of-period wealth V that maximizes the value of the preference functional $R(.)$. Of course the limitation to wealth distributions excludes a number of problems such as, for example, decisions of life and death, but for typical economic problems under uncertainty, such as portfolio management, insurance demand, and speculation, the limitation is normally of no consequence.

In line with the formulation of the decision problem given in chapter one, we assume in principle that V is a discrete random variable which takes on alternative variates v with known probabilities $W(v)$. For

[1] A more precise definition of the concept of wealth utilized in this book is given at the beginning of chapter three.

analytical purposes, however, it is normally more convenient to utilize continuous distributions, which may be interpreted as approximations of underlying discrete distributions[2]. It is thus assumed that V may also be a continuous random variable which takes on a particular variate v with known density $f(v)$. The random variable V will be called 'probability distribution' without, however, implying anything in advance about the kind of distribution it is[3].

Instead of end-of-period distributions of wealth it is equally possible to consider the period income distributions. Let a denote the decision maker's non-random initial wealth and assume there is no consumption (or, in the case of a firm, no dividends). Then

$$Y = V - a.$$

Thus a wealth distribution and its corresponding income distribution can be constructed from each other by a simple shift of size a. The type of distribution chosen in modelling choice under uncertainty is a matter of taste. However, because of a particular wealth dependence of risk evaluation, it will become clear in the next chapter that, in general, it is better to refer to the distribution of end-of-period wealth. Nevertheless, for the presentation of some of the preference functionals that have been proposed in the literature, we prefer to refer to the distribution of period income. The decision problem will therefore be formulated as $\max R(Y)$ where this seems appropriate.

It is certainly implausible to assume that there are no withdrawls from wealth for consumption. A realistic assumption would be that, at the beginning of the period, the decision maker simultaneously chooses the optimal risk project and his consumption over the period. For the time being this problem is neglected. In chapter IV the consumption decision will be taken into account in a full intertemporal approach and it will be

---

[2] A possible procedure for this approximation is as follows. First, the set of real numbers is divided into classes of size Δ. Then the probabilities of all wealth levels falling into such a class are added and the sum is associated with this class. Finally, a function $f(v)$ is chosen such that for a class extending symmetrically around $\bar{v}$ and for the probability $W(\bar{v})$ the equation

$$W(\bar{v}) = \int_{\bar{v}-\Delta/2}^{\bar{v}+\Delta/2} f(v)dv$$

will hold, which implies that

$$f(v) = \frac{W(\bar{v})}{\Delta}.$$

[3] In contrast to our definition, the integral $\int_0^v f(u)du$ is often called 'probability distribution of the random variable V'.

shown that our abstraction is not so severe as it might appear at the moment.

Various proposals for specifying the preference functional $R(V)$ have been made in the literature. There are three types of decision criteria under risk which seem fairly incompatible.

- The *two-parametric substitutive criteria*. From the probability distribution two characteristic numbers are generated to indicate 'risk' and 'return'. The numbers are then evaluated by means of a substitutive preference function.
- The *lexicographic criterion*. A preference function is formulated to evaluate the probabilities of wealth exceeding some critical levels.
- The *expected-utility criterion*. By means of a given utility function the end-of-period wealth distribution is transformed into a distribution of utilities whose mathematical expectation serves as the preference functional.

Table 1 gives an overview of the decision criteria that have been proposed in the literature. The meanings of the symbols used are set out below Table 1 and, as well, are explained by means of an example of a probability distribution illustrated in Figure 1. Where appropriate, the table was constructed by reference to end-of-period wealth V, even for cases where the preference functional was originally designed for income Y. For criteria b) and e), however, the reference to end-of-period wealth was not appropriate since negative and positive changes in wealth must be distinguished.

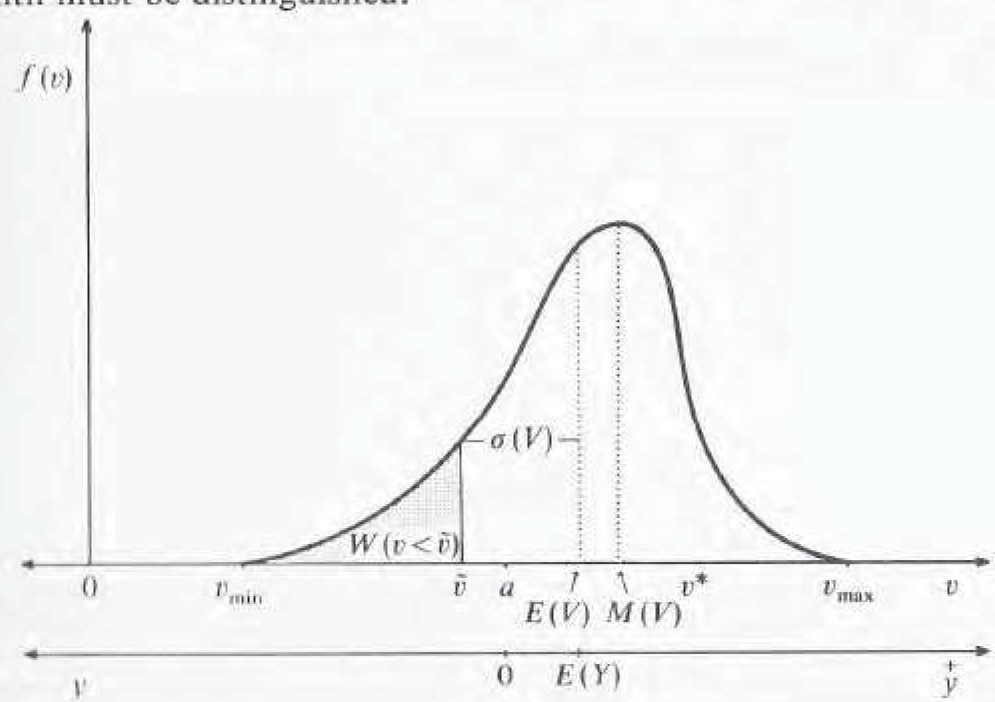


Figure 1

Table 1
Preference Functionals

two-parametric substitutive criteria	a) $U[M(V), v_{\max} - v_{\min}]$	LANGE (1943)
	b) $U[E(Y), -\int_{-\infty}^{0} yf(y+a)dy]$	DOMAR and MUSGRAVE (1944)
	c) $U[\mu, \sigma]$	FISHER (1906, pp. 406-410) HICKS (1933), MARSCHAK (1938), STEINDL (1941), TINTNER (1941), LUTZ (1951, pp. 179-192), MARKOWITZ (1952a), TOBIN (1958)
	d) $U[\mu, \int_{-\infty}^{v^*} (v-v^*)^2 f(v)dv]$	MARKOWITZ (1970, pp. 188-201)
	e) $U(\bar{y}^*, \overset{+}{y}^*)$	SHACKLE (1952, pp. 9-31), KRELLE (1957), SCHNEIDER (1964, pp. 89-133)
lexicographic criterion	f) $U[W(v \geq \bar{v}), \ldots]$	H. CRAMÉR (1930, pp. 10 and 38), R. ROY (1952), ENCARNACIÓN (1965), HAUSSMANN (1968/69), NACHTKAMP (1969, pp. 117-123, 145)
expected-utility criterion	g) $E[U(V)]$	G. CRAMER (1728), D. BERNOULLI (1738), VON NEUMANN and MORGENSTERN (1947, pp. 17-29, 617-632)

V	random variable 'end-of-period wealth'	$\bar{Y}, \bar{y}$	loss (absolute value of strictly negative values of Y or y respectively)
v	variate of V	$\overset{+}{Y}, \overset{+}{y}$	gain (positive values of Y or y respectively including a 'gain' of zero)
a	wealth at the beginning of period	$\bar{y}^*$	focus loss, equivalent loss
$Y = V - a$	random variable 'period income', 'change in wealth'	$\overset{+}{y}^*$	focus gain, equivalent gain
y	variate of Y	$M(V)$	mode (most dense value) of V
$v_{\max}, v_{\min}$	upper, lower boundary of wealth distribution		
$\bar{v}$	level of disaster		
v^*	critical wealth level		

$W(v > \bar{v}) = \int_{\bar{v}}^{+\infty} f(v)dv$ probability of survival

$\mu \equiv E(V) = \int_{-\infty}^{+\infty} vf(v)dv$ mathematical expectation of V ($E(Y)$ is analogously defined)

$\sigma^2(V) = \int_{-\infty}^{+\infty} [v - E(V)]^2 f(v)dv$ variance of V

$\sigma \equiv \sigma(V) = \sqrt{\sigma^2(V)}$ standard deviation of V

The criteria mentioned above will be discussed in the following sections[4]. We shall not consider those preference functionals which were constructed for the evaluation of unknown probability distributions[5]. They are ruled out from the beginning because, as shown in chapter one, it is always possible to find equivalent objective probabilities.

The so-called *expected-value* or *mean-value criterion* $R(V) = E(V) = \mu$, too, is out of the running. In the version $E(Y)$ this criterion is the classical preference functional for the evaluation of games of chance and its popularity is due to the fact that, when a game is continuously repeated, the average gain converges stochastically towards the expected gain[6]. However, since multiple risks are excluded for the time being, this argument does not count. Of course, even for unique choice situations, the mean-value criterion has a certain degree of plausibility since it chooses the center of gravity of a probability distribution to be the preference functional. Similarly plausible are other parameters of position such as the mode or the median. The usefulness of such simple position parameters must be doubted, however, since they imply that the decision maker is indifferent between a perhaps widely dispersed probability distribution and a non-random amount equal to the size of the position parameter. Such indifference cannot be justified from a normative point of view and contradicts all experience. The existence of insurance companies gives a clear indication that the mathematical expectation is defective, for, in the long run, the premium revenue has to exceed the indemnification payments. From the view point of the insured this means that the premium he pays is larger than the expected indemnification he receives, i.e., that a game with a negative expected gain is being played or that, among two end-of-period wealth distributions, the one with the lower expected value is chosen. Unless this aspect is explained by the hypothesis that the insurance purchaser systematically overestimates the objective loss probability, preference functionals have to be constructed that allow for risk aversion by including the dispersion of the end-of-period wealth distribution in the evaluation procedure. All criteria discussed below satisfy this requirement.

---

[4] Cf. also the overviews of ARROW (1951), SCHNEEWEISS (1967a, pp. 20–26), and MARKOWITZ (1970, pp. 286–297).

[5] Cf. the introduction of section I B 3.1.

[6] Cf. ch. IV A.

Section A
The Two-Parametric Substitutive Criteria

An obvious way of taking undesirable dispersions into account is to represent the probability distribution by one parameter measuring a mean return (K_1) and another parameter measuring risk (K_2), and then to assume a utility function over these parameters:

$$(1) \qquad R(V) = U(K_1, K_2).$$

This is the way with criteria a) through e) of Table 1. Of course it is always assumed that $U_1 > 0$[1]. The role of the second argument is not so self-evident, so that it is better to distinguish the general cases

$$(2) \qquad U_2 \begin{cases} <0 & \text{risk aversion,} \\ =0 & \text{risk neutrality,} \\ >0 & \text{risk loving.} \end{cases}$$

It is, however, usual, in the light of the insurance phenomenon, to consider the case $U_2 < 0$ as the only one of practical relevance. In what follows, therefore, no more time will be wasted on the other possibilities.

Often the preference structure is illustrated graphically in a (K_1, K_2) diagram by means of *indifference curves* on which, by definition, $R(V) = U(K_1, K_2) = \text{const}$. For $K_2 > 0$ they are positively sloped[2]:

$$(3) \qquad \left.\frac{dK_1}{dK_2}\right|_{U=\text{const.}} = -\frac{U_2}{U_1} > 0.$$

In addition, they are also usually assumed to be convex because of a 'decreasing willingness to bear uncertainty'[3]. Examples of such curves are shown in Figure[4] 2.

---

[1] We denote by f_i the derivative of a function $f(.)$ with respect to its ith argument. Accordingly f_{ij} indicates a derivative with respect to i and j.

[2] It will be shown later that, for the (μ, σ) criterion, the indifference curves have to enter the μ axis perpendicularly. Cf. equation (II D 52).

[3] LANGE (1941, p. 183).

[4] Because of

$$\left.\frac{d^2K_1}{dK_2^2}\right|_U = \frac{\partial}{\partial K_2}\left(\left.\frac{dK_1}{dK_2}\right|_U\right) + \left.\frac{dK_1}{dK_2}\right|_U \frac{\partial}{\partial K_1}\left(\left.\frac{dK_1}{dK_2}\right|_U\right) = \frac{-U_1^2U_{22} + 2U_{12}U_1U_2 - U_2^2U_{11}}{U_1^3} > 0$$

the convexity of the indifference curves requires a cardinal function $U(K_1, K_2)$ with at least one negative second-order partial derivative or a negative partial cross derivative. Only DOMAR and MUSGRAVE (1944, p. 402) mention the sufficient conditions $U_{22} < 0$, $U_{11} < 0$, and $U_{12} = 0$.

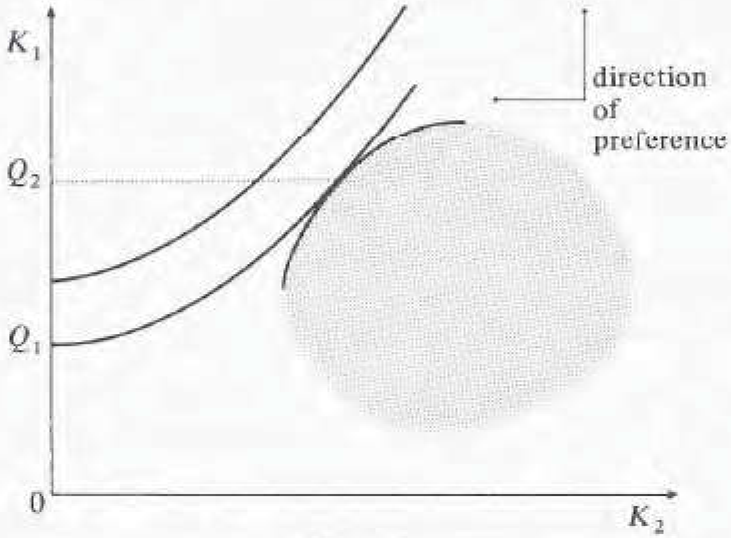


Figure 2

In addition to the indifference curves, the (K_1, K_2) diagram contains an *opportunity locus* consisting of a number of points, each of which represents one of the attainable end-of-period wealth distributions and thus one of the rows of the decision matrix. By use of the indifference curves it is easy to find the best distribution. Where there is a continuum of distributions, the procedure is that first the so-called *efficiency frontier* is determined by taking the north-east boundary of the opportunity set and then, by means of a tangency solution, one point at least on this boundary is found to be optimal.

For criteria a) through d) in Table 1, the starting point (e.g. Q_1) of an indifference curve at the K_1 axis is the so-called *certainty equivalent*, $S(V)$, of those wealth distributions which are located on this indifference curve. The certainty equivalent is a non-random level of end-of-period wealth that is equal in value to these distributions. The difference between the value of the position parameter K_1 and the certainty equivalent (e.g. $\overline{Q_1 Q_2}$) is called the *subjective price of risk*, π. The subjective price of risk measures the largest reduction of the position parameter which the decision maker would be willing to accept if his wealth distribution were exchanged for a non-random level of wealth.

1. *Lange's Criterion*

$$R(V) = U[M(V), v_{\max} - v_{\min}]$$

Lange (1943) assumes that, in general, the decision maker is confronted with a subjective probability distribution. For the sake of analytical simplicity, however, he further assumes that only the *range* and the *mode* are known (p. 182).

The question is whether essential information will be overlooked when such a qualification is made. This will indeed be the case as can easily be shown with the aid of Figure 3. The two probability distributions A and B in this figure are mirror images of each other on either side of a vertical line through the mode. Thus they can be considered equal with respect to Lange's distribution parameters. Nevertheless, distribution B is likely to be preferred by the typical decision maker, since values of $v<M(V)$ occur with a lower, the value $v=M(V)$ with the same, and values of $v>M(V)$ with a higher probability density than in distribution A.

It is not surprising, therefore, that Lange's criterion has not been taken up in the literature.

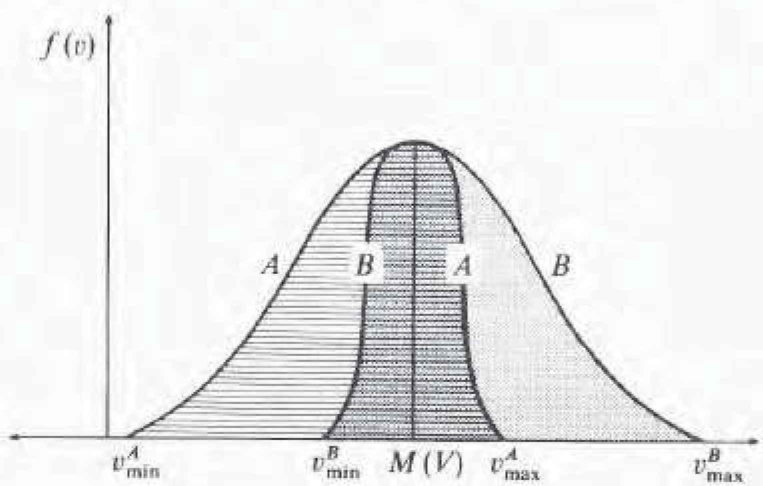


Figure 3

2. *The Domar–Musgrave Criterion*

$$R(V)=U[E(Y), -\int_{-\infty}^{0} yf(y+a)dy]$$

Domar and Musgrave (1944) maintain the classical measure[5] $E(Y)$ as the measure of returns (K_1). With regard to the risk measure (K_2), they consider that both the probability of loss and the expected size of loss are relevant[6]. Thus they feel it is natural to take the sum of the products

[5] Domar and Musgrave formulate their preference functional with reference to a percentage change in initial wealth a. This peculiarity is unimportant in the present context since $a=$const.

[6] A preference functional that is based on the expected value and the probability of loss is discussed by Schneeweiss (1967a, pp. 57–60) and has been mentioned by A.D. Roy (1952, p. 433). Allais (1952, pp. 317–320) also took it into consideration; cf. the discussion by Marschak, Allais, and Savage of Samuelson (1952a, pp. 151–155).

of all possible losses and their probabilities as a risk measure. Since the probability of loss is defined as

$$(4) \qquad W(y<0) \equiv \int_{-\infty}^{0} f(y+a)dy$$

and the loss expectation as

$$(5) \qquad E(\bar{Y}) \equiv \frac{-\int_{-\infty}^{0} yf(y+a)dy}{W(y<0)},$$

the Domar–Musgrave risk measure can be expressed as the product of loss expectation and loss probability:

$$(6) \qquad K_2 = E(\bar{Y})\, W(y<0).$$

It is not difficult to guess why the authors chose this approach when it is realized that, among other aspects, they intended to study the influence of an income tax without loss offset on the evaluation of given probability distributions. With their definition, the size of 'risk' is not affected by the tax, which is an important analytical simplification. Most of the other measures of risk would have created greater problems.

It therefore is not difficult to show the deficiencies in the approach. Analogously to (5), define the mathematical expectation of positive changes in wealth as

$$(7) \qquad E(\overset{+}{Y}) \equiv \frac{\int_{0}^{+\infty} yf(y+a)dy}{W(y \geq 0)}.$$

Then the preference functional becomes

$$(8) \qquad R(V) = U[W(y \geq 0)E(\overset{+}{Y}) - W(y<0)E(\bar{Y}), W(y<0)E(\bar{Y})].$$

This formulation shows that the shape of the density function may be arbitrarily modified for positive and negative changes in wealth as long as the respective partial expectations and the probability of loss remain unchanged. We are thus back to the deficiency of the mean-value criterion although now on another level. So the Domar–Musgrave approach did not become popular either[7].

[7] The approach was applied by BROWN (1957/58). An extensive critique is given by RICHTER (1959/60).

3. *The* (μ, σ) *Criterion*

$$R(V) = U[E(V), \sigma(V)]$$

Irving Fisher (1906, pp. 406–410) seems to have been the first to suggest an evaluation of economic probability distributions by means of their *mean value* ($\mu \equiv E(V)$) and their *standard deviation* ($\sigma \equiv \sigma(V) \equiv \sqrt{E\{[V-E(V)]^2\}}$). Later this approach was also discussed by Hicks (1933), Marschak (1938), Steindl[8] (1941), Tintner (1941), and F. and V. Lutz (1951, pp. 179–192). Since its application to problems of portfolio analysis by Markowitz (1952a) and Tobin (1958), the (μ, σ) criterion has become the most frequently used two-parametric approach.

In comparison to the risk measure of Domar and Musgrave, the standard deviation has the advantage of reacting to a change in the dispersion given the partial expectations for positive and negative changes in wealth and given the loss probability. This can easily be seen by splitting up the total variance into[9]

$$\sigma^2(V) = W(y<0)[\sigma^2(\overset{-}{Y}) + E^2(\overset{-}{Y})] + W(y \geq 0)[\sigma^2(\overset{+}{Y}) + E^2(\overset{+}{Y})] - E^2(Y). \tag{9}$$

If, given $W(y<0)$, $E(\overset{-}{Y})$, and $E(\overset{+}{Y})$, the partial variances $\sigma^2(\overset{-}{Y})$ and $\sigma^2(\overset{+}{Y})$ are changed, then the total variance changes in the same direction. The Domar–Musgrave risk measure would not have detected this change.

However, other cases can be constructed where the (μ, σ) criterion

---

[8] Steindl develops the (μ, σ) criterion in a somewhat disguised form. If, in line with our initial assumption, his multiperiod approach is reduced to one period, then it can be expressed as

$$R(V) = S(V) = a + \frac{\mu - a}{1 + r + h(\sigma)} \equiv U(\mu, \sigma),$$

where r is the rate of interest and $h(.)$ is a subjective risk evaluation function. Since Steindl assumes (pp. 50 f.) $h'(.)>0$ and $h''(.)>0$, this approach implies the normal properties of indifference curves in a (μ, σ) diagram:

$$\left.\frac{d\mu}{d\sigma}\right|_U = \frac{\mu - a}{1 + r + h(\sigma)} h'(\sigma) > 0$$

and

$$\left.\frac{d^2\mu}{d\sigma^2}\right|_U = \frac{\mu - a}{1 + r + h(\sigma)} h''(\sigma) > 0.$$

[9] The equation is developed in appendix 1 of this chapter.

does not appear to be so attractive. Mirror, for example, distribution A of Figure 4 on a vertical line at its expected value so that distribution B is created. Then the mean, the standard deviation, and hence also the value of the preference functional are unchanged. The Domar-Musgrave criterion however points to a clear improvement since given $E(Y)$ the risk measure is reduced via $W(y<0)$ and $E(\bar{Y})$ at the same time.

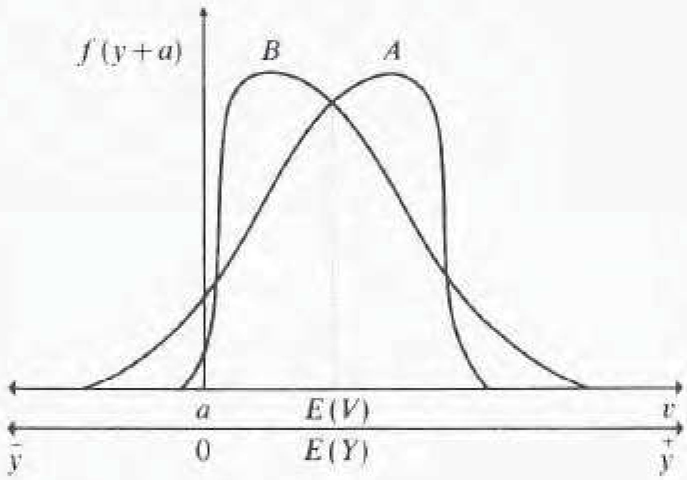


Figure 4

4. *The Mean-Semivariance Criterion*

$$R(V)=U[E(V), \int_{-\infty}^{v^*} (v-v^*)^2 f(v)dv]$$

Since, according to the (μ, σ) criterion, it does not matter whether changes in dispersion occur in the range of high or low levels of wealth, Markowitz (1970, pp. 188–201) suggests replacing the variance by the semivariance

$$(10) \qquad \sigma^2_{v^*}(V) \equiv \int_{-\infty}^{v^*} (v-v^*)^2 f(v)dv.$$

He does not say very much about the position of the critical wealth level v^*. It is, however, worth nothing that v^* may arbitrarily take on a level that is independent of the distribution to be evaluated or may take on a level equal to the expected value of the respective distribution $(v^*=\mu)$.

If the semivariance is written as[10]

$$(11) \qquad \sigma^2_{v^*}(V)=W(v<v^*)\{\sigma^2(V^*)+[E(V^*)-v^*]^2\},$$

[10] See appendix 2 of this chapter.

where V^* denotes wealth levels smaller than v^*, then it is obvious that the mirroring illustrated in Figure 4 leads to a diminution of $W(v<v^*)$, $\sigma^2(V^*)$, and $[E(V^*)-v^*]^2$ if $v^*=a$. Thus, for the example considered, the semivariance indicates a clear improvement.

Moreover, like the variance itself, the semivariance reflects changes in the dispersion of losses given the expectation and probability of loss. This follows immediately from (11).

Thus it seems that the semivariance combines the advantages of the Domar–Musgrave risk measure and the variance. However, it also has its deficiencies. Although the disperson above v^* might be relatively unimportant to the decision maker, does it make sense to assume that he will ignore it completely?

5. *The Criterion of Equivalent Gains and Losses*

$$R(V)=U(\overset{-}{y}^*,\overset{+}{y}^*)$$

As distinct from approaches that rely on the other two-parametric criteria, the approaches of Shackle (1952, pp. 9–31), Krelle (1957), and H. Schneider (1964, pp. 96–186) do not use statistical distribution parameters as risk (K_2) and return (K_1) measures. They prefer *subjectively* assessed index numbers of the distributions to be evaluated. Thus there are two ways in which the decision maker's preferences affect the evaluation of a probability distribution. The first is the usual one via the shapes of indifference curves in a (K_1, K_2) diagram. The second is through the formulation of these index numbers.

5.1. *Shackle's Approach*

Maintaining the interpretation indicated in chapter one, namely that Shackle's degree of potential surprise is basically a converse probability[11] we can illustrate his theory with reference to Figure 5.

The bell-shaped curve represents a usual probability distribution. By a suitable choice of two parameters, called *focus gain* ($\overset{+}{y}^*$) and *focus loss* ($\overset{-}{y}^*$), the probability distribution is measured and then, by the use of these parameters, a point in the (K_1, K_2) diagram of Figure 2 is determined. The measuring is done by means of so-called *contour lines* that, in the above figure, enter the auxiliary line parallel to the abscissa. In the range of gains, the contour lines indicate combinations of the level of gain and the probability density which the decision maker considers

[11] Cf. chapter I B 1 in this book and Krelle (1957, pp. 648–651).

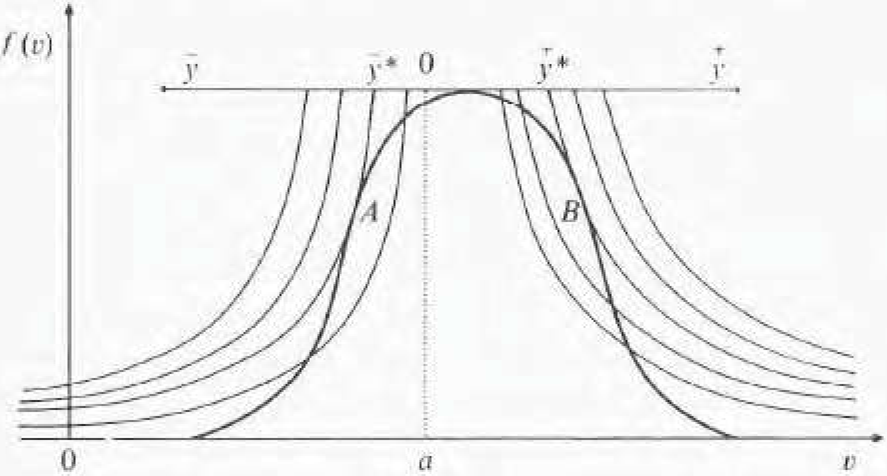


Figure 5

to be equally attractive. Similarly, in the range of losses, the contour lines represent points of equal deterrence. Applying the proverb that a chain is only as strong as its weakest link, Shackle contends that a probability distribution should only be measured by the positions of the outermost contour lines they reach. In Figure 5 these positions are characterized by $\bar{y}^*$ and $\overset{+}{y}^*$.

However, it seems that this contention is the weakest link in Shackle's chain of reasoning. It is hard to believe that, as long as the points of tangency A and B remain unchanged, the decision maker is indifferent to arbitrary modifications of the probability distribution.

5.2. *The Krelle–Schneider Approach*

Thus it is easy to understand why Krelle and Schneider suggest employing the total shape of the probability distribution in the assessment of the index numbers $\overset{+}{y}^*$ and $\bar{y}^*$, which they call *equivalent gain* and *equivalent loss*. They transform the probability distribution to a subjectively equivalent binary distribution which has one variate on each of the positive and the negative parts of the income axis with given, but arbitrary probabilities $\overset{+}{w}$ and $\bar{w}$, $0<\overset{+}{w}<1$, $0<\bar{w}<1$, $\overset{+}{w}+\bar{w}=1$. The two variates are the index numbers sought.

The uniqueness of the equivalent binary distribution is a problem with this approach. If a point in the $(\overset{+}{y}^*, \bar{y}^*)$diagram is found that represents a particular probability distribution, then the indifference curve passing through this point is the geometrical locus of all other points that represent this probability distribution just as well as the original point does. The opportunity locus would then itself consist of a set of indifference curves, and the conceptions of equivalent gain and equivalent loss would lose their meaning.

Krelle and Schneider try to circumvent the problem. They develop procedures, similar to each other, by which it is possible to construct, in a step-wise fashion, unique[12] equivalent binary distributions. However, which equivalent gains and losses are determined depends on the arbitrary characteristics of the procedures.

Consider, for example, Schneider's transformation procedure[13]. There are different transformation steps. On each step the probability distribution is altered without changing its value from the viewpoint of the decision maker. First the gain and loss distribution are both replaced by one variate each, without altering the gain or loss probabilities as such. Then these two probabilities are compared with $\overset{+}{w}$ and $\bar{w}$. If the probability of gaining is greater than $\overset{+}{w}$, it is reduced to $\overset{+}{w}$, the difference being assigned to a *nullchance*[14], and the variate on the positive income axis is altered to create the required indifference to the original distribution. Then the nullchance is removed and its probability is utilized to increase the probability of the variate on the negative income axis. Finally, the variate on the negative income axis has to be adjusted appropriately, so as to ensure indifference. If the loss probability exceeds $\bar{w}$, the procedure is analogous.

Why is there a nullchance and not a \$ 14.25 chance? Why are the probabilities not adjusted to $\overset{+}{w}$ and $\bar{w}$ in one step only? If one of these possibilities were chosen, different values for the equivalent gains and losses would be calculated.

The equivalent gains and losses, therefore, cannot be interpreted as subjective central values. Would it not be better to dispense with the whole idea of constructing an equivalent 'two-point' distribution and to construct instead an equivalent 'one-point' distribution, that is, the certainty equivalent of the distribution to be evaluated? What is the advantage of the intermediate step?

---

12 SCHNEIDER (1964, pp. 110 f.) does not exclude the possibility that, with his procedure, a probability distribution may be represented by two different points in a $(\overset{+}{y}^*, \bar{y}^*)$ diagram depending on how the transformation procedure starts. A similar problem does not arise in Krelle's case since he assumes the (additive!) von Neumann–Morgenstern utility concept. However, since Schneider (pp. 98–101) accepts CHURCHMAN'S (1961, pp. 225–232) axiom system, from which such a utility concept follows, the ambiguity cannot appear in his model either. Cf. section II D 2.1.2.

13 SCHNEIDER (1964, pp. 101–103, 106 f.). The same could be shown for KRELLE (1957).

14 Cf. KRELLE (1961, pp. 90 f.).

KRELLE (1968, esp. p. 143) obviously no longer sees an advantage, for he replaces the conception by the expected-utility approach that in any case was underlying his analysis[15]. However, Schneider, uses the intermediate step for analyzing the influence of income taxes on the evaluation of risk projects. Since income taxes normally only affect the positive part of income distribution ($y>0$), it certainly makes sense in this case to distinguish between gains and losses.

However, there are fundamental problems of application. Under the influence of taxation the opportunity locus depicted in Figure 2 is subject to shift. For example, an income tax of 50% without loss offset reduces all profits of the original distributions by 50%. To find out how this affects the equivalent gains, it would be necessary to carry out the transformation procedure for each after-tax distribution of Y. Unfortunately, however, because of the generality of the described transformation procedure[16] not very much can be said about the relationship between the equivalent gains before and after taxation. Thus the applicability of the theory is not shown. What has been shown is that it needs to be supplemented by an ad-hoc assumption. It is assumed, although only for the sake of approximation (pp. 136 a. 156), that, under an income tax of 50%, the equivalent gains will also fall by 50%.

6. *Limits and Possibilities of the Statistical Criteria*

Thus we return to the criteria utilizing statistical distribution parameters although, as mentioned earlier, these also have their disadvantages. These disadvantages have the same general cause. *A priori* it is unlikely that people formulate their preference orderings over probability distributions with respect only to statistical parameters[17]. It therefore seems evident that a completely correct preference functional can be constructed only if the arguments of this functional can perfectly describe the distribution. For arbitrarily shaped probability distributions such a perfect description is generally impossible, however.

Of course, an attempt could be made to increase the number of para-

---

[15] The retreat from the original concept had already begun in KRELLE's 'Preistheorie' (1961, pp. 81–107, 588–610) where a $(\overset{+}{y}{}^*, \overset{-}{y}{}^*)$ diagram can no longer be found.

[16] Since, in Krelle's approach, there is a given 'Chancenpräferenzfeld', it would, in principle, be possible to calculate the equivalent net gains.

[17] It can be shown that, for a preference functional according to *Weber's law*, which will be derived in chapter three of this book, this case can indeed be excluded.

meters. For example, the (μ, σ) criterion could be extended by adding the third moment[18]

$$(12) \qquad E\{[V-E(V)]^3\},$$

which is a measure of skewedness, the fourth moment

$$(13) \qquad E\{[V-E(V)]^4\},$$

that measures particular aspects of curvature, and still other moments. But, although in this way it is possible to describe the distribution more accurately, a perfectly exact description of all possible distributions cannot be given with a finite number of moments.

Fortunately, in reality, the problem does not always arise in this severe form so that our criticism loses much of its force. The reason is that it is often possible to indicate a typical class of distributions for all elements of the opportunity set. To establish a preference ordering for the members of this class, a signficantly smaller number of moments is normally sufficient. Linear distribution classes in particular seem to occur very frequently.

It is said that the random variables $V_1, V_2, \ldots$ form a *linear class* if their standardized values

$$(14) \qquad Z_i = \frac{V_i - E(V_i)}{\sigma(V_i)} \quad \forall i$$

$$\text{with} \quad E(Z) = 0$$

$$\text{and} \quad \sigma(Z) = 1$$

have the same density function $f_z(z;0,1)$. Within a linear class all distributions can be transformed into one another merely by a shift and a proportional extension. For example, since $E(V)$ acts as a measure for

[18] Often the moments are used in their standardized forms

$$M_3 \equiv E\left\{\left(\frac{V-E(V)}{\sigma(V)}\right)^3\right\} \quad \text{and} \quad M_4 \equiv E\left\{\left(\frac{V-E(V)}{\sigma(V)}\right)^4\right\}.$$

A positive sign of the third moment indicates a right skewed distribution. The fourth moment is a measure of 'peakedness'. Since a normal distribution has $M_4 = 3$, the measure $M_4 - 3$ can be interpreted as follows:

$$M_4 - 3 \begin{Bmatrix} > \\ < \end{Bmatrix} 0 \quad \text{means} \quad \begin{Bmatrix} \text{'more peaked} \\ \text{'flatter} \end{Bmatrix}$$

than with a normal distribution'. Cf. STANGE (1970, pp. 86–89).

the shift and $\sigma(V)$ as a measure for the extension around the mean, these two parameters are sufficient to characterize the whole distribution, given the shape of the standardized distribution. They can thus be used to construct arbitrary preference functionals[19,20].

In chapter V of this book two practical problems will be considered where linear classes occur in an *exact* form. Linear classes are, however, important also because the class of normal distributions, which is one of these classes, can often be used as a good approximation for the class of distributions occurring in real-life decision problems[21]. The reason is that the real-life distributions often originate by adding independent random variables. If many such random variables are added, then, according to the *Central Limit Theorem* of Lyapunow, the standardized distribution of the sum approximates the standard normal distribution[22]. A necessary assumption is that the variances of the added

[19] The argument has been raised by TOBIN (1958, p. 12) as a defense of the (μ, σ) approach. FISHER (1906, p. 408), too, seems to have had it in mind. It is not clear whether he was thinking of a linear class in general or of the special case of a normal distribution. His numerical examples suggest the latter. Cf. also SCHNEEWEISS (1967a, pp. 121–129).

[20] THORP (1971, p. 262) criticizes the (μ, σ) criterion since it does not discriminate *a priori* between two distributions V_1 and V_2 that spread evenly between the values (1, 3) and (10, 100) (such that $E(V_1) < E(V_2)$ and $\sigma(V_2) > \sigma(V_1)$) although clearly $V_2 > V_1$. This criticism is not substantial. Since both the distributions he assumes belong to a linear class of rectangle distributions, the members of which can all be exactly represented in a (μ, σ) diagram, the better distribution can be selected by reference to suitably constructed indifference curves.

[21] An example is the distribution of returns in a well diversified portfolio. Cf. ch. V A.

[22] In the case of a normal distribution the density function of a random variable standardized according to (14) is

$$f_z(z;0,1) = \frac{1}{\sqrt{2\pi}} e^{-0,5z^2}.$$

By integration this yields

$$W(z < z^*) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z^*} e^{-0,5z^2} dz.$$

This expression is the subject of the Central Limit Theorem. This theorem says that, for a sum of independent (arbitrarily distributed) random variables $X_1, X_2, \ldots, X_n$ with variates $x_1, x_2, \ldots, x_n$, the following relationship holds:

$$\lim_{n\to\infty} W\left(\sum_{i=1}^{n} x_i < \varepsilon\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z_n^*} e^{-0,5z_n^2} dz_n$$

$$\text{with } z_n^* \equiv \frac{\varepsilon - E\left(\sum_{i=1}^{n} X_i\right)}{\sigma\left(\sum_{i=1}^{n} X_i\right)} = \frac{\varepsilon - \sum_{i=1}^{n} E(X_i)}{\sqrt{\sum_{i=1}^{n} \sigma^2(X_i)}}.$$

→

variables exist; since, in economic problems, all conceivable distributions are bounded from above and below, this assumption is always satisfied[23].

Although the preceding discussion referred to the parameters μ and σ, it supports equally well the use of other statistical parameters[24]. If the standardized form of the distribution is given, then, for known $E(V)$, the exact form of a distribution V can, for example, also be determined if the Domar–Musgrave risk measure or the semivariance is known[25]. Even Lange's criterion would be completely satisfactory, for the shape of V could be completely described with the range and the mode[26].

Therefore, the statistical two-parametric criteria after all appear in a very favorable light. Nevertheless, it must be admitted that there are a number of problems in practical decision making under uncertainty where linear distribution classes do not prevail. Moreover, concerning the approximation by a normal distribution, it should not be forgotten that the relevant sums are often rather small and that there are mutual dependencies between the variables added. In this case the Central Limit Theorem is not applicable. Thus a preference ordering based on two parameters often cannot be more than an approximation of the true preference ordering. This approximation can be improved by the use of additional distribution parameters. How far to proceed with this improvement is a problem of the economics of economic research. Most authors do not think it worth-while to employ a third parameter[27]. Moreover, the majority of the founders of economic theory limited their attention to only one parameter, without, however, saying which one. Perhaps these remarks put into the right perspective what Tobin (1969, p. 14) called 'the modest endeavour of doubling the number of parameters'.

A proof is given in Fisz (1970, pp. 241–251). It must be stressed once again that the *standardized* version of the distribution of the sum becomes normal. It is also possible to say that the non-standardized distribution will become *relatively* more similar to a distribution which is developed from the standard distribution and has the same mean and variance. This, however, does not mean that the distance between points of equal density in the two distributions will vanish *absolutely*.

[23] Cf., however, Fama (1968, p. 30) and Fama and Miller (1972, pp. 259–265).

[24] Cf. Tobin (1958, p. 12).

[25] It must be assumed, however, that the distributions are such that the risk measures do not have a value of zero.

[26] In the case of the unbounded normal distribution the range has to be replaced by another measure of risk; for example, as Lange (1943, p. 182, footnote) suggests, by the distance between the lowest and the highest percentiles.

[27] Three parameters are considered by Cramér (1930, p. 50), Marschak (1938, p. 320), Hicks (1967, pp. 117–125), and Jean (1971).

Section B
The Lexicographic Criterion

The decision criteria examined above have one thing in common: the substitutability between risk and return. With the lexicographic criterion, it is different. Here the decision maker is supposed to maximize the probabilities of wealth exceeding some critical levels. *Safety first* is a suitable slogan for characterizing the underlying preference. In the simplest case, which will be discussed first, there is only one critical level, the disaster level ($\bar{v}$). A modified version with multiple critical levels will be discussed later.

1. *The Unconditional Maximization of the Probability of Survival*

1.1. *The Formal Approach*

The theory of lexicographic preferences was developed by René ROY (1943) but got its name from the 'lexikographische Anordnung von Mengen' (lexicographic ordering of sets) formalized by HAUSDORFF (1914, p. 78)[1]. ENCARNACIÓN (1965) and NACHTKAMP (1969) extended it to the case of uncertainty. Lexicographic criteria for decision making under uncertainty have also been employed by H. CRAMÉR (1930), A.D. ROY (1952), and HAUSSMANN (1968/69).

By referring to Hausdorff's formulation, the basic idea of the theory can be described as follows. The decision maker's task is to compare two commodity bundles (a, b) and (a', b'), where a, a' and b, b' indicate the quantities of goods A and B respectively in the two bundles. A lexicographic ordering then implies

$$
\begin{aligned}
(1)\qquad &\text{that} && (a,b)\ \{\lessgtr\}\ (a',b') \\
& && \text{if, and only if,} \\
& && \text{either} \quad a\{\lessgtr\}a' \\
& && \text{or} \quad a=a', \text{ but } b\{\lessgtr\}b' \\
&\text{and that} && (a,b) \sim (a',b') \\
& && \text{if, and only if, } a=a' \\
& && \qquad \text{and} \quad b=b'.
\end{aligned}
$$

Thus a choice is made first with reference to the quantity of goods of type A. The bundle with the larger quantity of these is preferred whether

[1] The theory was extended by GEORGESCU-ROEGEN (1954), CHIPMAN (1960), BANERJEE (1964), ENCARNACIÓN (1964a and b), FERGUSON (1966), and others.

or not the other bundle is better in terms of goods of type B. Only if both bundles are equal with respect to the *predominant* good A, is good B considered[2].

What are the implications of this basic idea for decision making under uncertainty? If, for example, we have the choice problem of the firm in mind, then A could plausibly be interpreted as the aim of economic survival that dominates some other aim B. Suppose there is no policy available that guarantees survival with absolute certainty. Then the best that can be done is to take the one that maximizes the survival probability $W(v > \bar{v})$. This probability corresponds to the parameters a and a' in the above formulation. Analogously, the parameters b and b' can be interpreted as probabilities of attaining the second-order aim B. This aim is not considered as long as it is possible to find a policy that is unambiguously better with respect to the aim of survival. Under this constraint the preference functional therefore is

$$(2) \qquad R(V) = W(v \geq \bar{v}) = \int_{\bar{v}}^{\infty} f(v)dv.$$

A.D. Roy (1952) introduced a certain modification to this basic concept. He assumed that the decision maker knows the variance and the expected value, but not the shape, of the probability distribution. With this limited information it is impossible to calculate the survival probability. It is, however, possible to determine a lower bound for it. This is done with the aid of *Chebyshev's* inequality[3]:

$$(3) \qquad W[|v - E(V)| \leq \varepsilon] \geq 1 - \left(\frac{\sigma(V)}{\varepsilon}\right)^2, \quad \varepsilon > 0.$$

If we set $\varepsilon \equiv E(V) - \bar{v}$ and drop the absolute-value operators, since we are only interested in the negative deviations from the expected value, we obtain the expression

$$(4) \qquad W[E(V) - v \leq E(V) - \bar{v}] = \mathrm{W}(v \geq \bar{v}) \geq 1 - \left(\frac{\sigma(V)}{E(V) - \bar{v}}\right)^2$$

which describes a lower bound to the probability of survival. Since the quotient $[E(V) - \bar{v}]/\sigma(V)$ can be derived from this lower bound through

---

[2] The term 'lexicographic' is used because the ordering is like that in a lexicon.

[3] For a derivation of this inequality see, e.g., Stange (1970, pp. 157 f.).

a strictly positive monotonic transformation, it can, in the absence of a better solution, be used as a preference functional[4]:

$$(5) \qquad \hat{R}(V) = \frac{E(V) - \bar{v}}{\sigma(V)}.$$

Although the idea that the rational (!) decision maker does not know subjective probabilities cannot be accepted, Roy's approach nevertheless suggests a very useful application of the (μ, σ) principle. Roy (p. 434) remarks that utilizing $\hat{R}(V)$ will not only maximize the lower bound of survival probability, but also the survival probability itself, if a choice from a class of normal distributions is made[5]. However, it is not in fact necessary to limit attention to this class. For each linear distribution class, the use of $\hat{R}(V)$ maximizes the probability of survival[6]. It is easy to understand this if equation (A 14) is recalled and (2) is written in the standardized form

$$(6) \qquad R(V) = \int_{-\hat{R}(V)}^{+\infty} f_z(z; 0, 1)\, dz.$$

This expression proves that $\hat{R}(V)$ and $R(V)$ can be produced from one another by a strictly positive monotonic transformation and will thus result in the same preference ordering[7].

Figure 6 shows how, with the aid of the (μ, σ) diagram, the probability distribution with the highest survival probability can be found. The rays starting from $\bar{v}$ connect points with equal probability of survival, for on these rays $[E(V) - \bar{v}]/\sigma(V) = \text{const.}$[8]. The survival probability is an increasing function of the angle α between such a ray and the

[4] Roy considers the upper limit of the disaster probability rather than the lower limit of the survival probability. Of course, this is the same problem.

[5] Cf. footnote 22 in section A.

[6] It was shown in section A 6 that, in the case of a linear distribution class, any kind of preference structure can be represented in a (μ, σ) diagram. Cf. Pyle and Turnovsky (1970) and Levy and Sarnat (1972) for the possibility of representing the lexicographic aim.

[7] There are parallels to Nachtkamp's (1969) model where a firm is able to evaluate probability distributions of demand quantities without knowing the shapes of these distributions (pp. 164–176).

[8] In Roy's case of unknown distributions, the positively sloped rays connect points of equal lower bound and the negatively sloped rays connect points of equal upper bound to the survival probability. The latter information is, of course, irrelevant. But even in the case of positive slopes, the probability information is only useful if the slope is >1, for otherwise (4) would yield the useless information that the survival probability is greater than zero (slope $=1$) or greater than a negative number (slope <1).

abscissa. Thus the safest project is indicated by the highest point of tangency of a ray with the opportunity locus, which is suggested in the figure by the curved segment with the shaded area underneath.

Since in the figure there is only one point of tangency, there is a unique risk project maximizing the probability of survival. With differently shaped opportunity loci, multiple points of tangency may occur so that lower-order aims have to be considered before a decision can be made. Thus the rays starting from $\tilde{v}$ are not indifference curves. 'Indifference' prevails with respect to the predominant aim but not necessarily with respect to lower-ranking aims. Following CHIPMAN's (1960, p. 202) suggestion we could therefore call these rays *pseudo indifference curves*.

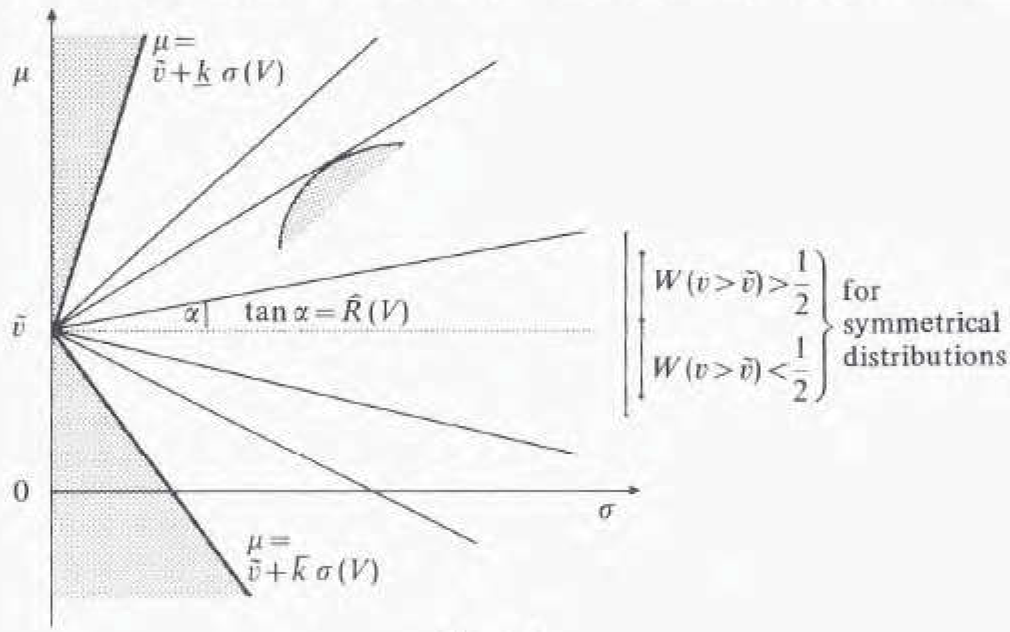


Figure 6

Another case where the aim of maximizing the survival probability does not lead to a unique solution is that where the opportunity locus either exceeds the upper boundary ray or is situated below the lower boundary ray. Such boundary rays will occur if the linear distribution class has a standardized distribution that is bounded from above and below as shown in Figure 7. If $-\underline{k}$ is the highest lower bound and $+\bar{k}$ the lowest upper bound to the standardized variable z then the survival probability is

$$(7) \qquad W(v \geq \tilde{v}) = \begin{Bmatrix} 1 \\ 0 \end{Bmatrix} \Leftrightarrow E(V) \quad \begin{Bmatrix} \geq \tilde{v} + \underline{k}\sigma(V) \\ \leq \tilde{v} - \bar{k}\sigma(V) \end{Bmatrix}.$$

So much for the formal aspects of the theory. A much more significant question is whether there is in fact an insolvency level of wealth

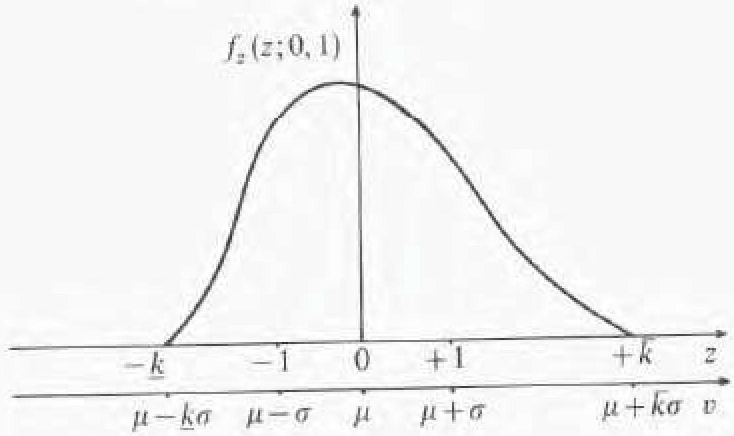


Figure 7

that plays a crucial role in economic decision making under uncertainty and, if there is, what is its value.

1.2. *The Problem of the Disaster Level*

To ensure that the unconditional maximization of the survival probability makes sense, there must be somewhere a disaster level of wealth below which total collapse occurs. The collapse must occur abruptly at this level. By how far wealth is below the disaster level does not really matter. The same applies to survival. The crucial thing is to survive at all, how well is not so important. The slightest reduction in the probability of survival is a change for the worse, even if all the treasures of the earth are gained in exchange.

For an individual, absolute disaster naturally means physical death. But we are not normally concerned with matters of life and death in typical economic decision making. Here disaster is better interpreted as, what for many people is nearly as bad as death, the destruction of their whole way of life through the loss of all their property. In this case $\bar{v}=0$. Whether at $v=\bar{v}=0$ there is in fact a level of disaster in the strict lexicographic sense is an empirical question that cannot be answered here. But if there is a lexicographic critical wealth level at all, it will be zero.

There are, however, reasons to believe that a lexicographic level of disaster is not a very typical feature of man's preferences. To gain pecuniary advantages most people drive a car and so risk their lives. Similarly, many people are unwilling to buy liability insurance although insurance is a protection against being reduced to beggary. But there are also people who do go out of their way to buy liability insurance and are frightened even at the thought of driving a car. Perhaps there are lexicographic wealth levels for a few people, although not for many. We leave this an open question.

If the decisions of the firm rather than those of the individual or the household are considered, disaster means insolvency: the firm cannot meet its liabilities, bankruptcy proceedings are instituted. Following FISHER (1906, p. 82), a distinction can be made between *pseudo insolvency* and *true insolvency*. A pseudo insolvency originates from short-run liquidity problems without the debt exceeding the value of the firm's capital. With patience on the part of creditors, pseudo insolvency normally can be avoided by a reorganization of the structure of debt. In the case of true insolvency, however, the equity capital is not sufficient to satisfy all creditors.

In the case of pseudo insolvency no critical level of wealth in the lexicographic sense can be established. Suppose, because of the intolerance of creditors, a pseudo insolvency could lead to bankruptcy. Then the firm's management, whose predominant aim it is to avoid this situation, will certainly include liquidity planning in its policy. As long as it is prepared to bear the interest costs, the management can always increase the degree of liquidity by, at the same time, lending short and borrowing long. Thus the level of insolvency itself is a planning variable and the concept of an unconditional maximization of the survival probability loses its meaning.

The situation is completely different if a further increase in the degree of liquidity is impossible, since no more security is available for the creditors. This is the case if the volume of long-term debt equals the firm's stock of capital. Thus true insolvency occurs when the total liable wealth of the firm's owners is lost. As in the case of a single individual, we therefore arrive at the conclusion: if there is a lexicographic critical wealth level, then it will occur at $v=\tilde{v}=0$.

2. *Aspiration Levels and Saturation Probabilities: A Pragmatic View*

The unconditional maximization of the survival probability implies a degree of risk averse behavior so extreme that it is hardly ever observable in reality. And, even if it occurs, the typical entrepreneur, so convincingly portrayed by SCHUMPETER (1942), definitely does not show it. Figure 8 demonstrates the behavioral implication of the degree of risk aversion contended by lexicographic preference theory. For the kind of opportunity set assumed in this figure, in the case where $\tilde{v}=0$, the decision maker chooses project *A* although it provides a dramatically low expected value of end-of-period wealth. This is not very plausible. Another implausible aspect is the lack of any subjective element in the choice of the optimal risk project[9]: according to the preceding section,

[9] In the case of uncertainty there is, however, a subjective scope insofar as people are differently informed and will therefore assess different objective probabilities.

$\bar{v}=0$ will hold for all decision makers and hence they will all choose the same project from a given opportunity set.

It is therefore an obvious modification of this approach to replace the disaster level with some higher *aspiration* level of wealth that is determined by the particular preferences of the individual decision maker. For example, for the decision problem of the firm we could follow NACHTKAMP (1969, p. 120) and choose a critical wealth level that would be achieved if the current profits reached the level of the previous year. Or we could, as CRAMÉR (1930, p. 10) and ENCARNACIÓN (1964a, p. 113) suggest, choose that level of wealth which allows the usual dividends to be paid. Arbitrarily chosen other wealth levels could also be considered. In all these cases a project would be chosen out of the opportunity set depicted in Figure 8 that is characterized not only by a higher expected wealth level, but also by a higher standard deviation. Revealed risk aversion obviously would be smaller. If, as an extreme case, a critical level of wealth equal to the maximum expected level is chosen, then risk aversion would be eliminated altogether[10]. This case is represented by point B in Figure 8.

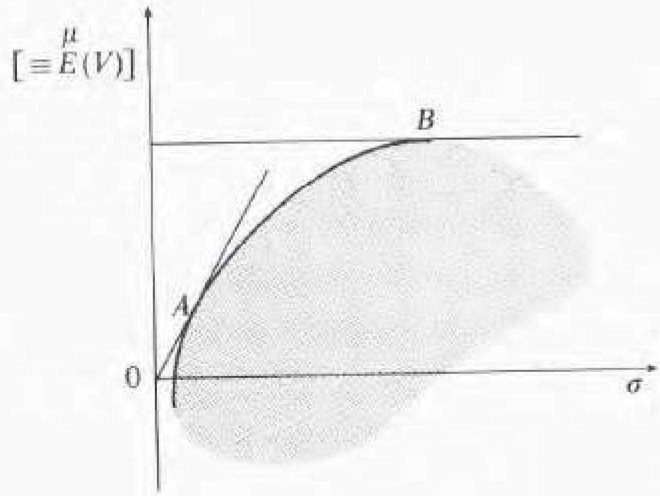


Figure 8

Although the basic idea of the aspiration level has some appeal, it is nevertheless questionable whether aspiration levels satisfy the strong requirements of lexicographic critical wealth levels. Is it plausible to assume that the decision maker would not be willing to accept a slight increase in the probability of failing to achieve the aspiration level if, in exchange, he gets a kingdom if he does achieve it? The answer that most people will give to this question seems obvious.

Moreover, even if there is an aspiration level ($\bar{v}_2$) in the lexicographic

[10] Cf. the definition of risk aversion given at the beginning of section A.

sense, this does not mean that the level of true insolvency, call it now $\tilde{v}_1$, loses its significance. The predominant aim still is to maximize the probability that wealth exceeds the insolvency level. Thus the introduction of aspiration levels does not in itself produce more plausible behavioral implications.

The way out of this dilemma is to introduce a *saturation* level $W^*(v \geq \tilde{v}_1)$ for the survival probability[11], analogous to the saturation level that R. Roy (1943) used for the case of commodity consumption. All projects with a higher probability of survival may then be considered as equivalent with respect to the predominant aim and the probability of reaching a particular aspiration level determines the ultimate choice among these projects. A more precise definition of this preference structure can be given by reference to Hausdorff's formulation (1) if, for a particular project, we set

$$(8) \qquad \begin{aligned} a &\equiv \min\,[W(v \geq \tilde{v}_1), W^*(v \geq \tilde{v}_1)], \\ b &\equiv W(v \geq \tilde{v}_2) \end{aligned}$$

and define a' and b' analogously for a comparable project.

Of course, there is no particular reason why there should be only two critical levels of wealth. There may be a saturation level for the probability of exceeding the aspiration level, so that another aim of lower order appears on the scene. If there is a saturation level for the probability of reaching this aim, a further aim may be considered and so on. The various aims do not have to be pecuniary. However, since we agreed to analyze wealth distributions, a hypothesis of Nachtkamp (1969, p. 120) according to which there are multiple critical wealth levels seems to be of interest.

For a linear class of distributions and three critical wealth levels $\tilde{v}_1$, $\tilde{v}_2$, and $\tilde{v}_3$, the preference structure is shown in Figure 9. For each of the critical wealth levels there is a bundle of pseudo indifference curves of the kind shown in Figure 6. The parallel lower boundaries of these bundles maintain their previous (cf. (7)) meaning: because there is an upper bound on the standardized probability distribution, below these

[11] Criticizing the approach by A.D. Roy, Telser (1955/56, pp. 2 f.) postulated a saturation probability, after the achievement of which the mathematical expectation is to be maximized. Baumol (1963) wanted to supplement the usual (μ, σ) approach for substitutive indifference curves by weeding out those projects that do not ensure exceeding a critical wealth level with a given minimum probability. Nachtkamp (1969) developed the hypothesis that the firm first wants to achieve a particular profit goal until its saturation probability is reached, then a sales goal until a further saturation probability is reached, and finally an expected-utility goal. A theoretical analysis of saturation probabilities is given by Encarnación (1965) and Nachtkamp (1969).

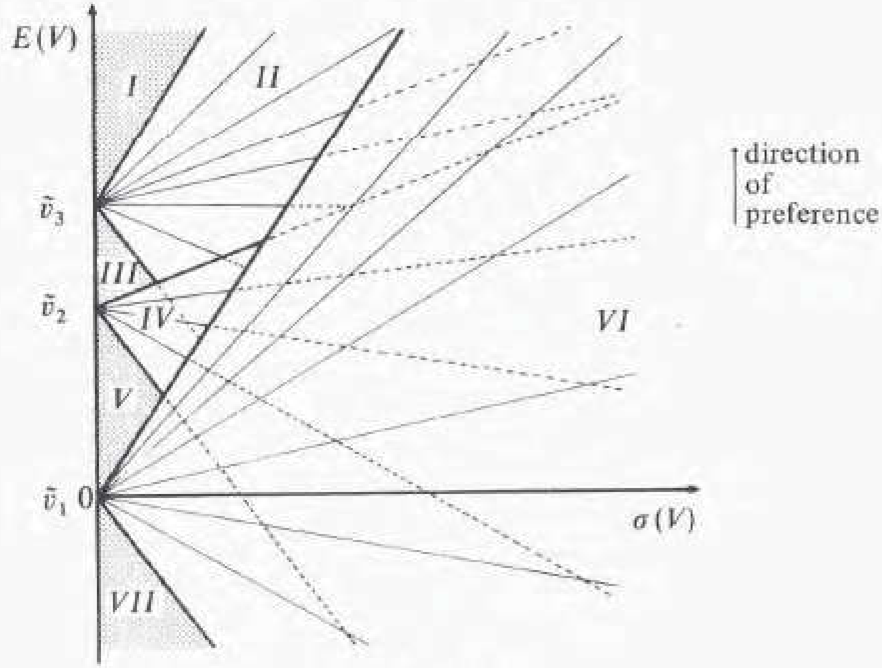


Figure 9

boundaries the probability of exceeding the corresponding critical wealth level is zero. (If the standardized distribution is not bounded from above, the lower boundaries of the indifference-curve bunches coincide with the ordinate.) The upper boundaries, however, gain a new meaning. Each of these boundaries is the geometrical locus of all those projects for which the probability of exceeding the corresponding critical wealth level just equals the saturation probability. Above such a boundary, the probability of exceeding the critical wealth level exceeds the saturation probability. Thus

(9) $$W(v \geq \bar{v}_i)\{\gtreqq\}W^*(v \geq \bar{v}_i) \Leftrightarrow E(V)\{\gtreqq\}\bar{v}_i + \underline{k}_i^*\sigma(V), \quad i=1,2,3,$$

if $-\underline{k}_i^*$ is that variate of the standardized distribution Z for which

(10) $$\int_{-\underline{k}_i^*}^{\infty} f_z(z;0,1)\,dz = W^*(v \geq \bar{v}_i).$$

If, as seems plausible, the critical wealth levels are inversely related to their ranks, the areas labelled with Roman numerals are achieved. The best project from a given opportunity set can then be found by the following procedure. First the intersection with the field with the lowest number has to be found. If the number is even then, from this intersection, the point(s) lying on the highest pseudo indifference curve must be selected. If there is more than one such point the broken-line pseudo indifference curves are consulted for a choice between them. If the number is odd, the choice is partly indeterminate since all projects within the intersection have the property of satisfying the higher-ranking aim with a probability higher than the saturation probability, but, at the same time, of not satisfying the lower-ranking aim at all.

Through the introduction of saturation probabilities lexicographic preference theory gained a high degree of flexibility. A large number of choices from a given opportunity set can be modelled by a suitable selection of critical wealth levels and saturation probabilities[12].

Nevertheless, the question of plausibility has to be asked. Why are there saturation probabilities? They have little in common with the kind of saturation associated with a full stomach. A possible, but not satisfactory, answer is given by the phenomenon of thresholds. An aim that is nearly reached with probability one by a number of projects is considered as practically certain and hence irrelevant for the choice between these projects. But why should there be thresholds only in the neighbourhood of one? Would it not be equally plausible to conjecture that probabilities close to 0.7 are practically 0.7 and others close to 0.1 practically 0.1[13]? It seems that thresholds can hardly be used to legitimate saturation probabilities[14].

But even if there are other reasons which may account for saturation probabilities, lexicographic ordering maintains the basic characteristic of an absence of substitutability between the various aims. Why are lower-ranking goals *completely* neglected if the higher-ranking aim is not achieved with a satisfactory probability?

One argument cannot be denied: the operational advantages of lexicographic orderings[15]. Even in its modified version with saturation probabilities, the search procedure for an optimal decision seems to be simpler with lexicographic preferences than with substitutive ones. For this

---

12 Flexibility, however, is of no value in itself. Cf. STIGLER's approach to a 'Theory of Economic Theories' in his review article (1950, pp. 114–119, esp. p. 115).

13 Cf. KRELLE's (1961, p. 611) attempt to quantify verbal probability judgements.

14 The uselessness of introducing thresholds into preference-theoretical analysis is well demonstrated by observing how SCHNEEWEISS (1967b) and GEORGESCU-ROEGEN (1954, p. 522) who, concerning the question of substitutability have opposite opinions, lay the blame for thresholds at each other's feet. Schneeweiß explains the existence of a lexicographic ordering between two aims by the assumption that there is a threshold for the aim with the lower rank that disguises a change in the degree of goal achievement. He is silent about why there should not be a threshold for the other goal. A very similar argument is presented by Georgescu-Roegen in order to explain the observation of substitutive choice. A sign of the substitutability, he says, is that a person does not feel a disadvantage if the degree of goal achievement rises for the goal without a threshold but falls for the other one with a threshold. Here also, a legitimation for assuming only one threshold is missing.

15 Cf. CHIPMAN (1960, p. 222), HAUSSMANN (1968/69, p. 33), and NACHTKAMP (1969, pp. 325 f.).

reason the delegated preference orderings utilized in administration hierarchies are typically lexicographic. The reader should, for example, think of the sticklers for the rules that seem to be indispensable in government administration. There are however reasons to suspect that lexicographic orderings observable in reality are merely simplified models of underlying substitutive orderings[16]. Somebody has to write the rule books and it may be worth his while to consult a substitutive ordering. This suspicion is also supported by the obvious fact that in administration the simplified lexicographic preference ordering is frequently not considered to be sufficient. Why else does the institution of limited competence exist? Whenever a decision problem on one level of the hierarchy is outside the area of competence of a particular department, it is passed up through the hierarchy until it becomes possible to make a decision according to a more flexible, i.e., more substitutive, preference ordering.

Anyone can see for himself that not only in administrative hierarchies is it useful to decide with the aid of a simplified picture of the true preference ordering. It would certainly involve too much effort to make a federal case out of every little day-to-day decision. Possibly, but certainly not necessarily, therefore, the simplified preference ordering we consult in our daily decisions is lexicographic. But here also the simplified ordering is not sufficient for extraordinary decisions; sometimes we have to think long and hard before coming to a decision.

Thus, for a theoretical analysis of decision fraught with grave consequences, it seems wise not to deal with lexicographic preferences, but rather to consider the underlying substitutive preference ordering itself.

Section C
The Expected-Utility Criterion

Two and a half centuries ago G. CRAMER[1] (1728) and D. BERNOULLI (1738) developed an idea that, after its axiomatic foundation by VON NEUMANN and MORGENSTERN (1947, pp. 17–29, 617–632), became the most popular approach to formulating a preference functional for the evaluation of probability distributions. The approach is to transform,

[16] This does not exclude the possibility that the usual assumption of a *continuous* substitutability of goals is itself an approximation of an ordering that in reality is *discrete*, as is to be expected in the light of the threshold phenomenon. Cf. however NACHTKAMP (1969, p. 325).

[1] Swiss mathematician (1704–1752). G. Cramer formulated his thoughts in a letter he sent to Nicolas Bernoulli, a cousin of Daniel Bernoulli. From there the letter went to D.B. who reproduced it in his article.

with the aid of a suitably chosen monotonically increasing index function $U(.)$, the probability distribution of end-of-period wealth into a probability distribution of index values and then to choose the expected value of this index distribution to be the preference functional:

$$(1) \qquad R(V)=E[U(V)].$$

If, as the Axiom of Ordering[2] requires, $R(V)$ is defined up to a monotonically increasing transformation, then $U(.)$ is determined up to a positive linear transformation, that is, $U(.)$ is measured by an interval scale[3]. The reason is that for two random variables V and V' we have[4]

$$(2) \qquad E[a+bU(V)]\left\{\substack{>\\=\\<}\right\}E[a+bU(V')] \Leftrightarrow E[U(V)]\left\{\substack{>\\=\\<}\right\}E[U(V')],\ b>0,$$

while a similar operation is impossible for arbitrary monotonic transformations of $U(.)$.

1. *The Approach of G. Cramer and D. Bernoulli*

1.1. *The Basic Idea*

The *expected-utility criterion* is formally similar to the mean-value criterion $R(V)=E(V)$. Indeed, Cramer and Bernoulli developed it from this criterion. In principle, they agreed that the preference functional should be the mathematical expectation of a value quantity. However, they argued that this value should be of a subjective rather than of an objective nature[5]. Thus they employed the index function $U(.)$ in the sense of a cardinal utility function for non-random wealth. A special version of this function favored by Bernoulli (p. 35) is $U(v)=ln\ v$, while Cramer (pp. 58–60) assumed alternately $U(v)=U(a+y)=\sqrt{y}$ and $U(v)=\min(y,y^*)$, where y^* is a saturation level of income. All three functions are concave, the first two strictly ($U''(v)<0$), and thus exhibit

---

[2] Cf. ch. I A 1.

[3] On the interval scale, equal utility steps can be determined. However, it is meaningless to relate two levels of utility to one another as is possible in the case of ratio scales that are defined up to the multiplication with a positive constant.

[4] The second line follows from the first by subtracting a and dividing through b.

[5] D. Bernoulli (1738, §3); G. Cramer (1728, §19). Laplace (1814, pp. XVI and 432–445) therefore uses the terms *esperance physique* as opposed to *esperance morale* and Allais (1952, pp. 271 ff.) refers to the *valeur monétaire* and the *valeur phychologique*.

the plausible property of diminishing marginal utility which later became popularly known as *Gossen's First Law*.

The concavity has a particular significance for the evaluation of risk[6]. As is well-known, the concepts 'concavity', 'convexity', and 'linearity' of a function are defined by comparing the function value associated with a linear combination of values of its argument with the linear combination of the corresponding function values. Thus, for a linear combination formed by applying the expectation operator to a given probability distribution, the following relationship holds:

$$(3) \qquad \left\{\begin{matrix}\text{concavity}\\ \text{linearity}\\ \text{convexity}\end{matrix}\right\} \Rightarrow E[U(V)]\left\{\begin{matrix}<\\=\\>\end{matrix}\right\}U[E(V)].$$

From this it is easy to conclude that a decision maker with a concave utility function should be willing to exchange a distribution of wealth levels for a non-random level of wealth the size of the expected value of the distribution. This interesting phenomenon can be elucidated even further by looking for the lowest non-random level of wealth the decision maker is willing to accept in exchange for the probability distribution. This certainty equivalent[7], $S(V)$, is defined by

$$U[S(V)]=E[U(V)]$$

and, after applying the inverse function[8] $U^{-1}(.)$, by

$$(4) \qquad S(V)=U^{-1}\{E[U(V)]\}.$$

As mentioned above[9], the difference between the expected value and the certainty equivalent is called the subjective price of risk:

$$(5) \qquad \pi(V)\equiv E(V)-S(V).$$

The subjective price of risk thus is that deduction from the expected value the decision maker is just willing to pay to have the dispersion completely eliminated. It is therefore suitable for distinguishing the

[6] That concavity, but not other particular aspects of the utility function, is relevant for risk aversion was perceived by MARSHALL (1920, p. 693, note IX (appendix)).

[7] Cf. the introduction to section A and SCHNEEWEISS (1967a, pp. 42–46).

[8] This operation requires that $U^{-1}(.)$ is continuous at $E[U(V)]$.

[9] Cf. the end of the introduction to section A.

various attitudes towards risk in accordance with the classification given with reference to the two-parametric criteria[10]:

$$(6) \qquad \pi(V)\{\gtreqless\}0 \Leftrightarrow \left\{\begin{array}{l}\text{risk aversion}\\ \text{risk neutrality}\\ \text{risk loving}\end{array}\right\}.$$

With the application of the inverse function to relationship (3) it can finally be stated[11] that

$$(7) \qquad \left\{\begin{array}{l}\text{concavity}\\ \text{linearity}\\ \text{convexity}\end{array}\right\} \Rightarrow \pi(V)\{\gtreqless\}0.$$

Thus, it is possible to derive from the hypothesis of diminishing marginal utility, which is plausible in itself, the hypothesis that people are willing to pay a price $\pi(V)$ for an elimination of risk. We now proceed to confront this hypothesis with reality by studying three simple examples.

1.2. *Examples*

The first example was of particular interest to Bernoulli, for Cramer it was the reason for formulating the expected-utility criterion. It is the determination of the maximum stake people are willing to put up in a gamble. Despite our initial doubts concerning the compatibility of the Supplement to the Axiom of Ordering with an evaluation of gambles, it is worth-while considering this example[12].

Both mathematicians tried to solve the so-called *St. Petersburg Paradox*. This 'paradox' is that no one could be found who was willing to risk his wealth, or even a considerable part of it, to participate in a particular gamble, although the mathematical expectation of its outcome was infinite. Whether the solution of Cramer and Bernoulli was really appropriate to the problem will be discussed later in connection with Arrow's Utility Boundedness Theorem. Here, however, we shall be content with finding out why the maximum stake (p_{max}) may be smaller than the expected prize.

---

[10] Cf. equations (A 1) and (A 2). There is a complete coincidence if $K_1 = E(V)$ or $K_1 = E(Y)$. For the criteria of Lange, Shackle, Krelle, and Schneider it is only possible to find a classification that is analogous but not identical.

[11] Note that, with a linear utility function, the expected-utility criterion coincides with the mean-value criterion.

[12] Cf. ch. I B 1.

The maximum stake is determined in a way that renders the decision maker indifferent between playing the game, where he pays the stake and receives the distribution of prizes X, and not playing at all. If, because the game only lasts a short time, we neglect interest payments, p_{max} is implicitly determined by the equation

$$(8) \qquad U(a)=E[U(a+X-p_{max})],$$

where a is the level of wealth without participation in the game. Applying the inverse function of $U(.)$ and subtracting $E(a+X-p_{max})$ we obtain

$$(9) \qquad a-E(a+X-p_{max})=-[E(a+X-p_{max})-S(a+X-p_{max})]$$

and, utilizing the definition (5), we have

$$(10) \qquad p_{max}=E(X)-\pi(a+X-p_{max}).$$

Because the concavity of the utility functions assumed by Cramer and Bernoulli implies $\pi>0$, this equation shows that $p_{max}<E(X)$.

The result may well explain why, for various oddly constructed gambles, no one is willing to put up a stake the size of the expected prize, but it does not describe the typical gambling situation. Unlike the game described above, the typical situation is characterized by gamblers who put up stakes higher than the expected prize; otherwise gambling casinos would not exist. The typical behavior of gamblers can be made compatible with the formal apparatus of expected-utility theory if a convex utility function is assumed and the plausible hypothesis of diminishing marginal utility is discarded. But this would not be the true explanation of gamblers' behavior. This behavior is better explained by the pleasure derived from aspects of gambling that cannot be seen simply by using information about the probability distribution of prizes[13].

The theory seems to fit the next two examples, taken from the insurance business, much better. Unlike in the first example, the problem of time now has to be taken into account. Without the passing of time, no loss can occur. It is therefore assumed that a premium is paid at the beginning of the period and that all indemnification payments are made at the end of the period. The initial wealth, net of the premium received or paid, is invested in the capital market at a non-random interest rate $q-1$.

---

[13] Cf. ch. III B 1.3.

We find first the minimum volume of premiums $p_{\min}$ that an insurance company would require for accepting the loss distribution C of its total stock of underwritten contracts[14]. If the company leaves the business, it achieves an end-of-period capital stock of size $v = aq$. If it stays in business, the end-of-period stock of capital is $V = aq + pq - C$ where p is the premium revenue. Thus the minimum premium revenue is determined by

$$(11) \qquad U(aq) = E[U(aq + p_{\min} q - C)],$$

so that, analogously to (10), the result

$$(12) \qquad p_{\min} q = E(C) + \pi(aq + p_{\min} q - C)$$

obtains. The minimum premium revenue, augmented by interest, consists of the expected loss $E(X)$ and a risk loading $\pi(.)$ that is positive in the case of a concave utility function. The result corresponds closely to the way insurance companies actually calculate.

What about the insurance purchaser? The question was asked by Bernoulli (1738, § 15[15]), but it was not until a century later that Barrois (1834, pp. 259–282, esp. pp. 260 f.) calculated the maximum premium from the point of view of the purchaser[16,17]. Without insurance, the potential purchaser faces the random end-of-period wealth $V = aq - C$, where C now indicates the *individual* random level of losses. If p denotes the *individual* premium paid, with insurance wealth becomes $aq - pq$. Thus, the contract is attractive from the point of view of the purchaser if $E[U(aq - C)] < U(aq - pq)$; the maximum premium, $p_{\max}$, is therefore determined by the condition

$$(13) \qquad E[U(aq - C)] = U(aq - p_{\max} q)$$

from which

$$(14) \qquad p_{\max} q = aq - S(aq - C)$$

---

[14] Cf. Bühlmann (1968, p. 268) and Helten (1973, pp. 208 f.).

[15] Bernoulli asked about the critical wealth level where the insurance purchaser is indifferent between buying insurance and bearing the risk himself. This is a meaningful question, since, as we shall see in ch. III A 2.3, $U(v) = \ln v$ implies that risk aversion is a decreasing function of wealth.

[16] Independently of Barrois, the problem was taken up by Mossin (1968). Cf. also Schneeweiss (1967a, p. 45) and Helten (1973, pp. 210 f.), and the remarks in ch. V C of this book.

[17] The approach of Barrois, who utilizes the logarithmic utility function, is here slightly generalized. Barrois's (p. 261) equation (3), incidentally, contains a mistake. The correct version of this equation is:

$$\Delta = F + A'S - (F+S)^{A'} F^{(1-A')}.$$

can be derived. If we now write

(15) $$p_{\max} q = aq - E(aq - C) + [E(aq - C) - S(aq - C)],$$

we get, because of (5),

(16) $$p_{\max} q = E(C) + \pi(aq - C).$$

This is the equation determining the maximum premium the insurance company can demand from a customer. With a concave utility function, therefore, the interest-augmented premium will exceed the expected idemnification payment. From the viewpoint of the insurance company this is a *conditio sine qua non*, since the expected indemnification payment for all underwritten contracts equals the sum of the single expectations just as the total premium income equals the sum of the single premiums.

It is useful to consider the case of insurance demand by inspecting the graph shown in Figure 10[18]. The figure refers to the simple case of a binary loss distribution with

$$C = \begin{pmatrix} w & 1-w \\ l & 0 \end{pmatrix},$$

where l and 0 are the variates of C, and w is the probability of loss. Since $E(V)$ is formed as a linear combination from aq and $aq - l$ with

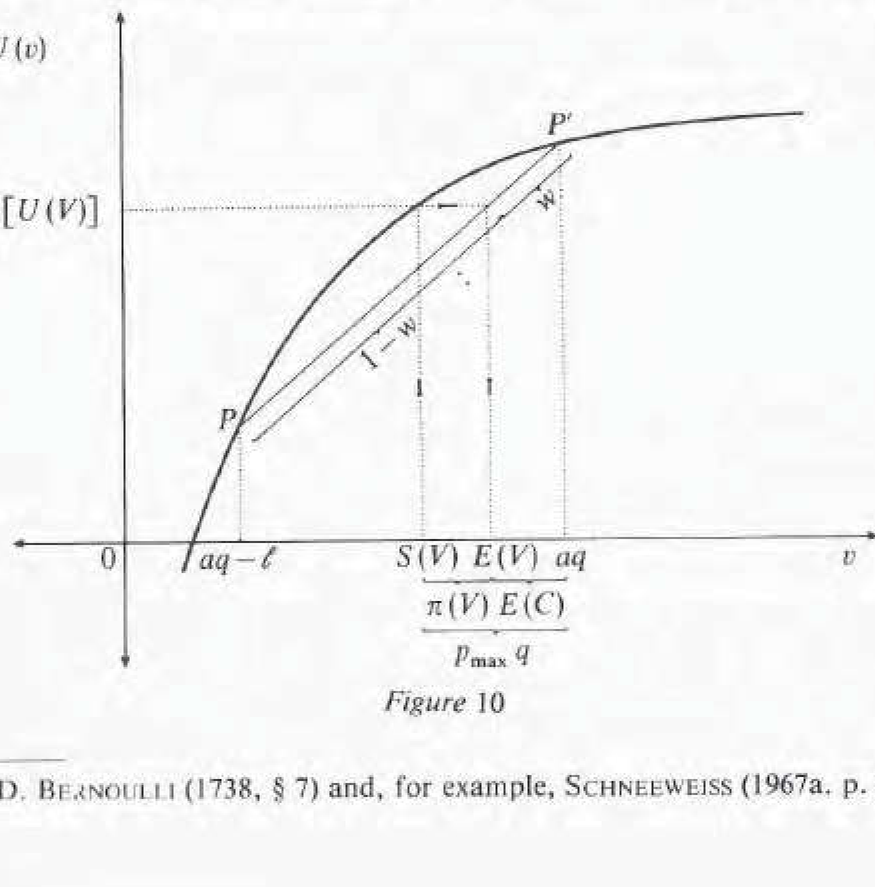


Figure 10

[18] Cf. D. BERNOULLI (1738, § 7) and, for example, SCHNEEWEISS (1967a, p. 62).

the same weights as those used to form $E[U(V)]$ from $U(aq)$ and $U(aq-l)$, the point with coordinates $(E[U(V)], E(V))$ is situated on the chord $\overline{PP'}$. To find the value of $S(V)$ given the value of $E(V)$, from $E(V)$ move vertically to the chord, from the chord horizontally to the utility curve, and from this curve back downwards to the abscissa. Obviously, for the example in this figure, there is a positive subjective price of risk and hence the maximum willingness to pay exceeds the expected indemnification payment. The reader may verify for himself the outcome of a linear or even convex utility curve.

1.3. *An Illustrative Measure of Risk Aversion: The Intensity of Insurance Demand*

Barrois's analysis of insurance demand gives a very graphic example of the possible applications of expected-utility theory. With reference to this example, we want to define a standardized measure of risk aversion that, in the course of this study, will be used a number of times for interpretation purposes. It is called the *intensity of insurance demand, g*. The intensity of insurance demand is the maximum interest-augmented willingness to pay for a full-coverage insurance contract divided by the expected level of loss[19]:

$$g \equiv g(aq-C) \equiv \frac{p_{\max} q}{E(C)} = \frac{E(C)+\pi(aq-C)}{E(C)}. \tag{17}$$

Because of (6), this measure satisfies the relationship

$$g \{ \gtreqless \} 1 \Leftrightarrow \left\{ \begin{array}{l} \text{risk aversion} \\ \text{risk neutrality} \\ \text{risk loving} \end{array} \right\}. \tag{18}$$

The intensity of insurance demand allows for a straightforward confrontation between the expected-utility approach and reality. Suppose an insurance company insures n persons each of whom brings the same loss distribution C and has the same maximum willingness to pay $p_{\max}$. Then $nE(C)$ is the expected value of total losses and, with independent risks and n sufficiently large, it also approximates the company's total sum of indemnification payments. The maximum aggregate premium volume is $np_{\max}$. Since the ratio between the actual sum of indemnifica-

[19] This measure resembles FISHER's (1906, p. 76) *coefficient of caution* that refers to the case of gambling and is the ratio of the maximum stake the decision maker is willing to put up and the expected prize. In the case of risk aversion the coefficient of caution has a value <1, in the case of risk neutrality a value $=1$, and in the case of risk loving a value >1.

tion payments and the actual total premium income is the *loss quota*, the inverse of the intensity of insurance demand $1/g = [nE(C)]/[nqp_{\max}]$ may be therefore interpreted as the minimum of the discounted loss quota. In general, empirical loss quotas significantly fall short of unity[20] and thus support the hypothesis $g>1$, which can be derived from the law of diminishing marginal utility by the use of the expected-utility theory.

This result puts the theory of Cramer and Bernoulli in a very favorable light. Nevertheless, it must defend itself against at least two objections, that have been raised in the course of scientific discussion. We shall see how successful it is in doing this.

1.4. *The Problem of Cardinal Utility*

It may well be doubted whether the cardinally measurable utility postulated by Cramer and Bernoulli exists at all. According to PARETO (1906, p. 169 f.), the assumption of cardinality belongs to metaphysics. The only thing certain, he said, is the fact that individuals are able to identify classes of equivalent commodity bundles and order these classes according to their desirability, i.e., that individuals have an *ordinal* utility function of income or wealth. This view, which incidentally had been clearly put forward by WUNDT (1863, esp. p. 26) for the psychology of stimulus sensations, became dominant among professional economists, particularly after the 'reconsideration' by HICKS and ALLEN (1934).

The reason is not that this view can be supported by clear-cut arguments, but that the theory of consumer choice, originally built on cardinal utility, could stand up to such a relaxation of assumptions[21,22]. It is true that there are critics like LITTLE (1950, pp. 14–52) who particularly dislike the indifference assumption and so prefer an even more general approach[23]. On the other hand, there are critics like FRISCH (1932), SCHULTZ (1933), ALT (1936), ALLAIS (1952, pp. 271, 273 f.), SCHNEEWEISS (1963), KRELLE (1968, pp. 10–12), and VAN PRAAG (1968, esp. pp. 6–10)

---

[20] Cf., for example, BUNDESAUFSICHTSAMT FÜR DAS VERSICHERUNGSWESEN (1974, pp. 96 ff. (appendix)).

[21] With respect to the cardinal utility functions or sensation functions postulated in psychophysics (cf. ch. III A) this point of view was also expressed by M. WEBER (1908, esp. pp. 389–392).

[22] The utilitarian welfare theory, however, did not survive this relaxation. Its recommendation that, in order to reach a welfare maximum, all income should be equalized, was based on the assumption of cardinal utility where even the unit is known. This probably did not bother the 'Fascist and critic of socialism and democracy', as Pareto was called by AMOROSO (1938).

[23] Cf. fn. 5 in ch. I A.

who find the approach too general. They all[24] plead for an interval scale of utility as required by the expected-utility rule. Indeed such a weakly cardinal function is not at all metaphysical. If individuals, as Pareto assumes, are able to draw their indifference curves, then they are revealing ratios of marginal utilities of different commodities. Why, then, should they not also be able to indicate the ratios of changes in utility from successive changes in the quantity of a single commodity or commodity bundle[25]? If they are, cardinality is ensured.

They are indeed able to do so, and can even do much more. That, at any rate, is the result of hundreds of series of experiments carried out by S.S. Stevens and co-workers at the Harvard Laboratory of Psychophysics[26]. According to these series of experiments, man can associate stimuli and sensations closely enough to construct even ratio scales.

Moreover, if this criticism were the only one raised against the expected-utility criterion, the cardinality of the utility function could be even more simply legitimated. Provided the preference functional is of the form (1) then, as shown by (2), the utility function must be measurable by an interval scale, as long as people can evaluate probability distributions in an ordinal way, as would be suggested by an analogous application of Pareto's postulates. Thus the cardinality of the utility function in itself cannot be subjected to significant criticism.

1.5. *Specific Risk Preference*

The weak point of the expected-utility rule lies elsewhere. Even if there is a cardinal utility function for non-random levels of wealth, there is no obvious reason for risk to be evaluated by using this function. This was vigorously pointed out by ALLAIS (1952 and 1953)[27]. Although two people have the same utility function for non-random wealth, they may well differ with regard to their *plaisir du risque*[28]. According to Allais, the expected-utility rule should therefore be corrected by introducing some measure of the dispersion of utilities. Unfortunately, Allais is not very specific about this[29]. Probably KRELLE's (1968, pp. 148–163) axiom system *B*, that explicitly refers to a parameter

---

[24] Alt does not want to be labelled as a protagonist of cardinality, but his axioms nevertheless lend strong support for it.

[25] Cf. KRELLE (1961, p. 140).

[26] A detailed overview is given in S.S. STEVENS (1975). Cf. also ch. III A of this book.

[27] Cf. also SCHNEEWEISS (1967a, p. 70), KRELLE (1968, p. 174), and HELTEN (1973, p. 197).

[28] ALLAIS (1952, pp. 130 f.); in contrast to the *plaisir du jeu*, the pure pleasure in participating in the game procedure.

[29] In one place, however, he mentions the second moment: ALLAIS (1953, p. 513).

measuring the dispersion of utilities, approaches what Allais had in mind. More elegant and, in content, not very different[30] is the way that Krelle (pp. 138–147) suggests with his axiom system A. There, $U(V)$ is split into a risk-preference function $\varphi(.)$ and a utility function $u(.)$ for non-random wealth such that

$$(19) \qquad U(v) = \varphi[u(v)].$$

In a similar way as was shown above for $U(.)$, concavity, linearity, and convexity of $\varphi(.)$ indicate a love of, indifference to, and aversion to dispersions in utility.

The intermediate case of linearity Krelle calls the 'normal' case. Since, when $\varphi(.)$ is linear, the concavity of $u(.)$ is sufficient to produce risk aversion, the choice of this name suggests a good deal of relevance for the expected-utility theory. But unless hypotheses are available to legitimate the assertion that the behavior defined as 'normal' is normal from an empirical point of view, (19) leads to an elimination of the expected-utility theory in the form used by Cramer and Bernoulli. The derivation of risk aversion from the hypothesis of diminishing marginal utility that at first glance seemed so plausible, loses much of its force.

This is a pity, particularly since up to now no other explanation of risk aversion has been offered. We therefore have no choice but to accept risk aversion as an empirically observable fact and to forgo its explanation. What remains from the proposal of Cramer and Bernoulli is the idea as such that the preference functional should be the expected value of an index function $U(.)$ however it is constructed and whatever its meaning is. *A priori*, this idea does not seem very reasonable. A legitimation similar to the one that could be given for the two-parametric criteria in the case of linear distribution classes is impossible. There is, however, an argument that throws a whole new light on the preference functional $E[U(V)]$ that over the years had become rather dusty. This argument is considered in the following section.

2. *The von Neumann-Morgenstern Index*

Beginning with an approach different from that of Cramer and Bernoulli, von Neumann and Morgenstern (1947) also developed a kind of expected-utility rule. However, the utility they consider has little in common with utility as a measure of the intensity of satisfaction, and it could be argued therefore that another word should be used. However,

---

[30] Cf. Krelle (1968, pp. 161–163).

it is common usage to denote the *von Neumann-Morgenstern index* by the term *expected utility*, so the name will also be used here.

2.1. *The Axioms*

The special characteristic of the von Neumann–Morgenstern approach is to derive the preference functional $R(V)=E[U(V)]$ from a few axioms postulating rational behavior[31]. The presentation of axioms is always largely a matter of taste. Thus it is not surprising that the axioms originally presented by von Neumann and Morgenstern, in the course of time, underwent considerable alteration.

The decisive step was made at the beginning of the fifties, with the introduction of the Axiom of Independence which, as we saw, also plays the crucial role in the foundation of the Principle of Insufficient Reason. It was such a big step that even SAMUELSON (1952a, p. 147) did not understand the relationship with the original axioms: 'Quelque mathématicien devrait éclairer tout cela.' The mathematician was soon found. It was MALINVAUD (1952).

A simple axiom system that leads to the expected-utility rule is obtained as soon as two further axioms are added to the two introduced in chapter one. These are the Archimedes Axiom and the Axiom of Non-Saturation[32]. For the sake of clarity, the complete axiom system is presented here. Deviating from the original formulation, we assume equivalent objective probabilities for the first two axioms, thus taking into account the result of the first chapter.

(1) Axiom of Ordering: *The decision maker has a complete weak ordering of all attainable probability distributions of end-of-period wealth.*

(2) Axiom of Independence: *Suppose for two probability distribution* e_1 *and* e_2 *it holds that*

$$e_1\{\succsim\}e_2.$$

Then it follows that distributions, built up by combining e_1 *and* e_2 *with another distribution* e_3*, satisfy*

$$\begin{pmatrix} w & 1-w \\ e_1 & e_3 \end{pmatrix}\{\succsim\}\begin{pmatrix} w & 1-w \\ e_2 & e_3 \end{pmatrix},$$

if $0<w\leq 1$.

---

31 VON NEUMANN and MORGENSTERN (1947, pp. 26–29, 617–632).

32 Cf. ch. I A 1, I B 1, and I B 3.1.2. A discussion of various axiom systems can be found in MARKOWITZ (1970, pp. 228–242) and KRELLE (1968, pp. 121–195). Our axiom system resembles that presented by FRIEDMAN and SAVAGE (1952, pp. 464–469).

(3) Archimedes Axiom: *Let there be three variates of wealth* $v_1 < v < v_2$. *Then there is one and only one probability* w, $0 < w < 1$, *such that*

$$v \sim \begin{pmatrix} w & 1-w \\ v_2 & v_1 \end{pmatrix}.$$

(4) Axiom of Non-Saturation: *If* $v_2 > v_1$, *then* $v_2 \succ v_1$.

We forgo a discussion of axioms (1) and (2) since they are known already. The most important aspect of the Archimedes Axiom[33] is the exclusion of lexicographically ordered ranges of wealth, an aspect that provoked vigorous criticism by GEORGESCU-ROEGEN (1954, esp. p. 525). If, for example, there is a lexicographic critical level of wealth $\tilde{v}$, $v_1 < \tilde{v} < v_2$, then for each probability in the range $0 < w < 1$ we have

$$v \left\{ \begin{smallmatrix} \prec \\ \succ \end{smallmatrix} \right\} \begin{pmatrix} w & 1-w \\ v_2 & v_1 \end{pmatrix} \Leftrightarrow v \left\{ \begin{smallmatrix} < \\ > \end{smallmatrix} \right\} \tilde{v}. \tag{20}$$

Thus there is no probability in the open unit interval that is able to produce the indifference required by the Archimedes Axiom. In the light of the doubts concerning the validity of the lexicographic criterion remaining after the discussion in section B, we should not place too much weight on this criticism in the case of careful economic decision making. If, however, despite these doubts, there is a lexicographic critical level of wealth, then the axioms given above have to be restricted to probability distributions that do not extend beyond this level of wealth. The remaining Axiom of Non-Saturation is, at least for wealth levels occurring in the real world, self-evident. If it were not true, then people would not mind being robbed.

2.2. *The Derivation of the Expected-Utility Rule from the Axioms*

It is now shown that the four axioms introduced above imply the expected-utility rule.

---

33 Reference is often made to this axiom under the name of 'Continuity Axiom' that was first used by MARSCHAK (1950, p. 117). If we consider the indifference probability w as a function of the type $h(v)$ then this name suggests that $h(v)$ has to be continuous. This, however, is unnecessary. For example, the function

$$h(v) = \begin{cases} 0.1 \text{ if } v < v^* \\ 0.5 \text{ if } v = v^* \\ 0.9 \text{ if } v > v^* \end{cases}$$

is not continuous although, for any v, it gives a unique $h(.)$ as required by the Archimedes Axiom.

Step 1: Assessment of the Indifference Probability

First we define two wealth levels $v_{\min}$ and $v_{\max}$ that are chosen generously enough to ensure that all distributions to be evaluated fall into the open interval they limit. Then, by the use of the Archimedes Axiom, an indifference probability $h(v)$ is assessed for all v in this interval. Analogously to the formulation of the axiom, $h(v)$ is implicitly defined by

$$(21) \qquad v \sim \begin{pmatrix} h(v) & 1-h(v) \\ v_{\max} & v_{\min} \end{pmatrix}$$

or, in other words, by

$$(22) \qquad \begin{pmatrix} h(v) & 1-h(v) \\ v & v \end{pmatrix} \sim \begin{pmatrix} h(v) & 1-h(v) \\ v_{\max} & v_{\min} \end{pmatrix}.$$

Figure 11 shows an example of the shape of the function $h(v)$. Note that

$$(23) \qquad v = \begin{Bmatrix} v_{\max} \\ v_{\min} \end{Bmatrix} \Rightarrow h(v) = \begin{Bmatrix} 1 \\ 0 \end{Bmatrix}.$$

This relationship originates from the fact that, according to axiom (4), we have

$$v_{\max} > v_{\min} \Rightarrow v_{\max} \succ v_{\min}$$

and that, because of axiom (2), in the case $v = v_{\max}$ and $h(v) < 1$ the right side of (22) would be worse and in the case $v = v_{\min}$ and $h(v) > 0$ better than the left side. We leave it open for a moment whether or not $h(v)$ is monotonically increasing.

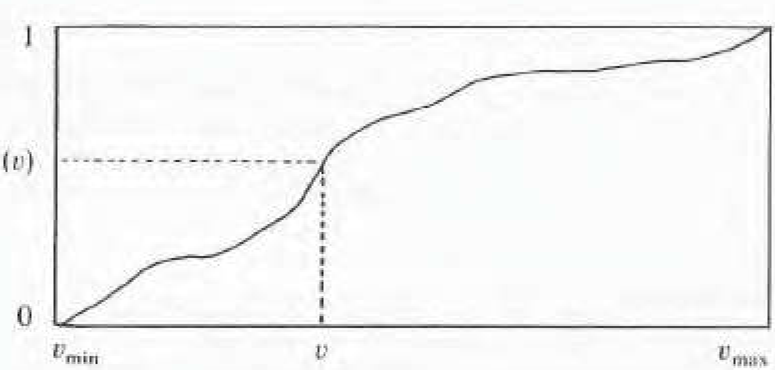


Figure 11

Step 2: Transformation of Chances
Consider one of the probability distributions to be evaluated,

$$(24)\qquad \begin{pmatrix} w_1 & w_2 \dots w_n \\ v_1 & v_2 \dots v_n \end{pmatrix},$$

and express it in the form

$$(25)\qquad \begin{bmatrix} w & 1-w \\ \\ e_1 & e_3 \end{bmatrix} = \begin{bmatrix} w_1 & 1-w_1 \\ & \begin{pmatrix} \frac{w_2}{1-w_1} \dots \frac{w_n}{1-w_1} \\ v_2 \quad \dots \quad v_n \end{pmatrix} \\ v_1 & \end{bmatrix}$$

in order to achieve the formulation of the Independence Axiom. Then, by the use of this axiom, the degenerated subdistribution $e_1 = v_1$ is replaced by the binary distribution

$$(26)\qquad e_2 \equiv \begin{pmatrix} h(v_1) & 1-h(v_1) \\ v_{\max} & v_{\min} \end{pmatrix}$$

that is equivalent according to step 1. Thereby the probability distribution (24) takes on the shape

$$(27)\qquad \begin{bmatrix} w_1 & w_2 \dots w_n \\ \begin{pmatrix} h(v_1) & 1-h(v_1) \\ v_{\max} & v_{\min} \end{pmatrix} & v_2 \dots v_n \end{bmatrix}.$$

Maintaining this transformation we now, in an analogous way, replace $v_2, v_3, \dots, v_n$ step by step by equivalent binary distributions similar to (26). Thus, the probability distribution finally becomes

$$(28)\qquad \begin{bmatrix} w_1 & \dots & w_n \\ \begin{pmatrix} h(v_1) & 1-h(v_1) \\ v_{\max} & v_{\min} \end{pmatrix} & \dots & \begin{pmatrix} h(v_n) & 1-h(v_n) \\ v_{\max} & v_{\min} \end{pmatrix} \end{bmatrix}$$

$$= \begin{bmatrix} \sum_{j=1}^{n} w_j h(v_j) & \sum_{j=1}^{n} w_j [1-h(v_j)] \\ \\ v_{\max} & v_{\min} \end{bmatrix}.$$

The procedure can easily be illustrated if the graph of the initial probability distribution (24) is added to the diagram of Figure 11. For example, for a distribution with three variates Figure 12 is obtained. The columns over v_1, v_2, and v_3, depicted in this figure, represent the corresponding probabilities w_1, w_2, and w_3. Each of these columns is divided in the same proportion as the curve $h(v)$ divides the distance between the upper and the lower bound of the figure for the corresponding level of v. Step by step, the lower parts of these columns are shifted to the wealth level v_{max} and the upper ones to v_{min}. In this way, a binary distribution, as represented by the columns over v_{max} and v_{min}, is constructed that is equivalent to the initial distribution.

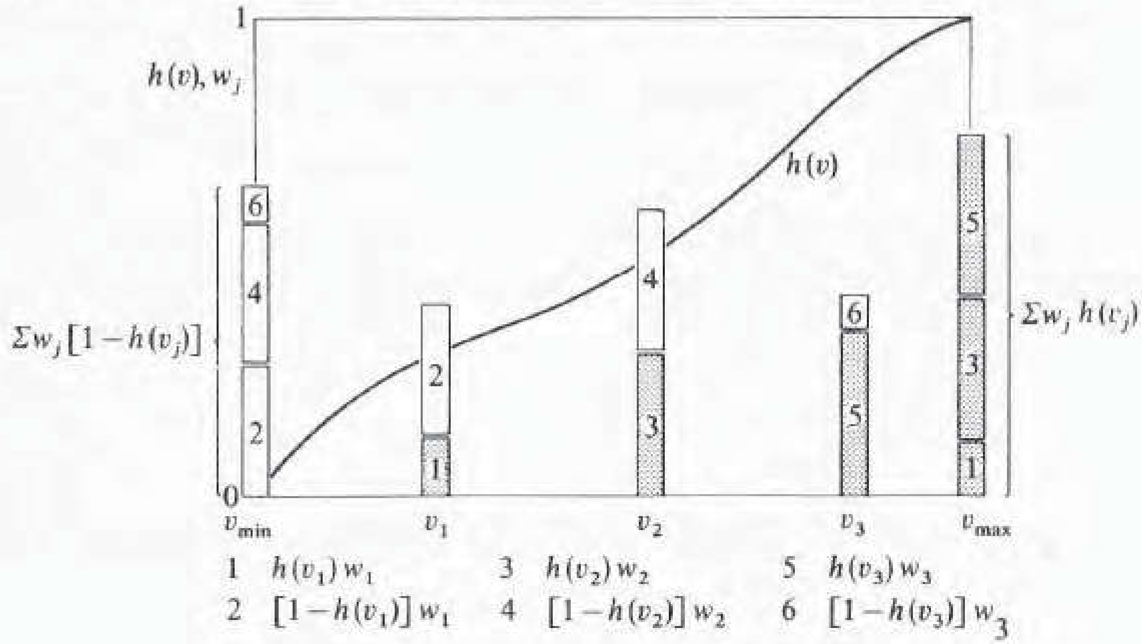


Figure 12

In the way just described all probability distributions from the decision maker's opportunity set can be transformed into equivalent binary distributions with the variates v_{max} and v_{min}. It seems wise to choose that distribution for which the probability

$$\sum_{j=1}^{n} w_j h(v_j)$$

of the occurrence of the variate v_{max} is maximal. But it has not yet been proved that this choice follows from the axioms introduced above.

Step 3: The Comparison of Binary Distributions

Suppose there are two distributions

$$e \equiv \begin{pmatrix} w & 1-w \\ v_{max} & v_{min} \end{pmatrix} \text{ and } e' \equiv \begin{pmatrix} w' & 1-w' \\ v_{max} & v_{min} \end{pmatrix}, \tag{29}$$

where $w' > w$,

to be compared. Defining the probability

$$(30)\qquad w'' \equiv \frac{w' - w}{1 - w}$$

we can write the two distributions in the form

$$(31)\qquad e = \begin{bmatrix} w'' & 1-w'' \\ \begin{pmatrix} w & 1-w \\ v_{\max} & v_{\min} \end{pmatrix} & \begin{pmatrix} w & 1-w \\ v_{\max} & v_{\min} \end{pmatrix} \end{bmatrix}$$

and

$$e' = \begin{bmatrix} w'' & 1-w'' \\ v_{\max} & \begin{pmatrix} w & 1-w \\ v_{\max} & v_{\min} \end{pmatrix} \end{bmatrix}.$$

By direct application of the Independence Axiom (2) the following relationships are obtained:

$$(32)\qquad e\left\{\begin{smallmatrix} > \\ < \end{smallmatrix}\right\}e' \Leftrightarrow \begin{pmatrix} w & 1-w \\ v_{\max} & v_{\min} \end{pmatrix}\left\{\begin{smallmatrix} > \\ < \end{smallmatrix}\right\} v_{\max}$$

$$\Leftrightarrow \begin{pmatrix} w & 1-w \\ v_{\max} & v_{\min} \end{pmatrix}\left\{\begin{smallmatrix} > \\ < \end{smallmatrix}\right\}\begin{pmatrix} w & 1-w \\ v_{\max} & v_{\max} \end{pmatrix}$$

$$\Leftrightarrow v_{\min}\left\{\begin{smallmatrix} > \\ < \end{smallmatrix}\right\} v_{\max}.$$

According to this formulation e' is better than e if, and only if, the non-random wealth level $v_{\max}$ is preferred to the smaller, also non-random, wealth level $v_{\min}$. As required by the Axiom of Non-Saturation this is the case.

Result

Thus it has been shown that, from any pair of distributions, the one with the higher value of $\sum_{j=1}^{n} w_j h(v_j)$ is to be preferred. The preference functional therefore is

$$(33)\qquad R(V) = \sum_{j=1}^{n} w_j h(v_j)$$

$$= E[U(V)]$$

where the 'indifference function' $h(v)$ turns out to be the utility function $U(v)$.

The result in turn allows the initial question of whether $h(v)=U(v)$ is monotonically increasing to be answered. The answer is in the affirmative. Suppose that, on the contrary, for two non-random levels of wealth v and v' we have $v'>v$ and because of $U(v')<U(v)$ at the same time $v'<v$. Then, obviously, there is a contradiction with the Axiom of Non-Saturation.

Section D
Comparison of Preference Functionals

1. *Expected Utility versus Lexicographic Preference: The Decision for a Decision Criterion*

Thanks to its axiomatic foundation, the expected-utility criterion plays a dominant role among the decision criteria discussed. Following SCHNEEWEISS (1967a, p. 78) it could therefore be called a 'quasi-logical principle'. This sounds favorable, perhaps a bit too favorable, because from the lexicographic side the question is asked[1]: 'Is it though the greatest of all irrationalities to assume that any given individual, be he a cardinalist, is ex definitione rational in the above sense?'

On the other hand, the alternative of a preference structure based on aspiration levels and saturation probabilities offered by lexicographic theory is not very convincing. As long as this preference structure is interpreted as being derived from an underlying substitutive ordering for the sake of simplifying short-run decision making it certainly has its merits. But as a guide for weighty economic decisions it is not acceptable. Nevertheless, the lexicographic theory in its simplest version with a critical wealth level below which there is the absolute disaster, cannot be altogether rejected. Such a level which, if it exists at all, was shown to be at $\bar{v}=0$ would of course have some bearing on careful decision making.

Thus, since the level of disaster is incompatible with the Archimedes Axiom, Schneeweiß's pink champagne seems to acquire an aftertaste of bitters. Fortunately ARROW (1951, p. 29) and ROY (1952, pp. 432 f.) save the situation. If the utility function has the shape (cf. Fig. 13)

$$(1) \qquad U(v)=\begin{Bmatrix}1, & v\geq\bar{v}\\ 0, & v<\bar{v}\end{Bmatrix},$$

[1] GEORGESCU-ROEGEN (1954, p. 505).

then a maximization of expected utility is identical with a maximization of the probability of survival, for in the present case

$$(2) \qquad R(V) = E[U(V)] = \int_{-\infty}^{+\infty} U(v) f(v) dv$$

$$= \int_{\tilde{v}}^{-\infty} f(v) dv$$

$$= W(v \geq \tilde{v}).$$

The surprising result, therefore, is that the expected-utility criterion is compatible with the aim of maximizing the probability of survival. The only aspect that may be disturbing is that the utility function described by (1) contradicts the Axiom of Non-Saturation and the Archimedes Axiom.

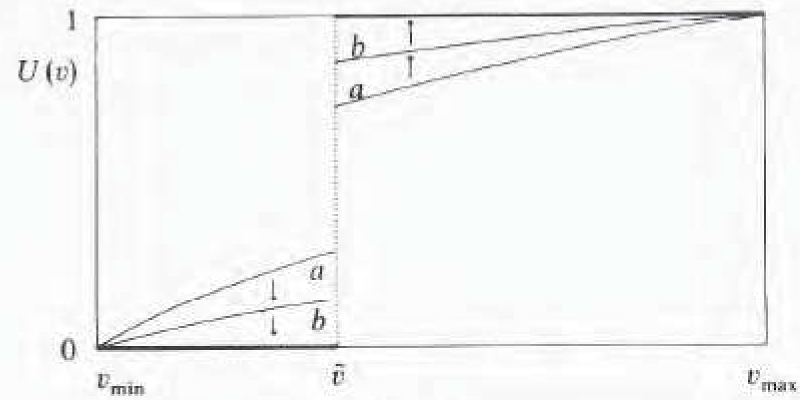


Figure 13

The contradiction, however, arises only from mathematical sophistication and has no real meaning. Consider the curve a in Figure 13 that shows the utility function $U(v)$ where the left part is valid for $v < \tilde{v}$ and the right part for $v \geq \tilde{v}$. Because of

$$0 < U(v) < 1 \Leftrightarrow v_{\min} < v < v_{\max}$$

and the strictly positive slope, this curve is compatible with the axioms of expected utility[2]. The corresponding preference functional can be written as[3]

$$(3) \qquad R(V) = E[U(V)]$$

$$= W(v < \tilde{v}) E[U(V)] \,|_{v < \tilde{v}} + W(v \geq \tilde{v}) \mathrm{E}[U(V)] \,|_{v \geq \tilde{v}}.$$

---

[2] Cf. fn. 33 in section C.

[3] $E(X)|_y$ means: 'expected value of X conditional on the event y'.

Now suppose, by a suitable transformation of $U(v)$, curve a is shifted, via position b, towards the thick curve that coincides with the lower boundary of Figure 13 for $v<\tilde{v}$ and with the upper one for $v>\tilde{v}$. Then $E[U(V)]\,|_{v<\tilde{v}}\to 0$, since $U(v)\to 0$ for $v<\tilde{v}$, and $E[U(V)]\,|_{v\geq\tilde{v}}\to 1$, since $U(v)\to 1$ for $v\geq\tilde{v}$. Accordingly, in the limit, we have for (3):

$$(4)\qquad R(V)\to W(v\geq\tilde{v}).$$

Thus the lexicographic aim of maximizing the probability of survival can be approximated as closely as we wish by a shape of $U(v)$ that satisfies the Archimedes Axiom and the Axiom of Non-Saturation.

These considerations show up another feat of the expected-utility approach. We have seen that, by a suitable choice of $U(v)$, it is possible to depict the lexicographical aim of survival. The question remains, however, how to discriminate between two distributions with equal probability of survival. The lexicographic approach decides this question by referring us to an additional, lower-ranking, aim. If attention is limited to pecuniary problems this aim could, for example, be the maximization of expected utility of wealth by the use of a given continuous utility function $\tilde{U}(v)$. At first glance the expected-utility approach as we have come to know it seems unable to handle such a two-dimensional aim. Searching for a suitable shape of the utility function $U(v)$, the first idea that occurs is either to choose $U(v)=\tilde{U}(v)$, so that only the lower-ranking aim is taken into account, or to choose

$$U(v)=U^*(v)\equiv\begin{Bmatrix}1, & v<\tilde{v}\\ 0, & v\geq\tilde{v}\end{Bmatrix},$$

so that, according to (1), only the predominant aim is depicted. There is, however, also the possibility

$$(5)\qquad U(v)=\begin{Bmatrix}\lambda\tilde{U}(v), & v<\tilde{v}\\ 1-\lambda[1-\tilde{U}(v)], & v\geq\tilde{v}\end{Bmatrix},\quad 0<\lambda\leq 1,$$

where the function $\tilde{U}(v)$ is assumed to be standardized so that $\tilde{U}(v_{\min})=0$ and $\tilde{U}(v_{\max})=1$. It can be shown that, with the aid of the function $U(v)$ described in (5), it is possible to approximate as closely as we wish the two-dimensional ordering with the predominant aim max $W(v\geq\tilde{v})$ and the lower-ranking aim $\max E[\tilde{U}(V)]$, if λ is chosen sufficiently close to zero.

Concerning the predominant aim, this contention is proved by (4), since, figuratively speaking, $U(v)$ is constructed from $\tilde{U}(v)$ by cutting the latter in two at $v=\tilde{v}$ and, with $\lambda\to 0$, shifting the left part towards the lower and the right part towards the upper boundary in Figure 13. The shape of $\tilde{U}(v)$ does not matter for this result.

Suppose now that the predominant aim does not allow a discrimination to be made between the two distributions V' and V'' since they are both characterized by the same survival probability. In this case the lower-ranking aim would have to determine choice. The function $U(v)$ would have to represent $\tilde{U}(v)$ so that, in the case $W(v'\geq\bar{v})=W(v''\geq\bar{v})$, the relationship

$$(6)\qquad E[U(V')]\{\gtreqless\}E[U(V'')]\Leftrightarrow E[\tilde{U}(V')]\{\gtreqless\}E[\tilde{U}(V'')]$$

will hold, either exactly for arbitrary values of λ or, at least approximately, for sufficiently small values of λ. The former is the case. This can easily be shown by the following chain of identical transformations:

$$
\begin{aligned}
(7)\qquad & E[U(V')]\{\gtreqless\}E[U(V'')]\\
\Leftrightarrow\quad & W(v'<\bar{v})E[U(V')]\,|_{v'<\bar{v}}+W(v'\geq\bar{v})E[U(V')]\,|_{v'\geq\bar{v}}\\
& \{\gtreqless\}\,W(v''<\bar{v})E[U(V'')]\,|_{v''<\bar{v}}+W(v''\geq\bar{v})E[U(V'')]\,|_{v''\geq\bar{v}}\\
\Leftrightarrow\quad & W(v'<\bar{v})E[\lambda\tilde{U}(V')]\,|_{v'<\bar{v}}+W(v'\geq\bar{v})E[1-\lambda[1-\tilde{U}(V')]]\,|_{v'\geq\bar{v}}\\
& \{\gtreqless\}\,W(v''<\bar{v})E[\lambda\tilde{U}(V'')]\,|_{v''<\bar{v}}+W(v''\geq\bar{v})E[1-\lambda[1-\tilde{U}(V'')]]\,|_{v''\geq\bar{v}}\\
\Leftrightarrow\quad & W(v'<\bar{v})E[\tilde{U}(V')]\,|_{v'<\bar{v}}+W(v'\geq\bar{v})E[\tilde{U}(V')]\,|_{v'\geq\bar{v}}\\
& \{\gtreqless\}\,W(v''<\bar{v})E[\tilde{U}(V'')]\,|_{v''<\bar{v}}+W(v''\geq\bar{v})E[\tilde{U}(V'')]\,|_{v''\geq\bar{v}}\\
\Leftrightarrow\quad & E[\tilde{U}(V')]\{\gtreqless\}E[\tilde{U}(V'')].
\end{aligned}
$$

Thus it has been shown that, by a suitable choice of λ, the two-part utility function $U(v)$ as defined in (5) can be made to approximate the two-dimensional lexicographic preference structure as closely as we wish[4].

It is important to recognize that, with $\lambda\to 0$, the two-part utility function is able to incorporate the predominant and the lower-ranking aim at the same time. For arbitrary $\lambda>0$, and thus during the whole transition process towards the limit $\lambda=0$, the separated curve $U(v)$ ranks the two distributions V' and V'' in the same way as the original function $\tilde{U}(v)$ does, provided the distributions do not differ with respect to their probabilities of survival. For $\lambda\to 0$ the function $U(v)$, however,

[4] The aim could be generalized by taking the maximization of a weighted average of conditional expected utilities as the lower-ranking aim:

$$\max\{\alpha E[U(V)]\,|_{v<\bar{v}}+(1-\alpha)E[U(V)]\,|_{v\geq\bar{v}}\}.$$

has the additional property of leading to a better evaluation of the distribution with the higher survival probability irrespective of how $\tilde{U}(v)$ and the two probability distributions are shaped.

From a normative point of view, the lexicographic objection to the Archimedes Axiom was the only substantial criticism that survived the scientific discussion of the von Neumann–Morgenstern index. With the preceding demonstration of the flexibility of this index, this criticism also loses much of its force. We thus should accept the expected–utility criterion as a guide to wise action.

We could finish chapter two at this point and continue with the expected-utility criterion. Unfortunately, there is another aspect that has not yet been considered; it is how easy the preference functionals are to handle in theoretical and practical analysis. With regard to this aspect, the expected-utility criterion seems to be among the worst of the criteria considered. Its very flexibility makes it difficult to use.

For this reason, the question of whether the expected-utility approach is compatible with the various two-parametric approaches arises. Perhaps a sufficient compatibility is somewhere to be found that will allow us to choose a simpler decision criterion if we need one.

2. *Expected Utility and the Two-Parametric Substitutive Criteria: Searching for an Operational Alternative*

2.1. *Common Preference Structures*

This section compares the expected-utility criterion with some of the two-parametric substitutive criteria and attempts to answer a *rigorous* question[5]. Suppose no limitation of the classes of probability distributions in the decision maker's opportunity set is possible. Are there preference structures such that arbitrarily chosen probability distributions are ranked by the two-parametric criterion in the same order as by the expected-utility criterion? Or shorter: Which intersection of preference structures do the criteria have in common?

The search is hopeless with the criteria of Lange and Shackle, for they both neglect much of the information contained in a probability distribution. Thus, right from the beginning, we need only consider the other criteria[6].

---

[5] Cf. SCHNEEWEISS (1967a, pp. 89–117, ch. III) and MARKOWITZ (1970, pp. 287–294).

[6] SCHNEEWEISS (1967a, pp. 103–111) shows that preference structures the depend on ordinal parameters are incompatible with $R(V) = E[U(V)]$.

2.1.1. The Domar–Musgrave Criterion

$$R(V) = U\left[E(Y),\ -\int_{-\infty}^{0} y f(y+a)\,dy\right]$$

The Domar–Musgrave criterion that can be written as

$$R(V) = U(K_1, K_2) \text{ with } K_1 \equiv W(y \geq 0)E(\overset{+}{Y}) - W(y<0)E(\bar{Y})$$
$$\text{and } K_2 \equiv W(y<0)E(\bar{Y})$$

implies, as shown by RICHTER (1959/60, pp. 155–157), a utility curve that is composed of two linear parts:

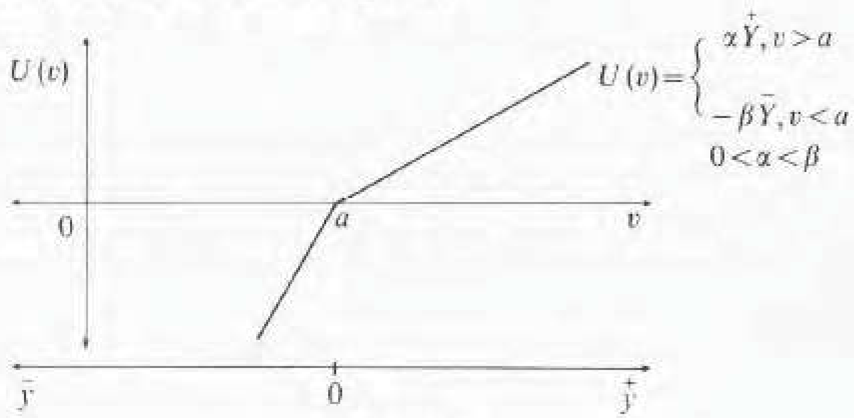


Figure 14

The reason is that dispersions of gains and losses around the given mean values $E(\bar{Y})$ and $E(\overset{+}{Y})$ must not influence the value of the preference functional, a requirement that holds only with linear utility. The preference functional can therefore be simplified to

(8) $$R(V) = E[U(V)] = W(y \geq 0)\,\alpha E(\overset{+}{Y}) - W(y<0)\,\beta E(\bar{Y})$$
$$= \alpha K_1 - (\beta - \alpha)K_2,$$

where α and β are some positive constants. As a first approximation, the shape of the utility function depicted in Figure 14 is not implausible. The concavity still ensures risk aversion. Less attractive is the implication that, as shown in Figure 15, the corresponding indifference curves in the (K_1, K_2) diagram are linear:

(9) $$\frac{dK_1}{dK_2} = \frac{\beta - \alpha}{\alpha} = \text{const.}$$

For problems of taxation in particular, for which Domar and Musgrave constructed their preference functional, this aspect is fatal. The substitution effects resulting from taxation fail to occur.

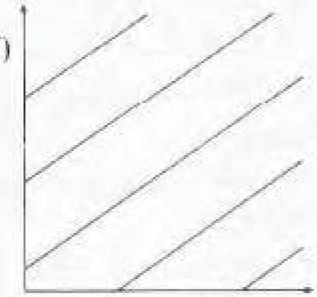


Figure 15

2.1.2. The Criterion of Krelle and Schneider

$$R(V) = U(\overset{+}{y}^*, \bar{y}^*)$$

The Krelle-Schneider criterion is general enough to include all kinds of preference structures. Indeed, it is the decision maker himself who constructs the equivalent gains and losses. With respect to its handling, however, in its general form this criterion is no better than the expected-utility criterion.

Nevertheless, Schneider succeeds in applying his criterion to the analysis of tax effects by introducing a supplementary assumption. This assumption is that an income tax at the rate t with no loss offset reduces the equivalent net gains to $1-t$ times the equivalent gross gains for all projects in the opportunity set. An analogous assumption is made concerning the relationship between equivalent net and gross losses when government bears the share z of the losses. An examination will be made of whether there are von Neumann-Morgenstern utility functions compatible with this *operationalized version* of the Krelle-Schneider criterion.

For this purpose the shape of the utility function above the gain axis is denoted by $\overset{+}{U}(\overset{+}{y})$ and above the loss axis by $\bar{U}(\bar{y})$ so that

$$\text{(10)} \qquad \overset{+}{U}(\overset{+}{y}) \equiv U(y) \quad \text{and} \quad \overset{+}{y} \equiv y, \quad \text{if} \quad y \geq 0,$$

and

$$\bar{U}(\bar{y}) \equiv -U(y) \quad \text{and} \quad \bar{y} \equiv -y, \quad \text{if} \quad y \leq 0,$$

where y is the period income. We assume that $U(y)$ is continuous at $y=0$ so that $\overset{+}{U}(0) = \bar{U}(0)$. If this assumption were not satisfied the notion of the nullchance would not make sense. Since the utility function $U(.)$ is only defined up to a strictly positive linear transformation, we may arbitrarily set

$$\text{(11)} \qquad \overset{+}{U}(0) = \bar{U}(0) = U(0) = 0.$$

Now consider Schneider's transformation procedure, given a particular distribution Y that is to be evaluated. According to the expected-utility rule it holds that

$$U(\overset{+}{y}{}^*, \bar{y}^*) = W(y \geq 0)\,\mathrm{E}[\overset{+}{U}(\overset{+}{Y})] - \mathrm{W}(y < 0)\,E[\bar{U}(\bar{Y})]. \tag{12}$$

If $W(y \geq 0) > \overset{+}{w}$, then, installing the nullchance, we have[7]

$$U(\overset{+}{y}{}^*, \bar{y}^*) = \overset{+}{w}\,\overset{+}{U}(\overset{+}{y}{}^*) + [W(y \geq 0) - \overset{+}{w}]\,U(0) - W(y < 0)\,E[\bar{U}(\bar{Y})] \tag{13}$$

and, after transferring the excess probability $W(y \geq 0) - \overset{+}{w}$ to the loss axis,

$$U(\overset{+}{y}{}^*, \bar{y}^*) = \overset{+}{w}\,\overset{+}{U}(\overset{+}{y}{}^*) - \bar{w}\,\bar{U}(\bar{y}^*). \tag{14}$$

If, however, $W(y < 0) > \bar{w}$, we first have

$$U(\overset{+}{y}{}^*, \bar{y}^*) = W(y \geq 0)\,E[\overset{+}{U}(\overset{+}{Y})] + [W(y < 0) - \bar{w}]\,U(0) - \bar{w}\,\bar{U}(\bar{y}^*) \tag{15}$$

and then again (14). By using (11) and setting (14) equal to (15), or (12) equal to (13), we now find

$$\overset{+}{w}\,\overset{+}{U}(\overset{+}{y}{}^*) = W(y \geq 0)\,E[\overset{+}{U}(\overset{+}{Y})] \tag{16}$$

and hence the following equation for the equivalent gain:

$$\overset{+}{y}{}^* = \overset{+}{U}{}^{-1}\left\{\frac{1}{\overset{+}{w}} W(y \geq 0)\,E[\overset{+}{U}(\overset{+}{Y})]\right\}. \tag{17}$$

Analogously we get

$$\bar{y}^* = \bar{U}^{-1}\left\{\frac{1}{\bar{w}} W(y < 0)\,E[\bar{U}(\bar{Y})]\right\} \tag{18}$$

from (13) and (14) or from (12) and (15).

---

[7] The first step of the transformation procedure, i.e., the replacement of the distributions of gains and losses by their certainty equivalents, is not of interest in the present context. For the definitions of $\overset{+}{y}{}^*$, $\bar{y}^*$, $\overset{+}{w}$, and $\bar{w}$ cf. section A 5.2.

Schneider contends that

$$(19)\qquad (1-t)\overset{+}{y}{}^{*}=\overset{+}{U}{}^{-1}\left\{\frac{1}{\overset{+}{w}}W(y\geq 0)E[\overset{+}{U}((1-t)\overset{+}{Y})]\right\},$$

$$(1-z)\bar{y}{}^{*}=\bar{U}^{-1}\left\{\frac{1}{\bar{w}}W(y<0)E[\bar{U}((1-z)\bar{Y})]\right\}.$$

According to a theorem of ACZÉL (1966, pp. 151–153)[8], the only shapes the utility functions $\overset{+}{U}(\overset{+}{y})$ and $\bar{U}(\bar{y})$ may then obtain are

$$(20)\qquad \overset{+}{U}(\overset{+}{y})=\alpha\overset{+}{y}{}^{\gamma};\ \alpha>0,\ y>0;$$

and

$$(21)\qquad \bar{U}(\bar{y})=\beta\bar{y}^{\delta};\ \beta>0,\ \delta>0;$$

if we take account of (11) and the monotonicity of $U(.)$. With the aid of these functions, equation (14) can be specified as

$$(22)\qquad U(\overset{+}{y}{}^{*},\bar{y}{}^{*})=\overset{+}{w}\alpha\overset{+}{y}{}^{*\gamma}-\bar{w}\beta\bar{y}{}^{*\delta}.$$

Setting $E[U(V)]=U(\overset{+}{y}{}^{*},\bar{y}{}^{*})\equiv c=\text{const.}$, we can even derive from this equation the explicit functional form of the indifference curves in a $\overset{+}{y}{}^{*}-\bar{y}{}^{*}$ diagram:

$$(23)\qquad \overset{+}{y}{}^{*}=[a+b\bar{y}{}^{*\delta}]^{1/\gamma},\ a=\frac{c}{\overset{+}{w}a},\ b=\frac{\bar{w}\beta}{\overset{+}{w}\alpha}>0.$$

As an example, Figure 16 illustrates the indifference curve system for

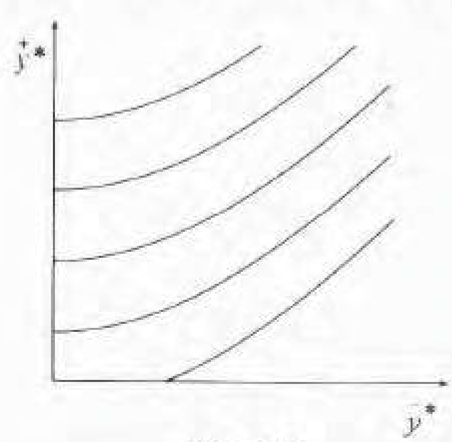


Figure 16

[8] Cf. ch. III A 2.1.

$\delta > \gamma = 1$. It shows that the indifference curves are convex, a plausible property that is indispensable for Schneider's analysis of taxation.

To achieve more general information on the question of convexity, the curvature of the indifference curves should be calculated explicitly. From (22) we get[9]

$$(24) \qquad \left.\frac{d^2\overset{+}{y}{}^*}{d\bar{y}^{*2}}\right|_{U(\overset{+}{y}{}^*,\bar{y}^*)} = \frac{-U_1^2U_{22}+2U_{12}U_1U_2-U_2^2U_{11}}{U_1^3}$$

$$= \frac{-(\overset{+}{w}\alpha\gamma\overset{+}{y}{}^{*\gamma-1})^2[-\bar{w}\beta\delta(\delta-1)\bar{y}^{*\delta-2}]-(-\bar{w}\beta\delta\bar{y}^{*\delta-1})^2[\overset{+}{w}\alpha\gamma(\gamma-1)\overset{+}{y}{}^{*\gamma-2}]}{[\overset{+}{w}\alpha\gamma\overset{+}{y}{}^{\gamma-1}]^3}.$$

Since we are only interested in the sign of this expression some simplifications are possible, and thus we find

$$(25) \qquad \left.\frac{d^2\overset{+}{y}{}^*}{d\bar{y}^{*2}}\right|_{U(\overset{+}{y}{}^*,\bar{y}^*)} \{\gtreqless\} 0 \Leftrightarrow \overset{+}{w}\alpha\gamma(\delta-1)\overset{+}{y}{}^{*\gamma}-\bar{w}\beta\delta(\gamma-1)\bar{y}^{*\delta}\{\gtreqless\}0.$$

The implications of this expression are summarized in the following table.

The Curvature of Indifference Curves $d^2\overset{+}{y}{}^*/d\bar{y}^{*2}\,|_U$

δ \ γ	<1	$=1$	>1
<1	(1) $\gtreqless 0$	<0	<0
$=1$	(2) >0	$=0$	<0
>1	(3) >0	(4) >0	(5) $\gtreqless 0$

If the indifference curves are to be convex, then only the parameter constellations (1) through (5) are relevant, where however cases (1) and (5) are not completely satisfactory since they yield partly convex and partly concave shapes of the indifference curves in the $\overset{+}{y}{}^*-\bar{y}^*$ diagram.

The following five sketches illustrate the implications of the admissible parameter constellations with reference to the utility-of-wealth function $U(v)$:

[9] Cf. fn. 4 in section A.

It is worth noting that the utility curve either has an infinite slope or a slope of zero if $v \to a$ (i.e., $y \to 0$). This aspect produces some implausible kinks and implies that *all* admissible curves are characterized by convex segments that indicate risk preference. Perhaps the most plausible shape is the fifth. It roughly fits a curve proposed by MARKOWITZ (1952b). This curve is convex for small positive changes in wealth to allow for an explanation of gambling and is concave for negative changes in wealth to depict the preference for insurance. Even apart from the problem that we cannot accept the expected-utility argument for gambling[10], the curve (5) must be rejected, however, since at $v = a$ it is characterized by a slope of zero, a property that is not in line with the Non-Saturation Axiom.

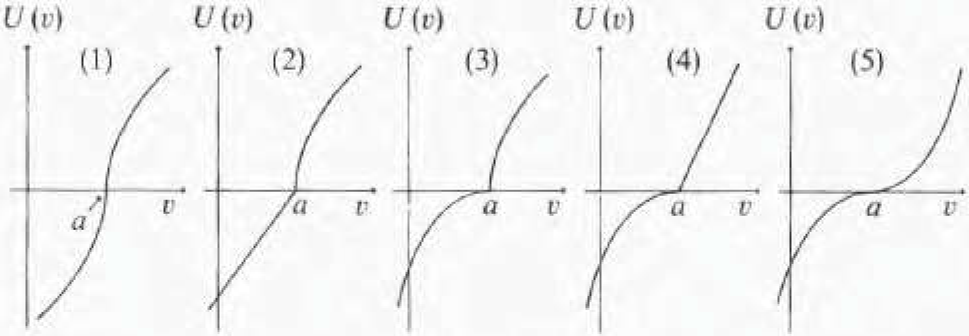


Figure 17

2.1.3. The (μ, σ) Criterion

$$R(V) = U[E(V), \sigma(V)]$$

As shown by RICHTER (1959/60, p. 153)[11] the (μ, σ) criterion coincides with the expected-utility criterion if the utility function is a polynomial of second degree[12],

$$U(v) = v - \alpha v^2. \tag{26}$$

Applying the expectation operator to this expression we have

$$E[U(V)] = E(V) - \alpha E(V^2). \tag{27}$$

[10] Cf. ch. III B 1.3.

[11] Cf. also BORCH (1962, 1968b, and 1969).

[12] This simple version may, for example, be obtained from

$$U(v) = \alpha_1 + \alpha_2 v + \alpha_3 v^2 \quad \text{or} \quad U(v) = (v - \alpha_1) - \alpha_2 (v - \alpha_3)^2$$

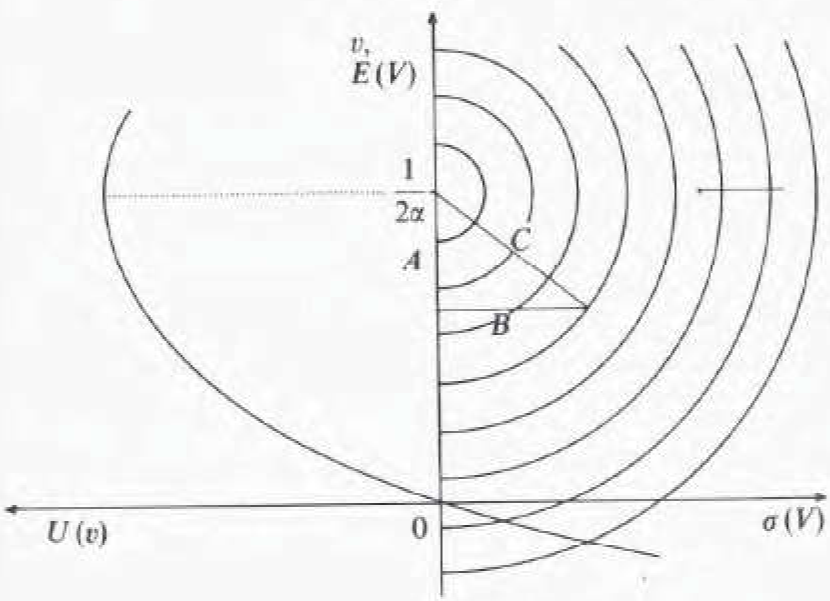
by linear transformations.

Because of[13] $\sigma^2(V) = E(V^2) - E^2(V)$ this implies[14]

$$E[U(V)] = E(V) - \alpha E^2(V) - \alpha\sigma^2(V). \tag{28}$$

Figure 18 shows the corresponding utility and indifference curves in a (μ, σ) diagram. The fact that the utility curve has a maximum[15] is as implausible as the fact that the indifference curves are concentric circles.

The shape of the utility curve follows directly from (26), while the



$A = \frac{1}{2\alpha} - E(V)$

$B = \sigma(V)$

$C = \sqrt{\left(\frac{1}{2\alpha}\right)^2 - \frac{E[U(V)]}{\alpha}}$

Figure 18

---

[13]
$$\begin{aligned}\sigma^2(V) &= E\{[V - E(V)]^2\}\\ &= E[V^2 - 2VE(V) + E^2(V)]\\ &= E(V^2) - 2E(V)E(V) + E^2(V)\\ &= E(V^2) - E^2(V)\end{aligned}$$

[14] The proof that this preference functional is in fact the only one compatible with both the expected-utility criterion and the (μ, σ) criterion is given by MARKOWITZ (1970, pp. 286 ff.) and SCHNEEWEISS (1967a, pp. 113–117). BORCH (1968b) even shows that the preference functional under consideration is the only one compatible with both the (μ, σ) criterion and the Axiom of Independence.

[15] For this reason SCHNEEWEISS (1968a) succeeds in demonstrating the incompatibility between the (μ, σ) criterion and the principle of stochastic dominance. (A random variable V_1 with density function $f_1(.)$ is said to dominate the random variable V_2 with density function $f_2(.)$ if

$$\int_{-\infty}^{v^*} v f_1(v)dv \geq \int_{-\infty}^{v^*} v f_2(v)dv \;\forall v^*, \; -\infty < v^* < +\infty,$$

and if this expression holds with strict inequality for at least one v^*.)

shape of the indifference curves can be found by dividing (28) by $-\alpha$ and adding $[1/(2\alpha)]^2$:

$$(29)\qquad \left(\frac{1}{2\alpha}\right)^2-\frac{E[U(V)]}{\alpha}=\sigma^2(V)+\left\{E^2(V)-2E(V)\frac{1}{2\alpha}+\left(\frac{1}{2\alpha}\right)^2\right\}$$
$$=\sigma^2(V)+\left[\frac{1}{2\alpha}-E(V)\right]^2.$$

On an indifference curve the left side of this expression is constant. Thus, according to *Pythagoras' Theorem*, the indifference curves are circles centered on a point with the coordinates ($\sigma=0$, $E(V)=1/(2\alpha)$)[16].

This implausible description of a preference structure can only act as a deterrant to using the (μ,σ) criterion. A negative marginal utility is certainly unrealistic for it implies that rich people throw their money away. This aspect becomes less important if the distributions to be evaluated are such that the probability of exceeding the maximum of the utility function is zero or at least very small. In this case, only the 'plausible' range of the indifference curves where $dE(V)/d\sigma(V)\,|_{E(U)}>0$ and $d^2E(V)/d\sigma^2(V)\,|_{E(U)}>0$ is relevant.

On the other hand, even in this range the indifference curves have a strange property that is illustrated in Figure 19. Suppose a given income

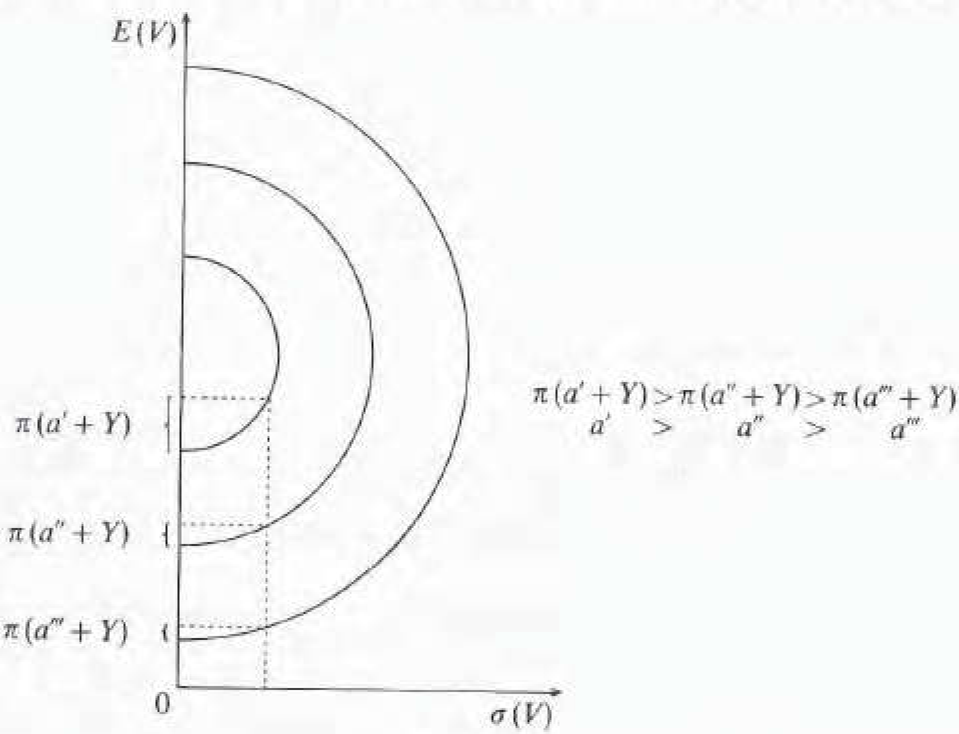


Figure 19

[16] Hicks (1965, p. 115) seems to have been the first to recognize the circular structure of the indifference curves.

distribution Y is to be evaluated. Then the subjective price of risk $\pi(a+Y)$ for this distribution is an *increasing* function of wealth[17]. To see this in Figure 19 it is necessary to move upward on a vertical line at $\sigma(V)=\sigma(Y)=\text{const.}$ and measure the vertical distance between each point reached and the point where the corresponding indifference curve enters the ordinate. In other words, this curious relationship between wealth and the subjective price of risk implies that the intensity of insurance demand

$$g(aq-C)=\frac{E(C)+\pi(aq-C)}{E(C)} \quad (\text{with } Y=a(q-1)-C)$$

for a given risk C rises if wealth is increasing. All experience suggests the opposite.

2.1.4. The Mean-Semivariance Criterion

$$R(V)=U[E(V), \int_{-\infty}^{v^*}(v-v^*)^2 f(v)dv]$$

With a preference functional based on the mean value and the semivariance, some of the implausible aspects of the (μ, σ) criterion can be removed.

Since values of $v>v^*$ do not enter the risk measure, the underlying utility function, if it exists, must be linear for $v>v^*$. However, for $v<v^*$ the function must be concave in order to depict risk aversion. Let us check

$$(30) \qquad U(v)=v-\alpha[\min(v-v^*, 0)]^2$$

by applying the expectation operator. Then, indeed, a suitable preference functional in terms of expected wealth and semivariance[18] can be found:

$$(31) \qquad E[U(V)]=E(V)-\alpha\int_{-\infty}^{v^*}(v-v^*)^2 f(v)dv$$

$$=E(V)-\alpha\sigma^2_{v^*}(V).$$

Figure 20 illustrates the indifference curves described by (31) and the

[17] HICKS (1962, p. 802) remarks pithily: 'That, I submit, is nonsense.' Cf. also ARROW (1965, pp. 35 f.) who speaks of an 'absurdity of the quadratic assumption'.

[18] MARKOWITZ (1970, p. 290).

utility function (30) (where $U(v)$ was shifted in such a way that $U(0)=0$)[19].

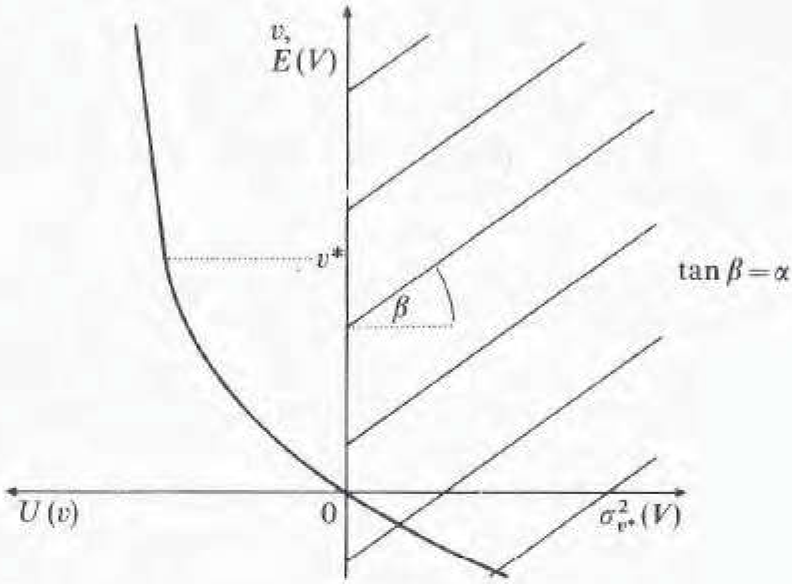


Figure 20

This picture of a preference structure looks much more realistic than the one fitting the (μ, σ) criterion. With the linear part of the utility curve the absurdity of negative marginal utility is avoided. There is also a more satisfactory answer to the question of how the subjective price of risk depends on the decision maker's wealth. Of course, with increasing wealth a, but given income distribution Y, the semivariance

$$\sigma^2_{v^*}(V) = \int_{-\infty}^{v^*} (v - v^*)^2 f(v)\,dv \tag{32}$$

$$= \int_{-\infty}^{v^*-a} (y + a - v^*)^2 f(a+y)\,dy$$

and, together with it, the subjective price of risk $\pi(V) = \alpha\sigma^2_{v^*}(V)$ decline as long as there is a positive probability of wealth falling short of v^*. The only implausible aspect is that risk aversion disappears completely if the whole distribution is situated beyond v^*. But, since v^* can be arbitrarily chosen, this is only a minor defect.

The preceding discussion referred to the version of the semivariance where v^* is a constant. Markowitz, however, considered in addition the

---

19 Analogously to the (μ, σ) criterion, instead of $\sigma^2_{v^*}(V)$, the root $\sigma_{v^*}(V)$ could be used. In this case a system of convex indifference curves would be obtained where the single curves can be transformed into one another by vertical shifts.

case $v^* = E(V)$. For this version no utility function is available since v^*, which is the border between the convex and linear segments of this curve, would vary with each distribution considered.

2.1.5. Result

To summarize, the Domar–Musgrave, the Krelle–Schneider, the (μ, σ), and the mean-semivariance criteria can all be shown to be compatible with the expected-utility criterion in the sense that there are preference structures that, without restricting the class of distributions to be compared, can be represented equally well by both types of criteria. The statement even holds for the operationalized version of the Krelle-Schneider criterion where it is assumed that proportional changes in the distributions of gains and losses imply proportional changes in equivalent gains and losses of equal size.

The resulting preference structures are, however, often not very plausible. The following aspects should be stressed in particular. Because of linear indifference curves, the Domar–Musgrave criterion cannot be used for the analysis of tax-induced behavior changes for which it was formulated. The operationalized version of the Krelle-Schneider criterion designed for the same purpose performs significantly better since convex indifference curves are possible. Less attractive, however, are the admissible utility curves. In the relevant range, they all have strictly convex segments indicating risk loving rather than risk aversion. The (μ, σ) criterion implies a partly negative marginal utility of wealth and risk aversion rising with wealth; both aspects are absurd. The mean-semivariance criterion performs better than the (μ, σ) criterion, but it has the implausible implication that, with bounded probability distributions, risk aversion vanishes completely if wealth is sufficiently large.

Broadly speaking, there seems to be an inverse relationship between the ease of handling of the various criteria and the plausibility of the preference structures that they have in common with the expected-utility criterion. This is a dilemma in our search for an operational alternative to the latter. For example, the rather appealing mean-semivariance criterion does not seem to have advantages in handling compared to the expected-utility criterion. An application of either criterion requires knowledge of the complete shapes of the probability distributions compared. Fortunately, however, there appears to be a way out of the dilemma that makes the (μ, σ) criterion the preferred one despite its apparent implausible implications. From the view point of handling, this criterion has attractive features. As a typical example, we should mention the calculation of μ and σ for a random variable that is the sum of other random variables. Both parameters can, in a very simple way, be calculated from the corresponding parameters of the single items without utilizing complicated and possibly numerical methods for determining the shape of the distribution of the sum variable.

The way out of the dilemma is to forgo the exact representation of implausible preference structures in a (μ, σ) diagram and to attempt instead an approximation of realistic utility functions. The following sections 2.2 and 2.3 deal with the problem.

2.2. *The Local Quadratic Approximation*

2.2.1. The Asymptotic Efficiency of the Variance

If the true utility function is not quadratic, we can nevertheless try to approximate it by a quadratic function (parabola). There are two possible ways of doing this. The first corresponds to the procedure in the previous section. In the range of the probability distributions to be evaluated the true utility function is *globally* replaced by a quadratic function, i.e., for

$$E[U(V)] = E(V) - \alpha E^2(V) - \alpha\sigma^2(V)$$

the parameter α is suitably chosen[20]. The second way is a *local* approximation. The true utility function is replaced, separately for each single distribution of the opportunity set, by a parabola such that, at the mean of this distribution, slope and curvature of both types of utility curves coincide. The difference between the two methods is illustrated in Figure 14 by reference to the marginal utility curves which are linear for the parabola[21]. The method of local approximation was first used by FARRAR (1962, pp. 20 f.) and later by many other authors[22]. In the following, we shall attempt to provide a theoretical legitimation for this method.

---

[20] The usefulness of *global* approximation for small dispersions was shown by SAMUELSON (1970). Cf. also SAMUELSON (1967, p. 9). Samuelson's approximation method is to fit a parabola to the true utility function independently of the decision maker's opportunity set where $v = a$ (initial wealth).

[21] Since the utility function is defined up to a strictly increasing linear transformation the marginal utility function is defined up to the multiplication with a strictly positive constant.

[22] Among these are PRATT (1964, p. 125) and ARROW (1965, pp. 32–35). MARKOWITZ (1970, pp. 120–125), MARKOWITZ and LEVY (1979), and TSIANG (1972, pp. 355–362) calculate clarifying examples in order to demonstrate the usefulness of this method. Other calculations carried out by LEVY (1974) and LOISTL (1976) bring about less optimistic results concerning the quality of approximation. Cf. also TSIANG's (1974) reply to Levy's criticism.

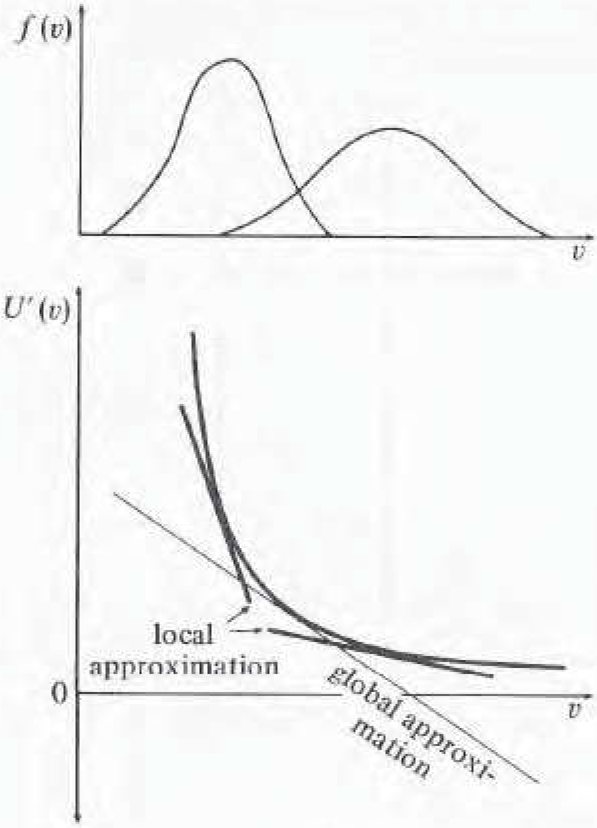


Figure 21

A basic assumption underlying the local approximation procedure is that around $v=\mu\equiv E(V)$ there is a range where the true utility function can be developed into a Taylor series. This means that there must be a range where the true utility function can be depicted through a polynomial, of possibly infinite degree, by calculating all derivatives of the true utility function at $v=\mu$ and setting them equal to the corresponding derivatives of the polynomial. If this basic assumption is not satisfied, a first step of approximation is necessary. This step involves representing as well as possible the true function through a polynomial. It is not considered here. Instead, we analyze the way in which the polynomial itself may be approximated.

The value a polynomial $U(v)$ obtains at $v=\mu+d$ can be calculated by developing a Taylor series at $v=\mu$[23]:

(33) $$U(\mu+d)=U(\mu)+d^1\frac{U^{(1)}(\mu)}{1!}+d^2\frac{U^{(2)}(\mu)}{2!}+d^3\frac{U^{(3)}(\mu)}{3!}+\dots.$$

[23] $U^{(n)}(v)$ denotes the nth derivative of the function $U(.)$ where the argument takes on the value v.

This formula can be used for the evaluation of a whole probability distribution if all variates of this distribution fall into the range where the polynomial fits the true utility function. Define

(34) $$v \equiv \mu + d \quad \text{and} \quad V \equiv \mu + D.$$

Then, applying the expectation operator, we get from (33):

(35) $$E[U(V)] = U(\mu) + E(D^1)\frac{U^{(1)}(\mu)}{1!} + E(D^2)\frac{U^{(2)}(\mu)}{2!} + E(D^3)\frac{U^{(3)}(\mu)}{3!} + \ldots.$$

Since, by construction, $E(D^1) = 0$ this in turn implies

(36) $$E[U(V)] = U(\mu) + E(D^2)\frac{U^{(2)}(\mu)}{2!} + E(D^3)\frac{U^{(3)}(\mu)}{3!} + \ldots,$$

where $E(D^2) = \sigma^2(V)$.

The level of expected utility can hence be expressed as a function of the moments μ, $E(D^2), E(D^3), \ldots$ of the probability distribution to be evaluated. This is an interesting parallel to a conclusion drawn after the discussion of the two-parametric criteria. The conclusion was that, in general, it is impossible to describe a preference ordering over arbitrary distributions with a finite set of statistical distribution parameters. Equation (36) shows where there is an exception to this rule. If the true utility function is a polynomial of degree i, then the derivatives of higher order than i vanish and hence it is possible to express the preference structure in terms of the first i moments only[24]. There is, however, no reason to believe that the utility functions of people form a polynomial of finite order.

The question we are trying to answer is whether, and in what sense, it is possible to approximate the polynomial by means of a parabola, i.e., to neglect the moments of higher order than two. Assume the decision maker knows that each probability distribution in his opportunity set belongs to one of a finite number of linear distribution classes[25]. Consider two arbitrary distributions $V_1 = \mu + D_1$ and $V_2 = \mu + D_2$ from the decision maker's opportunity set. Which of these distributions the decision maker prefers depends on the sign of the difference in their

[24] This result was first achieved by RICHTER (1959/60).

[25] For a definition of a linear distribution class see equation (II A 14).

expected utilities. Because of (36) this difference in expected utilities is given by

$$\Delta U = E[U(\mu + D_1)] - \mathrm{E}[U(\mu + D_2)] = \sum_{i=2}^{m} S_i, \tag{37}$$

where

$$S_i \equiv \frac{U^{(i)}(\mu)}{i!}[E(D_1^i) - E(D_2^i)]. \tag{38}$$

Suppose now the decision maker has decided to calculate the difference between expected utilities solely by reference to the variance and to neglect higher moments. Then he does not make a mistake if

$$|S_2| > \left| \sum_{i=3}^{m} S_i \right|. \tag{39}$$

Unfortunately, in general, we cannot assume that this inequality is satisfied. It is, however, possible to find out when it is valid. For this purpose we consider further pairs of distributions from the same two linear classes to which the distributions V_1 and V_2 belong. The pairs are chosen such that the ratio of their standard deviations equals that of the initial pair, i.e., $\sigma(V_1)/\sigma(V_2)$. Let λ denote the factor by which the standard deviations of the initial pair have to be multiplied to obtain the standard deviations of the new pair under consideration. Then, since

$$\lambda^i S_i = \frac{U^{(i)}(\mu)}{i!}\{E[(D_1\lambda)^i] - E[(D_2\lambda)^i]\}, \tag{40}$$

the difference in expected utilities for a new pair as defined by λ is

$$\Delta U = \sum_{i=2}^{m} \lambda^i S_i \tag{41}$$

and condition (39) becomes

$$|\lambda^2 S_2| > \left| \sum_{i=3}^{m} \lambda^i S_i \right| \tag{42}$$

or, equivalently,

$$(43) \qquad |S_2| > \left| \sum_{i=3}^{m} \lambda^{i-2} S_i \right|.$$

Obviously the right side of this inequality vanishes for $\lambda \to 0$. Thus, provided $|S_2| > 0$, i.e., provided $\sigma(V_1) \neq \sigma(V_2)$, there must be some critical level $\lambda^* > 0$ for the factor λ below which a choice solely with respect to the variance or standard deviation must lead to the correct decision, irrespective of the values the moments of higher order obtain. If the decision maker neglects the moments of orders higher than two when $\lambda > \lambda^*$ he may by chance pick the right distribution, but he may also be mistaken. Figure 22 illustrates this argument.

The important feature of the result is that a discrimination between the distributions considered is possible, in particular, when the standard deviations are small. This aspect is surprising since, for a small level of λ, the absolute difference between the standard deviations is also small, a property that by itself suggests that a discrimination is very difficult.

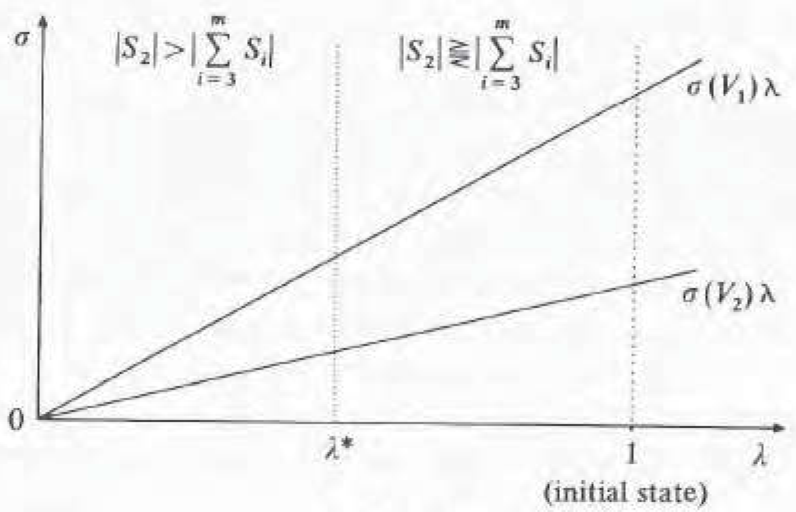


Figure 22

At the present stage of analysis the decision maker knows that for $\lambda < \lambda^*$ he may choose among distributions from two particular linear classes by referring to the standard deviations. His problem, however, is that, since he does not monitor the moments of higher order, he does not know *which* linear classes are involved. Fortunately we can help him. The procedure that was used for a comparison of distributions from two particular linear classes can be repeated for other, arbitrarily selected, pairs of linear classes with the initial distributions having the same standard deviations as before. Each time some $\lambda^*>0$ is found, although not usually the same one. Thus all classes of distributions can be compared and it turns out that, given the total set of linear classes that may occur, there is some lower boundary $\lambda^{**}>0$ for all the λ^*'s such that if $\lambda<\lambda^{**}$ the decision maker can rely on the standard deviations of two distributions compared *without* knowing to which particular linear classes these belong.

A further generalization of the result can be achieved by considering other ratios $\sigma(V_1)/\sigma(V_2)$ representing different degrees of accuracy in the evaluation. In each case standard deviations different from zero are sufficient for correct decision making, provided that $\sigma(V_1)\neq\sigma(V_2)$.

Thus the following conclusion can be drawn for a comparison of probability distributions with equal mean. Suppose there is a sequence of opportunity sets of $n\geq 2$ probability distributions each. Within each set the standard deviations differ, but the pattern of these standard deviations, as given by their *relative* differences, is the same for each set. Then, whatever the sizes of the moments of order higher than two, an expected-utility maximizer can rely on the (μ,σ) criterion for all those opportunity sets in the sequence for which the levels of standard deviations are sufficiently low.

A question not considered up to now is which choice should be made if the distributions to be compared have the same standard deviations. We cannot assume that in this case the decision maker is indifferent. Although it was shown that under certain conditions moments of order higher than two can be neglected in (41), it is not possible to conclude that these moments are also irrelevant in the present case. In fact, the third moment will now appear on the scene. Concerning the choice between distributions from two given linear classes we have, analogously to (42) and (43), the condition

$$|\lambda^3 S_3| > \left|\sum_{i=4}^{m} \lambda^i S_i\right| \tag{44}$$

or, equivalently,

$$|S_3| > \left|\sum_{i=4}^{m} \lambda^{i-3} S_i\right| \tag{45}$$

which allows us to rely on the third moment. Using the same argument as before we can infer that, even without knowing the particular linear classes to which the compared distributions belong, the decision maker who considers only the third moment will not make a mistake if λ is sufficiently small. Thus a conclusion completely analogous to that of the last paragraph emerges if the term 'standard deviation', i.e., the square root of the second moment, is replaced by 'cube root of the third moment'. If the third moments do not differ, the argument can be carried further to show that moments of an even higher order are to be consulted. This indicates that for sufficiently small standard deviations there is a *lexicographic order of moments* and that an 'indifference' found by consulting a limited number of moments may in fact be a pseudo indifference. The fact that, in this lexicographic ordering, the second moment is on a more important rank than all moments of 'higher' orders is the reason for the asymptotic efficiency of the (μ, σ) criterion.

2.2.2. Examples

For the sake of illustration and also for use later, some of the implications of the previous analysis for a local approximation of the particular utility curves

(46)
$$\text{(a)}\quad U(v) = -e^{-\beta v}, \quad \beta > 0,$$
$$\text{(b)}\quad U(v) = \ln v,$$
$$\text{(c)}\quad U(v) = \gamma v^{\gamma}, \quad \gamma \neq 0,$$

are now investigated. We first calculate the ith derivatives of these functions at $v = \mu$. These are

(47)
$$\text{(a)}\quad U^{(i)}(\mu) = -(-\beta)^i e^{-\beta\mu},$$
$$\text{(b)}\quad U^{(i)}(\mu) = \mu^{-i} \prod_{k=0}^{i-1} (-k),$$
$$\text{(c)}\quad U^{(i)}(\mu) = \mu^{\gamma - i} \gamma \prod_{k=0}^{i-1} (\gamma - k),$$

With the aid of these derivatives it is now checked whether the basic requirement that these functions can be represented by a polynomial is satisfied. By the use of the Lagrangean formula

$$R_i = \frac{1}{i!} d^i U^{(i)}(\mu + \Theta d), \; 0 \leq \Theta \leq 1,$$

we calculate the value of the remainder

$$\frac{d^i U^{(i)}(\mu)}{i!} + \frac{d^{i+1} U^{(i+1)}(\mu)}{(i+1)!} + \ldots.$$

If the basic requirement is satisfied then, for all admissible values of Θ, R_i has to vanish as $i \to \infty$. We obtain

$$(48) \qquad \text{(a)} \lim_{i \to \infty} \left[R_i = -\frac{(-\beta d)^i}{i!} e^{-\beta(\mu + \Theta d)} \right] = 0, \quad \text{if} \quad -\infty < d < \infty,$$

$$\text{(b)} \lim_{i \to \infty} \left[R_i = \frac{(-1)^{i-1}}{i} \left(\frac{d}{\mu + \Theta d} \right)^i \right] = 0, \quad \text{if} \quad \frac{1}{2} \le \frac{\mu + d}{\mu} \le 2,$$

$$\text{(c)} \lim_{i \to \infty} \left[R_i = \left(\frac{d}{\mu + \Theta d} \right)^i (\mu + \alpha d)^\gamma \gamma \prod_{k=0}^{i-1} \frac{\gamma - k}{k+1} \right] = 0,$$

$$\text{if} \quad \frac{1}{2} \le \frac{\mu + d}{\mu} \le 2.$$

Thus, in case (a), the whole function can be developed into a Taylor series. But, in cases (b) and (c), the standard deviations of the probability distributions to be compared have to be small enough to ensure that the ranges of these distributions do not extend above twice and below half their mean values[26].

Now we utilize the derivatives from (47) to specify the expression for S_i in (38). Condition (42) will then, according to the underlying utility function, become one of the three following inequalities:

$$(49) \qquad \text{(a)} \left| \lambda^2 \frac{-\beta^2}{2} x_2 \right| > \left| \sum_{i=3}^{m} \lambda^i \frac{-(-\beta)^i}{i!} x_i \right|,$$

$$\text{(b)} \left| \left(\frac{\lambda}{\mu} \right)^2 \left(-\frac{1}{2} \right) x_2 \right| > \left| \sum_{i=3}^{m} \left(\frac{\lambda}{\mu} \right)^i \frac{\prod_{k=1}^{i-1} -k}{i!} x_i \right|,$$

$$\text{(c)} \left| \left(\frac{\lambda}{\mu} \right)^2 \frac{\gamma^2(\gamma - 1)}{2} x_2 \right| > \left| \sum_{i=3}^{m} \left(\frac{\lambda}{\mu} \right)^i \frac{\gamma \prod_{k=1}^{i-1} (\gamma - k)}{i!} x_i \right|,$$

$$\text{with} \quad x_i \equiv E(D_1^i) - E(D_2^i).$$

---

[26] It is not surprising that LOISTL (1976) succeeds in demonstrating that a Taylor expansion of $\ln v$ and γv^γ for the normal (range from $-\infty$ to $+\infty$) and the log normal (range from 0 to $+\infty$) distributions gives a poor quality approximation. Loistl's overall rejection of the Taylor expansion (p. 909) cannot be accepted. In his examples, the coefficients of variation chosen are implausibly large. The reason is that he applies the two functions mentioned to income rather than to wealth. If his examples are recalculated by using the utility-of-wealth functions, it seems that significantly better degrees of approximation can be achieved. It should be mentioned also that Loistl (p. 906) erroneously assumes that the range in which the utility functions can be developed into a polynomial is $0 \le (\mu + d)/\mu \le 2$; cf. expression (48) above.

It is worth noting that, in (b) and (c), the mean μ of the distributions to be compared appears, while in (a) it could be eliminated. In (b) and (c) the mean plays an important role. Assume that, given μ and given a particular pair of linear distribution classes, for (b) or (c) we have calculated the critical level λ^*, below which a choice with respect to standard deviations is admissible. Then, with a change in μ, λ^* must change in strict proportion, for the inequalities (b) and (c) are unaffected if $\lambda/\mu=$ const. Since a similar result holds for a comparison of distributions from all possible linear classes it is clear that the minimum of all the λ^*'s, i.e. λ^{**}, is also proportional to the mean, as long as the set of possible linear distribution classes is given. This implies that in the (μ, σ) diagram we can plot a ray through the origin which together with the μ axis encloses a range where, with a given degree of approximation, the standard deviation can be used to discriminate between probability distributions of equal mean.

TSIANG (1972, esp. pp. 358 f.) claims a similar result for all utility functions listed in (46). Concerning the function $U(v)=-e^{-\beta v}$ we cannot, however, agree with this claim since the inequality (a) in (49) is independent of μ. If, for all μ, there is a given set of linear distribution classes that the decision maker considers possible, then the border line of the approximation range is not a ray through the origin, but a parallel to the μ axis. The explanation for this divergence is that Tsiang does not consider, as we do, the problem of approximating a *given* function $U(v)$. He assumes instead that the total shape of the utility curve to be approximated depends on the decision maker's expected wealth: he takes the parameter β in the exponential utility function to be given by the equation $\beta=k/\mu$, $k=$ const.>0. This, however, is not admissible since it contradicts the Archimedes Axiom, one of the axioms underlying the expected-utility rule. By construction, von Neumann-Morgenstern utility $U(v)$ equals, up to an increasing linear transformation, the 'indifference probability', whose existence and uniqueness is postulated by this axiom. With $\beta=k/\mu$, Tsiang assumes that the indifference probability depends on the decision maker's opportunity set. This violates the uniqueness postulate.

2.2.3. The Shape of the Pseudo Indifference Curves in the (μ, σ) Diagram

The preceding section provided the basis for applying Farrar's method of local quadratic approximation of the true utility curve. For small values of σ, equation (35) can thus be simplified to

$$\text{(50)} \qquad \mathrm{E}[U(V)]=U(\mu, \sigma) \approx U(\mu)+\frac{\sigma^2(V)\, U''(\mu)}{2}.$$

It is not difficult to draw from this equation some information that allows a graphical representation of the preference structure in a (μ, σ) diagram to be made. The preference structure is described by so-called pseudo indifference curves. The term 'pseudo' is chosen as a reminder of the lexicographic ordering of moments which implies that strict indifference cannot generally by ensured by considering only two distribution parameters.

Assume for a moment that $U''(v) = \text{const.}$ for all v. In this case, the local approximation is globally correct because for each v we find $U(v) = v - \alpha v^2$ where $\alpha = -U''/2$. There are genuine indifference curves that, as shown in section 2.1.3, are circles whose center is on the v axis at $v = 1/(2\alpha)$. What changes if U'' depends on v?

If $U''' \neq 0$, the logic of local approximation requires consideration of an alternative system of circles for each possible μ. Consider a particular distribution with mean $\mu = \mu^*$ and standard deviation $\sigma = \sigma^*$ as illustrated in Figure 23. Local approximation means setting $U''(\mu) = U''(\mu^*)$ in (50) and hence fitting the circles so that their center is on the μ axis at $v = -1/U''(\mu^*)$. The point where the circle that goes through (μ^*, σ^*) enters the μ axis (below μ^*) indicates the locally approximated certainty equivalent $S(V)$ of the distribution (μ^*, σ^*).

Now consider other points in the (μ, σ) diagram that bring about the same certainty equivalent as (μ^*, σ^*). The geometrical locus of these points is a pseudo indifference curve. Obviously, in the case $U''' \neq 0$, the pseudo indifference curve cannot coincide with the segment of the circle connecting points (μ^*, σ^*) and $(0, S(V))$. The reason is that for $\mu \neq \mu^*$ we have $U''(\mu) \neq U''(\mu^*)$ and hence another system of circles with a center at $v = -1/U''(\mu)$ has to be consulted. Suppose, for example, $U''' > 0$ so that

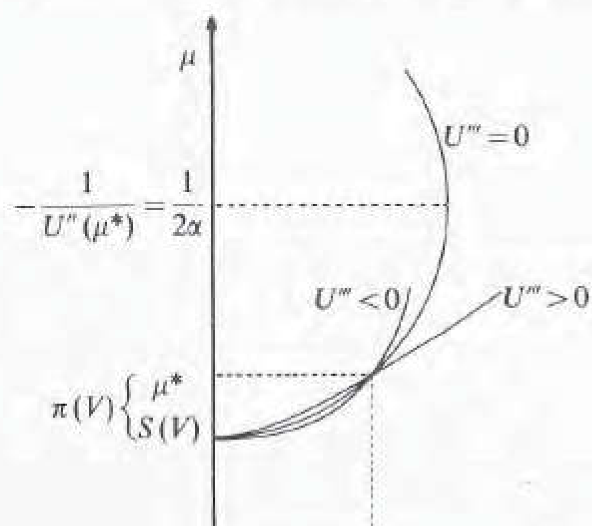


Figure 23

a rise in μ increases $-1/U''(\mu)$, i.e., shifts the center of the circles upwards. An inspection of Figure 23 shows that, in this case, at point (μ^*, σ^*) the pseudo indifference curve must be flatter than the corresponding circle segment, for otherwise an upward movement along the pseudo indifference curve increases the certainty equivalent, which is a contradiction.

The general formula for the slope of the pseudo indifference curves can easily be calculated from (50):

$$(51)\qquad \left.\frac{d\mu}{d\sigma}\right|_{U(\mu,\sigma)} = -\frac{\dfrac{\partial U(\mu,\sigma)}{\partial \sigma}}{\dfrac{\partial U(\mu,\sigma)}{\partial \mu}} \approx -\frac{\sigma U''(\mu)}{U'(\mu)+\dfrac{\sigma^2}{2}U'''(\mu)}.$$

In the special case $U'''=0$, where the, otherwise, pseudo indifference curves coincide with the concentric-circle indifference curves of section 2.1.3, the slope is $\sigma U''/U'$. If however $U'''>0$ or $U'''<0$ then, for a given point in the (μ, σ) diagram, the slope of the pseudo indifference curve is respectively lower or higher than that of the corresponding circle segment[27]. With reference to point (μ^*, σ^*), the possibilities are illustrated in Figure 23.

Apart from the comparison with the indifference curves relevant under global approximation, (51) also carries interesting information in itself. Since $U''<0$ and $U'>0$ we have

$$(52)\qquad \left.\frac{d\mu}{d\sigma}\right|_{U(\mu,\sigma)} > 0 \quad \text{for } 0<\sigma<\bar{\sigma}, \text{ if } \bar{\sigma} \text{ is sufficiently small,}$$

and

$$\lim_{\sigma\to 0} \left.\frac{d\mu}{d\sigma}\right|_{U(\mu,\sigma)} = 0.$$

The case $\sigma \to 0$ is of some significance for it indicates that the pseudo

---

[27] As will be shown below in this section, $U'''>0$ ($U'''<0$) implies a preference for (against) right skewed distributions. For small dispersions, however, this preference is unimportant since $E(D^3)\approx 0$. U''' only carries the information about how U'', and thus the circle center $-1/U''$, changes with μ. The effect represented by $U'''\neq 0$ is sometimes overlooked in economic model building. Cf. HOCHGESAND (1974, pp. 45 f.) and TSIANG (1972, pp. 356 and 364). The correct formula can, however, be found in TSIANG (1974, appendix).

indifference curves enter the ordinate perpendicularly[28,29]. This aspect extends the above remarks concerning the lexicographic ordering of moments to the first moment. In the case of sufficiently small dispersions, even the standard deviation can be neglected in evaluating proba bility distributions[30]. The mean or expected value ranks higher than any other moment. This phenomenon already had been noted by BERNOULLI (1738, § 9) who explained it by the fact that, for very small dispersions, the utility curve can be considered 'as an infinitesimally small straight line'[31].

It seems worth trying to see how the range of possible shapes of the pseudo indifference curves is reduced if a realistic restriction on the shape of the utility function is imposed. This restriction is that the intensity of insurance demand $g(aq-C)=[E(C)+\pi(aq-C)]/E(C)$ for a given risk C should not increase with a rise in wealth or, in the terminology of PRATT (1964) and ARROW (1965), that *absolute risk aversion* should be constant or decreasing. It can easily be shown that the restriction implies $U'''(.)>0$ so that only one of the three possible shapes illustrated in Figure 23 is relevant[32]. Following PRATT (1964), we assume that the subjective price of risk $\pi(V)$ can be calculated from the linear approximation

$$(53) \qquad U[\mu-\pi(V)]\approx U(\mu)-U'(\mu)\pi(V), \quad V=aq-C,$$

since its size is very small compared with the range of the corresponding probability distribution. Setting (53) equal to (50) we thus have

$$(54) \qquad \begin{aligned} &U[\mu-\pi(V)]=E[U(V)],\\ &U(\mu)-U'(\mu)\pi(V)\approx U(\mu)+\frac{\sigma^2(V)}{2}U''(\mu)\end{aligned}$$

28 This property sometimes does not show up in graphs of the (μ,σ) diagram. Cf. e.g. LUTZ (1951, pp. 190 f.) or FAMA and MILLER (1972, pp. 220, 222, 223, 282).

29 That $d\mu/d\sigma\to 0$ for $\sigma\to 0$ follows from TOBIN's (1958, p. 13; his figures 4 and 7 are wrong in this respect) equation (9) and is the subjet of SCHNEEWEISS's theorem 7 (1967a, p. 128). While (52) is derived without particular constraints on the distribution classes (except for the boundedness assumption), the results of Tobin and Schneeweiß are based on the assumption of a linear distribution class. Cf. equation (62) below.

30 The irrelevance of the variance for an evaluation of small probability distributions is a severe handicap for an experimental assessment of the von Neumann-Morgenstern function. This remark, for example, applies to the experiments of MOSTELLER and NOGEE (1951) where people were offered bets with prizes that were by some orders of magnitude lower than in real life decision problems. Cf., to this problem, SAMUELSON (1960, p. 35).

31 Cf. LAPLACE (1814, p. XVI).

32 The relationship between the wealth dependence of risk aversion and the sign of U''' was shown by STIGLITZ (1969a, p. 279) and HIRSHLEIFER (1970, p. 283, footnote).

and hence

$$\pi(V) \approx \frac{\sigma^2(V)}{2}\beta(\mu) \tag{55}$$

where

$$\beta(v) \equiv -\frac{U''(v)}{U'(v)} \tag{56}$$

is the Pratt–Arrow measure of absolute risk aversion. Since, given the loss distribution C, an increase in initial wealth leaves $\sigma^2(V) = \sigma^2(aq - C) = \sigma^2(C) = \text{const.}$, a rise in the subjective price of risk π will not occur if $\beta'(v) \leq 0$, i.e., if

$$\frac{\partial}{\partial\mu}\left(-\frac{U''(\mu)}{U'(\mu)}\right) \leq 0 \tag{57}$$

or, equivalently,

$$\frac{-U'''(\mu)U'(\mu) + U''^2(\mu)}{U'^2(\mu)} \leq 0$$

and hence, as was to be shown, if

$$U'''(\mu) \geq \frac{U''^2(\mu)}{U'(\mu)} > 0. \tag{58}$$

Expression (58) is an implication of the plausible requirement that risk aversion should not increase with wealth. But it can also be defended on its own grounds. By inspection of (36) we find that $U''' > 0$ implies a preference for right skewed distributions which are characterized by[33] $E(D^3) > 0$. Such a preference was already claimed by Marschak (1938, p. 320) and Hicks (1967, p. 119). Markowitz (1952a, pp. 87–91; 1952b, p. 156) also observed the preference, but he dismissed it as being a preference peculiar to gamblers. Gamblers tend to reduce their stakes when their gambling capital declines and to increase them when this capital rises, with the result that the distribution of the sum of prizes will automatically become skewed to the right. The phenomenon also showed up in the game experiments of Mosteller and Nogee (1951, p. 389). It does

[33] Cf. fn. 18 in section A.

not seem, however, that the preference for right skewed distributions is restricted to gambling. The institution of limited liability in stock holding or the stop-loss reinsurance contracts bought by insurance companies are clear signs of a much broader relevance.

The possibility of deriving indifference curves by local quadratic approximation is immune to the usual criticism of the (μ, σ) criterion. For small dispersions of the probability distributions in the decision maker's opportunity set, this criterion, in practice, coincides with the expected-utility criterion. The method of local approximation is flexible enough to represent a large variety of aspects of the decision maker's preference structure, without imposing any restrictions other than that the linear distribution classes the decision maker thinks possible are bounded in v.

For large dispersions, however, the quality of approximation may be poor. In this case the method of local quadratic approximation cannot do more than hint at the optimal solution.

2.3. *Indifference Curves in the (μ, σ) Diagram for Linear Distribution Classes*

The deficiency of the method of local approximation in the case of wide dispersions does not mean that it is impossible to construct indifference curves in the (μ, σ) diagram that will lead to an optimal choice. Actually, as is known from section A 6, it is generally possible to represent exactly in a (μ, σ) diagram any preference structure over distributions from a linear class. Thus, it makes sense to try to find out what the relationship is between the indifference curves and the von Neumann-Morgenstern function in the presence of such a linear class.

The analysis is based on the assumption that it is possible to write expected utility in the form

$$(59) \qquad E[U(V)] = \mathrm{E}[U(\mu + \sigma Z)] \text{ with } Z = \frac{V-\mu}{\sigma},\ E(Z) = 0,\ \sigma(Z) = 1.$$

If the utility function is continuous in the range from $-\infty$ to $+\infty$, no constraints have to be imposed on the range of[34] Z. If, however, the range over which the function is defined is limited in a particular direction or if there is a discontinuity or even a lexicographic boundary, Z has to be constrained in this direction. The following results are then only valid if the variates of the wealth distributions to be evaluated cannot go beyond the range where $U(.)$ is continuous and well-defined.

[34] Cf. fn. 16 in ch. III B.

What the (pseudo) indifference curves look like when a lexicographic border can be crossed was shown in section B 1.1.

By implicit differentiation of (59) for $E[U(V)] = U(\mu, \sigma) = \text{const.}$, we find that the slope of an indifference curve is given by

$$(60)\qquad \left.\frac{d\mu}{d\sigma}\right|_{U(\mu,\sigma)} = -\frac{E[ZU'(\mu+\sigma Z)]}{E[U'(\mu+\sigma Z)]} = \frac{-\operatorname{cov}[Z, U'(\mu+\sigma Z)]}{E[U'(\mu+\sigma Z)]}$$

where $\operatorname{cov}(Z, U') = E(ZU') - E(Z)E(U')$ denotes the convariance[35] between Z and U' and $E(Z) = 0$ by definition. The assumption of risk aversion, $U'' < 0$, implies $\operatorname{cov}(Z, U') < 0$ if $\sigma > 0$. Hence

$$(61)\qquad \left.\frac{d\mu}{d\sigma}\right|_{U(\mu,\sigma)} > 0 \quad \text{for } \sigma > 0.$$

Now, $\sigma \to 0$ implies $U'(\mu + \sigma Z) \to U'(\mu)$ for all variates of Z. This gives

$$(62)\qquad \lim_{\sigma \to 0} \left.\frac{d\mu}{d\sigma}\right|_{U(\mu,\sigma)} = 0.$$

Obviously these aspects of the indifference curve in the (μ, σ) diagram are perfectly compatible with the results of local approximation as given by expression (52) above.

Other interesting aspects that may be useful for economic model building based on the (μ, σ) approach require answers to the questions of how the indifference-curve slope changes if

(1) μ rises given σ $(\sigma > 0)$,
(2) μ and σ change, as represented by a movement along a given indifference curve, and
(3) σ rises given μ $(\mu > 0)$.

We shall now consider these questions.

---

[35] Note that

$$\begin{aligned}\operatorname{cov}(Z, X) &= E\{[Z - E(Z)][X - E(X)]\} \\ &= E(ZX) - E[ZE(X)] - E[XE(Z)] + E[E(Z)E(X)] \\ &= E(ZX) - E(Z)E(X).\end{aligned}$$

(1) Differentiation of (60) yields

$$\frac{\partial}{\partial\mu}\left(\frac{d\mu}{d\sigma}\bigg|_U\right) = -\frac{E[ZU''(\mu+\sigma Z)]E[U'(\mu+\sigma Z)] - E[U''(\mu+\sigma Z)]E[ZU'(\mu+\sigma Z)]}{E^2[U'(\mu+\sigma Z)]}.$$

In equation (56) the Pratt-Arrow measure of absolute risk aversion $\beta(v)>0$ was defined. Utilizing this measure we have in abbreviated notation[36]

$$\text{(63)}\qquad \begin{aligned} &\operatorname{sgn}\frac{\partial}{\partial\mu}\left(\frac{d\mu}{d\sigma}\bigg|_U\right)\\ &=\operatorname{sgn}[E(Z\beta U')E(U')-E(ZU')E(\beta U')]\\ &=\operatorname{sgn}\left[E\left(Z\frac{\beta U'}{E(\beta U')}\right)-E\left(Z\frac{U'}{E(U')}\right)\right]\\ &=\operatorname{sgn}\left\{\left|\operatorname{cov}\left(Z,\frac{U'}{E(U')}\right)\right|-\left|\operatorname{cov}\left(Z,\frac{\beta U'}{E(\beta U')}\right)\right|\right\}.\end{aligned}$$

Suppose the utility function is such that the subjective price of risk, as given in (55), is independent of wealth, i.e., suppose with $\beta'=0$ we have constant absolute risk aversion in the Pratt-Arrow terminology. Then both covariance terms in (63) take on the same value and hence $\partial(d\mu/d\sigma\,|_U)/\partial\mu=0$. If $\beta'<0$ (decreasing absolute risk aversion) the second term dominates the first, and if $\beta'>0$ (increasing absolute risk aversion) the first dominates the second. Hence

$$\text{(64)}\qquad \frac{\partial}{\partial\mu}\left(\frac{d\mu}{d\sigma}\bigg|_{U(\mu,\sigma)}\right)\left\{\begin{smallmatrix}\geq\\=\\<\end{smallmatrix}\right\}0 \Leftrightarrow \beta'\left\{\begin{smallmatrix}\geq\\=\\<\end{smallmatrix}\right\}0.$$

It will be argued in chapter III A 2 that the intensity of insurance demand is declining with a rise in wealth, i.e., that the utility function exhibits the property of decreasing absolute risk aversion. According to (64), in this case the indifference-curve slope in the (μ,σ) diagram is declining with a rise in μ, given $\sigma>0$.

[36] We define

$$\operatorname{sgn} x = \begin{cases} +1 & \text{for } x>0,\\ 0 & \text{for } x=0,\\ -1 & \text{for } x<0.\end{cases}$$

(2) Usually the indifference curves in the (μ, σ) diagram are assumed to be convex so that

$$(65) \qquad \left.\frac{d^2\mu}{d\sigma^2}\right|_{U(\mu,\sigma)} > 0.$$

Indeed, as shown by TOBIN (1958), this property follows from the assumption of a strictly concave von Neumann-Morgenstern function, i.e., from $U''<0$. The argument runs as follows[37]. Consider two different points (μ_1, σ_1) and (μ_2, σ_2) which are both situated on the same indifference curve. Obviously this indifference curve is strictly convex if, and only if, for any pair of such points

$$(66) \qquad (\mu_1, \sigma_1) \sim (\mu_2, \sigma_2) < \left(\frac{\mu_1 + \mu_2}{2}, \frac{\sigma_1 + \sigma_2}{2}\right).$$

The assumption of a strictly concave utility function in turn implies

$$(67) \qquad \frac{U(\mu_1 + z\sigma_1)}{2} + \frac{U(\mu_2 + z\sigma_2)}{2} \leq U\left(\frac{\mu_1 + z\sigma_1}{2} + \frac{\mu_2 + z\sigma_2}{2}\right)$$

where the inequality sign holds strictly for all z except for the special case where $\mu_1 + z\sigma_1 = \mu_2 + z\sigma_2$. Applying the expectation operator we thus have

$$(68) \quad \frac{E[U(\mu_1 + Z\sigma_1)]}{2} + \frac{E[U(\mu_2 + Z\sigma_2)]}{2} < E\left[U\left(\frac{\mu_1 + \mu_2}{2} + \frac{\sigma_1 + \sigma_2}{2} Z\right)\right].$$

By assumption, (μ_1, σ_1) and (μ_2, σ_2) are chosen such that $E[U(\mu_1 + Z\sigma_1)] = E[U(\mu_2 + Z\sigma_2)]$. Hence

$$(69) \quad E[U(\mu_1 + Z\sigma_1)] = E[U(\mu_2 + Z\sigma_2)] < E\left[U\left(\frac{\mu_1 + \mu_2}{2} + \frac{\sigma_1 + \sigma_2}{2} Z\right)\right].$$

This expression has the same meaning as (66) and, since it holds for any pair of different points on an indifference curve, it proves (65).

---

[37] SCHNEEWEISS (1967a, pp. 126–128) and FELDSTEIN (1969) prove that not all classes of two-parametric distributions bring about convex indifference curves. They both refer to the class of logarithmic normal distributions, where a particular distribution of this class is completely determined by μ and σ, and show that, in connection with some special utility functions, the indifference curves will be concave for σ sufficiently large. This result does not contradict the one reported in the text since the class of logarithmic normal distributions is not a linear class as defined by equation (14) in section II A.

(3) The change in the indifference-curve slope brought about by an increase in σ, given μ, can be considered to be composed of two parts, the change resulting from a movement along a given indifference curve $d\mu^2/d\sigma^2|_U$, and the change resulting from a movement from this indifference curve downward to compensate for the change in μ, $-[d\mu/d\sigma|_U][\partial(d\mu/d\sigma|_U)/\partial\mu]$. Hence[38]

$$(70) \qquad \frac{\partial}{\partial\sigma}\left(\frac{d\mu}{d\sigma}\bigg|_U\right)=\frac{d^2\mu}{d\sigma^2}\bigg|_U-\frac{d\mu}{d\sigma}\bigg|_U\,\frac{\partial}{\partial\mu}\left(\frac{d\mu}{d\sigma}\bigg|_U\right).$$

If we bring together the pieces of information given by (61), (64), and (65), then this expression implies

$$(71) \qquad \frac{\partial}{\partial\sigma}\left(\frac{d\mu}{d\sigma}\bigg|_U\right)>0 \quad \text{if } \beta'\leq 0.$$

Thus, in the realistic case where the intensity of insurance demand for a given risk does not rise with an increase in wealth or, equivalently, where the utility function exhibits the property of constant or decreasing absolute risk aversion, a rise in σ, given μ, increases the indifference-curve slope.

At the expense of the assumption of a linear distribution class, these results confirm and extend the findings of the last section to the case where the decision maker has to choose between probability distributions with large dispersions. The derived properties of indifference curves in a (μ,σ) diagram may, and indeed will later in this book, be helpful for constructing models of economic behavior under uncertainty.

2.4. *Conclusions: The (μ,σ) Criterion as Proxy for the Expected-Utility Criterion*

From the above analysis, the (μ,σ) criterion appears to be the practical alternative to the expected-utility criterion that we were looking for. Although the intersection of preference structures that the (μ,σ) criterion has in common with the expected-utility rule does not appear to be a very plausible one, a good case can be made for this criterion by referring to the asymptotically lexicographic ordering of moments. With the aid of the parameters μ and σ that are on the two highest ranks in this ordering, it is possible to approximate almost arbitrarily shaped von

[38] Cf. fn. 4 in section A.

Neumann-Morgenstern utility functions, if the dispersions of the probability distributions in the decision maker's opportunity set are sufficiently small. In addition, like many other two-parametric criteria, the (μ, σ) criterion can represent exactly the expected-utility criterion even in the large, if choice is constrained to selecting a distribution from a linear class. Last, but not least, by using the tools provided by mathematical statistics, the (μ, σ) approach can easily be handled in practical and theoretical analysis.

It is not possible to indicate in general whether, in decision problems under risk, the indirect (μ, σ) approach is preferable to the direct use of the expected-utility criterion. If the builder of an economic model wants to take advantage of the comparative simplicity of the (μ, σ) criterion, he has to consider the following possibilities before making up his mind.

(1) The (μ, σ) criterion coincides with the expected-utility criterion because all distributions in the opportunity set belong to the same linear class.
(2) The (μ, σ) criterion approximates the expected-utility criterion since
  i) the dispersions of the distributions to be compared are small,
  ii) the distributions in the opportunity set approximately form a linear class (e.g. the class of normal distributions).
(3) The (μ, σ) criterion cannot be applied since a choice among widely dispersed distributions from very different linear classes is to be modelled.

Although (3) may be the possibility with the highest practical relevance, abstract economic models studying only a very limited number of choice problems at a time will often be able to take advantage of possibility (1). Actually it seems that most of the published expected-utility approaches to economic decision problems under uncertainty belong to this category. In the majority of cases, therefore, despite views to the contrary expressed occasionally by some authors, there is no justification for claiming a higher degree of generality for the expected-utility approach than for the (μ, σ) approach.

Appendix 1 to Chapter II

$$\begin{aligned}\sigma^2(V) &= \int[v-E(V)]^2 f(v)dv = \int[a+y-E(a+Y)]^2 f(a+y)dy\\ &= \int[y-E(Y)]^2 f(a+y)dy\\ &= \int[y^2-2yE(Y)+E^2(Y)]^2 f(a+y)dy\\ &= \int y^2 f(a+y)dy - 2E(Y)\int y f(a+y)dy + E^2(Y)\int f(a+y)dy\end{aligned}$$

$$= E(Y^2) - E^2(Y)$$

$$= \int_{-\infty}^{0} \{[y - E(-\bar{Y})] + E(-\bar{Y})\}^2 f(a+y)dy$$

$$+ \int_{0}^{+\infty} \{[y - E(\overset{+}{Y})] + E(\overset{+}{Y})\}^2 f(a+y)dy - E^2(Y)$$

$$= \int_{-\infty}^{0} \{[y - E(-\bar{Y})]^2 + 2[y - E(-\bar{Y})]E(-\bar{Y}) + E^2(-\bar{Y})\} f(a+y)dy$$

$$+ \int_{0}^{+\infty} \{[y - E(\overset{+}{Y})]^2 + 2[y - E(\overset{+}{Y})]E(\overset{+}{Y}) + E^2(\overset{+}{Y})\} f(a+y)dy - E^2(Y)$$

Now define $\bar{w} \equiv W(y<0)$ and $\overset{+}{w} \equiv W(y \geq 0)$; then we get

$$\sigma^2(V) = \frac{\int_{-\infty}^{0} [y - E(-\bar{Y})]^2 f(a+y)dy}{\bar{w}} \bar{w}$$

$$+ \int_{-\infty}^{0} \{2yE(-\bar{Y}) - 2E^2(-\bar{Y}) + E^2(-\bar{Y})\} f(a+y)dy$$

$$+ \frac{\int_{0}^{+\infty} [y - E(\overset{+}{Y})]^2 f(a+y)dy}{\overset{+}{w}} \overset{+}{w}$$

$$+ \int_{0}^{+\infty} \{2yE(\overset{+}{Y}) - 2E^2(\overset{+}{Y}) + E^2(\overset{+}{Y})\} f(a+y)dy - E^2(Y)$$

$$= \sigma^2(\bar{Y})\bar{w} + 2E(-\bar{Y}) \frac{\int_{-\infty}^{0} yf(a+y)dy}{\bar{w}} \bar{w} - E^2(-\bar{Y}) \int_{-\infty}^{0} f(a+y)dy$$

$$+ \sigma^2(\overset{+}{Y})\overset{+}{w} + 2E(\overset{+}{Y}) \frac{\int_{0}^{+\infty} yf(a+y)dy}{\overset{+}{w}} \overset{+}{w} - E^2(\overset{+}{Y}) \int_{0}^{+\infty} f(a+y)dy - E^2(Y)$$

$$= \sigma^2(\bar{Y})\bar{w} + 2E(\bar{Y})E(\bar{Y})\bar{w} - E^2(\bar{Y})\bar{w}$$

$$+ \sigma^2(\overset{+}{Y})\overset{+}{w} + 2E(\overset{+}{Y})E(\overset{+}{Y})\overset{+}{w} - E^2(\overset{+}{Y})\overset{+}{w} - E^2(Y)$$

$$= \bar{w}\sigma^2(\bar{Y}) + \bar{w}E^2(\bar{Y}) + \overset{+}{w}\sigma^2(\overset{+}{Y}) + \overset{+}{w}E^2(\overset{+}{Y}) - E^2(Y)$$

$$= W(y<0)\{\sigma^2(\bar{Y}) + E^2(\bar{Y})\} + W(y \geq 0)\{\sigma^2(\overset{+}{Y}) + E^2(\overset{+}{Y})\} - E^2(Y),$$

q.e.d.

Appendix 2 to Chapter II

Define

$$w^* \equiv W(V < v^*) = \int_{-\infty}^{v^*} f(v)\,dv,$$

$$E(V^*) \equiv E(V)\big|_{v<v^*} = \frac{\int_{-\infty}^{v^*} v f(v)\,dv}{w^*},$$

$$\sigma^2(V^*) \equiv \sigma^2(V)\big|_{v<v^*} = \frac{\int_{-\infty}^{v^*} [v - E(V^*)]^2 f(v)\,dv}{w^*},$$

then

$$\sigma^2_{v^*}(V) = \int_{-\infty}^{v^*} (v - v^*)^2 f(v)\,dv = \int_{-\infty}^{v^*} (v^2 - 2vv^* + v^{*2}) f(v)\,dv$$

$$= \int_{-\infty}^{v^*} v^2 f(v)\,dv - 2v^* \frac{\int_{-\infty}^{v^*} v f(v)\,dv}{w^*} w^* + v^{*2} \int_{-\infty}^{v^*} f(v)\,dv$$

$$= \int_{-\infty}^{v^*} [v - E(V^*) + E(V^*)]^2 f(v)\,dv - 2v^* E(V^*) w^* + v^{*2} w^*$$

$$= \frac{\int_{-\infty}^{v^*} \{[v - E(V^*)]^2 + 2[v - E(V^*)]E(V^*) + E^2(V^*)\} f(v)\,dv}{w^*} w^*$$
$$- 2v^* E(V^*) w^* + v^{*2} w^*$$

$$= \sigma^2(V^*) w^* + \int_{-\infty}^{v^*} [2vE(V^*) - 2E^2(V^*)] f(v)\,dv + E^2(V^*) w^*$$
$$- 2v^* E(V^*) w^* + v^{*2} w^*$$

$$= \sigma^2(V^*) w^* + 2E^2(V^*) w^* - 2E^2(V^*) w^* + E^2(V^*) w^*$$
$$- 2v^* E(V^*) w^* + v^{*2} w^*$$

$$= w^*[\sigma^2(V^*) + E^2(V^*) - 2v^* E(V^*) + v^{*2}]$$

$$= W(v < v^*)\{\sigma^2(V^*) + [E(V^*) - v^*]^2\}, \text{ q.e.d.}$$