# Economic Decisions under Uncertainty

by Hans-Werner Sinn

North Holland: Amsterdam, New York and Oxford 1983

Chapter 5: Areas of Application

# Chapter Five
# Areas of Application

The theory of decision making under uncertainty developed above has a large number of possible applications. Here we consider in more detail three of them only: optimal portfolio management, currency speculation, and insurance demand. These areas of interest seem to be good examples of problems that are difficult to handle with non-stochastic theories.

## Section A
## Portfolio Theory

Judging by the number of articles and books that were written after the fundamental studies of MARKOWITZ (1952a, 1970) and TOBIN (1958), portfolio theory is nowadays the most important field in risk theory. It therefore seems well worth-while investigating whether the portfolio theory can be integrated into our multiperiod model and, if so, which behavioral implications can be derived for the portfolio holder.

### 1. *The Decision Problem*

The aim of portfolio theory is to find rules by which an amount of capital $a$ available for investment should be distributed among different assets. Such assets are, for example, company shares and bonds. As shares bring in uncertain dividend payments and are subject to the risk of price fluctuations, the determination of the optimal portfolio structure is a problem of risk theory.

In line with our basic model, it is assumed that decisions are made period by period. At the point in time when a decision is made, the level of wealth available after period consumption has been subtracted is

invested in the best possible way. The portfolio structure is then maintained for a period and, at the end of that period, a decision is made, on the basis of the then available capital, about how much to consume in the next period and how to invest the remaining capital. This process is continued in subsequent periods until the planning horizon is reached. In the absence of transactions costs, it is useful to imagine that the total portfolio is sold at the end of each period so that the total wealth accumulated is available for reinvestment and consumption.

The approach just described does not need assumptions concerning the length of the periods. It is therefore suitable for representing the decision problem of the speculator who is frequently revising his portfolio as well as for the small saver who only bothers about his shares every few years.

For a further specification of the approach it is assumed that there are one safe and $n$ risk-bearing assets. A unit of money invested in the safe asset contributes to end-of-period wealth the amount $Q^s = q^s$ and a unit of money invested in the $j$th risk-bearing asset contributes the amount $Q_j^r$. Thus the variables $q^s$ and $Q_j^r$ are effective return factors generally defined as

$$(1) \qquad \left.\begin{matrix} q^s \\ Q_j^r \end{matrix}\right\} \equiv \frac{\text{final price + dividend (or interest)}}{\text{initial price}}, \quad j = 1, \ldots, n.$$

Let the proportion of capital safely invested be $\alpha^s$, let the proportion invested in all the risk-bearing assets be $\alpha^r = 1 - \alpha^s$, and let the proportion of the $j$th risky asset in the total amount invested in risky assets be $\alpha_j^r$, $\sum_{j=1}^{n} \alpha_j^r = 1$. Then the standard risk project utilized in the multi-period approach that was studied in chapter IV B is[1]

$$(2) \qquad Q = \alpha^s q^s + \alpha^r \sum_{j=1}^{n} \alpha_j^r Q_j^r$$

and according to equation (IV B 2) end-of-period wealth $V$ is

$$(3) \qquad V = a\left[\alpha^s q^s + \alpha^r \sum_{j=1}^{n} \alpha_j^r Q_j^r\right].$$

The proportions $\alpha^r$, $\alpha_j^r$, and $\alpha^s$ must be positive. Since wealth is defined so as to include human capital, $\alpha^s \geq 0$ means that it is possible for the decision maker to raise money on his human capital and to invest it in risk-bearing assets.

[1] Indices of time are omitted. It is assumed that $a$ refers to the beginning and $V$, $q^s$, and $Q_j^r$ $\forall j$ belong to the end of the current period.

In order to integrate the portfolio model into the multiperiod optimization approach studied in chapter IV, some of the assumptions underlying that approach have to be considered[2]. With (2) and (3), the basic assumption of stochastic constant returns to scale clearly is satisfied. Another assumption is that the opportunity set contains at least one alternative that, with certainty, brings about strictly positive wealth. Since

$$q^s > 0 \tag{4}$$

and since the assets are typically of the limited-liability variety[3],

$$Q_j^r \geq 0, \quad j = 1, \ldots, n, \tag{5}$$

the set of all those portfolios, for which $\alpha^s > 0$, brings about $Q > q_{\min} > 0$. Hence this second assumption is clearly satisfied. A further assumption is that the standard risk projects at two different points in time are stochastically independent. Whether this condition is satisfied primarily depends on whether stock prices perform *random walks* where the relative changes in prices up to the end of the current period are independent of the levels of prices at the beginning of the period. At least as an idealization, this independence assumption should be justifiable in the light of the fact that, as a rule, the undesired autocorrelation in the share price movement can be removed simply by increasing the length of the periods[4].

Thus the portfolio problem has been integrated into the basic approach developed previously. The task of the decision maker, therefore, is to determine the portfolio structure so that it satisfies the aim

$$\max_{\{\alpha^s, \alpha^r; \alpha_1^r, \ldots, \alpha_n^r\}} \{E[U(a(\alpha^s q^s + \alpha^r \sum_{j=1}^{n} \alpha_j^r Q_j^r))]\}. \tag{6}$$

---

[2] Assumptions (2) and (3) that were made at the beginning of chapter IV B are assumed to hold.

[3] An exception is, for example, the so-called 'Kux' issued by mining companies in Germany. Although the case $Q_j^r \gtreqless 0$ is covered, in principle, by our basic approach, for the sake of brevity it is not discussed here.

[4] For German share prices CONRAD and JÜTTNER (1973) found an autocorrelation in asset price changes, though in absolute ones. RONNING (1974a,b) succeeded in demonstrating that the extremely short period (one day) chosen by these authors was responsable for the result. The finding of Conrad and Jüttner is not compatible with similar American studies either, as the authors themselves concede. See the review article by FAMA (1965).

Here $U(.)$ is one of the Weber functions specified by the time-dependent measure of relative risk aversion $\varepsilon$.

In principle, $U(.)$ may also represent the preference structure for gross distributions of wealth as implied by the BLOOS rule. However, since $\alpha^s \geq 0$, the gross distributions coincide with their net counterparts in the present problem and the BLOOS rule is not operative[5].

## 2. *On the Applicability of the $(\mu, \sigma)$ Approach for Portfolio Management*

To investigate the implications of (6) it will be helpful to approach the problem of optimizing expected utility indirectly by employing the $(\mu, \sigma)$ diagram as suggested by Markowitz and Tobin. Before doing this, however, some remarks concerning the applicability of the $(\mu, \sigma)$ approach, that go beyond what has previously been said in this book, are appropriate[6].

The $(\mu, \sigma)$ approach is perfectly suitable for portfolio analysis only if all attainable distributions belong to the same linear class. The distribution classes $V$ or $Q$ defined by (3) thus have to be invariant with respect to changes in the portfolio structure as described by $\alpha^s$, $\alpha^r$, and $\alpha_1^r, \ldots, \alpha_n^r$. The only distribution class of the $Q_j^r$'s that satisfies this requirement and that is characterized by a finite variance is, because of its reproduction property[7], the class of normal distributions. However, a normal distribution for $Q_j^r$ is excluded by (5). Thus, obviously, a perfect precision cannot be achieved. To draw from this difficulty the conclusion that the $(\mu, \sigma)$ approach is not applicable at all to the portfolio problem would nevertheless be a mistake. The $(\mu, \sigma)$ approach is able at any rate to *approximate* the expected-utility approach.

First we may refer to the method of local approximation which, as shown before[8], is applicable if $E(Q)/2 \leq Q \leq 2E(Q)$. This condition can easily be interpreted for the sake of portfolio analysis if it is written as

$$(7) \qquad \tfrac{1}{2}[\alpha^s q^s + (1-\alpha^s)E(Q^r)] \leq \alpha^s q^s + (1-\alpha^s)Q^r \leq 2[\alpha^s q^s + (1-\alpha^s)E(Q^r)]$$

---

[5] Without the constraint $\alpha^s \geq 0$, in the case of weak risk aversion ($\varepsilon < 1$) it could well happen that in the optimum $\alpha^s < 0$. If, however, $\varepsilon \geq 1$ then, because of $\lim_{v \to 0} U(v) = -\infty$, it is never optimal to set $\alpha^s \leq 0$ if $Q_j^r$ may take on the variate zero with a strictly positive probability.

[6] Cf. chapter II A 3, A 6, and D 2.2–D 2.4.

[7] Cf. fn. 7 in chapter II A.

[8] In chapter II D 2.2.2.

where

$$(8) \qquad Q^r \equiv \sum_{j=1}^{n} \alpha_j^r Q_j^r.$$

Because of (5) the first two terms in this inequality imply the condition

$$(9) \qquad \alpha^s \geq \frac{1}{1+\frac{q^s}{E(Q^r)}}.$$

Since, in practice, $q^s$ and $E(Q^r)$ are rather close to one another, condition (9) requires that around 50% at least of the capital must be safely invested to ensure that the wealth distribution does not extend below the lower boundary of the approximation range[9]. Consider now the upper boundary of the approximation range. Since there is no clear-cut upper boundary of $Q_j^r$, the second inequality in (7) might seem to be more of a problem. To check this, transform this inequality to

$$(10) \qquad Q^r \leq \frac{1}{\frac{1}{\alpha^s}-1} q^s + 2E(Q^r).$$

Inserting the smallest admissible value fo $\alpha^s$ compatible with (9) we find that (10) reduces to

$$(11) \qquad Q^r \leq 3E(Q^r).$$

Provided that $E(Q^r)$ is close to unity and provided that condition (9) is met, this condition, roughly speaking, requires the decision maker to believe that it is impossible for the value of his risk portfolio to more than treble.

Apart from local approximation which is possible for arbitrary distribution classes, in the case of portfolio analysis, the $(\mu, \sigma)$ principle can be legitimated on yet another basis, one that was already mentioned in chapter II D 2.4. This is that the end-of-period wealth distributions attainable by alternative portfolio structures approximately seem to belong to the same linear class. Indeed, empirical data suggest that, with

[9] If the maximum capital loss considered possible by the decision maker is 50% rather than 100% then, even in the case $\alpha^s = 0$, the lower boundary of the approximation range is not binding.

the class of normal distributions, such a class prevails. FISHER and LORIE (1970) reported frequency distributions of returns for fictitious portfolios of well-known assets. Comparing these distributions with the normal distribution, MOSSIN (1973, pp. 60–62) found a high degree of coincidence. The degree of approximation was higher the larger the number of different assets in the portfolio. From a theoretical point of view, this result is to be expected when the various assets bring about returns that are stochastically independent of one another. According to (3) and (8), end-of-period wealth $V$ and the weighted return factor $Q^r$ are sum variables to which the Central Limit Theorem is applicable[10]. It implies that

$$\lim_{\substack{n\to\infty \\ \alpha_j^r\to 0}} \sum_{j=1}^{n} \alpha_j^r Q_j^r \qquad \forall j = 1, \ldots, n \tag{12}$$

is normally distributed regardless of the distributions of the $Q_j^r$'s provided that their variances exist. But of course, in reality, the independence assumption can hardly be satisfied in a strict sense. It is therefore remarkable that the normal distribution nevertheless turned out to be a good approximation empirically.

Despite these encouraging aspects, some doubts remain. Whatever the degree of similarity between the empirical distributions and the normal distribution, these types of distribution differ from one another definitely with respect to their left tails: empirical distributions are bounded to the left where $v = a\alpha^s$, but the normal distribution is unbounded. It is not obvious that this divergence is negligible, since the concavity of the utility function implies that a given difference between the two types of distribution affects expected utility more strongly when it occurs at the left tails than when it occurs at the right. The bias in expected utilities is stronger the higher the degree of risk aversion and the higher the variances of the distributions. The assumption of a normal distribution is completely misleading when relative risk aversion ($\varepsilon$) is equal to or above unity. Since, in this case, there is no lower bound to the utility function, the normality assumption would imply lexicographic pseudo indifference curves in the form of rays through the origin even though it is impossible for gross wealth to become zero or negative. On the other hand, in the analysis of the time dependence of risk aversion and of the demand for liability insurance, the possibility $\varepsilon < 1$ turned out to be the realistic one. This rescues the normality

[10] Cf. footnote 22 in chapter II A.

assumption from the implausible implication pointed out. Thus, after all, the similarity between the empirical distributions of portfolio returns and the normal distribution does suggest that the distributions among which the portfolio holder has to choose can be idealized as belonging to the same linear class.

For this reason and also because of the possibility of a distribution-free local approximation, the $(\mu, \sigma)$ approach seems to be appropriate for an analysis of practical portfolio problems. We therefore may replace (6) by

$$\max_{\{\alpha^s, \alpha^r; \alpha_1^r, \ldots, \alpha_n^r\}} U(\mu, \sigma) \tag{13}$$

where the function $U(\mu, \sigma)$ describes a system of convex indifference curves, as was derived in III A 2.2. For the needed distribution parameters we calculate

$$\begin{aligned} \mu \equiv E(V) &= a[\alpha^s q^s + \alpha^r \sum_{j=1}^{n} \alpha_j^r E(Q_j^r)] \\ &= a[\alpha^s q^s + \alpha^r E(Q^r)], \end{aligned} \tag{14}$$

$$\begin{aligned} \sigma \equiv \sigma(V) &= a\alpha^r \sqrt{\sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i^r \alpha_j^r \varrho_{ij} \sigma(Q_j^r) \sigma(Q_i^r)} \\ &= a\alpha^r \sigma(Q^r), \end{aligned} \tag{15}$$

where

$$\varrho_{ij} \equiv \frac{\operatorname{cov}(Q_i^r, Q_j^r)}{\sigma(Q_i^r)\sigma(Q_j^r))} \tag{16}$$

is the coefficient of correlation between assets $i$ and $j$.

### 3. *Implications of an Optimal Portfolio Structure*

#### 3.1. *The Advantage of Diversification*

Suppose the decision maker is risk neutral so that $U(.)$ is linear ($\varepsilon = 0$). Then (6) and (13) can be written in the form[11]

$$\max a[\alpha^s q^s + \sum_{j=1}^{n} \alpha^r \alpha_j^r E(Q_j^r)]$$

[11] This result depends on the institutional constraint $\alpha^s \geq 0$. If $\alpha^s$ is unconstrained, i.e., if borrowing beyond the value of human capital is allowed, the BLOOS rule appears on the scene and it may very well turn out that a risk neutral investor behaves as if he were a risk lover.

and it turns out that the best strategy is to invest all capital in the asset with the highest expected return factor $E(Q_j^r)$ or $q^s$ respectively. It should not be necessary to stress, however, that both the assumption and its implication are highly unrealistic. Therefore, the role of risk aversion will now be checked.

For this purpose, it is useful to represent the opportunity set of end-of-period wealth distributions, as described by (14) and (15), in a $(\mu, \sigma)$ diagram. Suppose first that no capital is invested safely so that $\alpha_s = 1 - \alpha_r = 0$. A variation of the proportions $\alpha_j^r$ may then bring about an opportunity set[12] like the one shown on the right in Figure 1. Choose now the other extreme and set $\alpha^s = 1 - \alpha_r = 1$; then the structure of risk assets is, of course, irrelevant and we find a degenerated distribution with $E(V) = a\,q^s$ and $\sigma(V) = 0$, which in the diagram is represented by point $P$. Provided with this information, the opportunity set achievable when $\alpha^s$ and $\alpha_j^r\ Vj$ are freely choosable can easily be found. Equations (14) and (15) show that we need only form linear combinations of the coordinates of the points in the opportunity set mentioned first and those of point $P$. Figure 1 illustrates how, for example, $A$ changes to $A'$, $B$ to $B'$, and $C$ to $C'$.

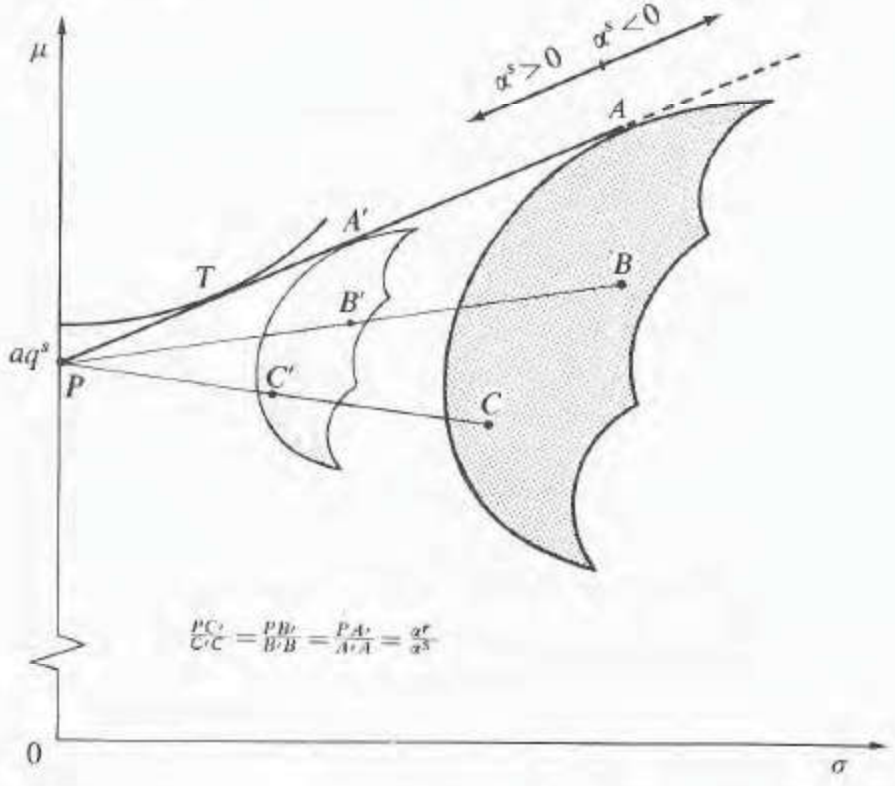


*Figure* 1

[12] It can be shown that, for $n \to \infty$, the left-hand border of the opportunity set approaches a hyperbola open to the right. See ROY (1952). For our purposes, the particular shape of the opportunity set is irrelevant.

In Figure 1 there is a line starting at point $P$ which is tangent to the opportunity set prevailing when $\alpha^s = 0$. This tangent characterizes the efficiency frontier of the total opportunity set achievable when $\alpha^s$ and $\alpha_j^r \; \forall j$ are variable. Since the points on the efficiency frontier differ from one another with respect to the share $\alpha^r$ of the 'risk portfolio' in the total portfolio but not with respect to the structure of the risk portfolio, the optimization process can be divided into two steps. First, independently of the particular aspects of the portfolio holder's preferences, the optimal structure $\alpha_1^r, \alpha_2^r, \ldots, \alpha_n^r$ of the risk portfolio is determined. Then, with $\alpha^s$ and $\alpha^r$, the division of funds between the safe asset and the risk portfolio is determined. This is a well-known result of TOBIN (1958) that is usually referred to as the *Separation Theorem*.

For convex indifference curves exhibiting risk aversion, we now find that the strategy of investing all funds in the asset with the highest expected return is, in general, suboptimal. If risk aversion is sufficiently large, i.e., if the indifference curves are sufficiently curved there will be a tangency solution as illustrated by point $T$ in Figure 1. It shows that the safe asset is held although[13] $E(Q^r) > q^s$. The decision maker is willing to pay a price for safety.

Figure 1 does not provide obvious information on the optimal structure of the risk portfolio. In order to achieve such information, it is useful to formulate (13) as a Lagrange approach,

$$\mathcal{L} = U(\mu, \sigma) + \lambda(1 - \alpha^s - \sum_{i=1}^{n} \alpha^r \alpha_i^r), \tag{17}$$

and to differentiate with respect to the proportions of the various assets in the portfolio:

$$\frac{\partial \mathcal{L}}{\partial \alpha^s} = \frac{\partial U}{\partial \mu} \frac{\partial \mu}{\partial \alpha^s} - \lambda = 0 \tag{18}$$

$$\frac{\partial \mathcal{L}}{\partial (\alpha^r \alpha_k^r)} = \frac{\partial U}{\partial \mu} \frac{\partial \mu}{\partial (\alpha^r \alpha_k^r)} + \frac{\partial U}{\partial \sigma} \frac{\partial \sigma}{\partial (\alpha^r \alpha_k^r)} - \lambda = 0 \qquad \forall k = 1, \ldots, n. \tag{19}$$

This yields the rule

$$\frac{\partial \mu}{\partial (\alpha^r \alpha_k^r)} - \frac{\partial \mu}{\partial \alpha^s} = - \frac{\dfrac{\partial U}{\partial \sigma}}{\dfrac{\partial U}{\partial \mu}} \frac{\partial \sigma}{\partial (\alpha^r \alpha_k^r)} \qquad \forall k = 1, \ldots, n. \tag{20}$$

[13] It can be shown that, with risk averse investors, the capital market equilibrium is such that $E(Q^r) > q^s$ and that the point $T$ characterizes the situation of the typical investor.

Since

$$\left.\frac{d\mu}{d\sigma}\right|_{U(\mu,\sigma)} = -\frac{\frac{\partial U}{\partial \sigma}}{\frac{\partial U}{\partial \mu}}, \tag{21}$$

equation (20) implies

$$\frac{\partial \mu}{\partial(\alpha^r \alpha_k^r)} - \frac{\partial \mu}{\partial \alpha^s} = \left.\frac{d\mu}{d\sigma}\right|_{U(\mu,\sigma)} \frac{\partial \sigma}{\partial(\alpha^r \alpha_k^r)} \qquad \forall k = 1, \ldots, n. \tag{22}$$

This equation compares the additional 'return' of an increase in the $k$th risk-bearing asset with its additional 'cost'. If the share of the safe asset is reduced by one percentage point in favor of the $k$th risk-bearing asset, then expected end-of-period wealth rises by one percent times $\partial\mu/\partial(\alpha^r\alpha_k^r) - \partial\mu/\partial\alpha^s$; this is the 'return'. The 'cost' of this portfolio restructuring on the one hand depends on how the standard deviation is affected by an increase in the portfolio share of the $k$th risk-bearing asset, $\partial\sigma/\partial(\alpha^r\alpha_k^r)$, and, on the other, on the 'price' of an additional unit of standard deviation, $d\mu/d\sigma\,|_{U(\mu,\sigma)}$. This 'price' can be interpreted in two ways. First, it indicates how much it is necessary for the expected level of end-of-period wealth to increase in order to compensate for an increase in risk. This interpretation, though correct, leads to the wrong conjecture that the size of $d\mu/d\sigma\,|_{U(\mu,\sigma)}$ depends on the decision maker's personal preference. In fact, this is not the case. Instead, $d\mu/d\sigma\,|_{U(\mu,\sigma)}$ equals the slope of the efficiency frontier[14], i.e., the maximum value of $[E(Q^r) - q^s]/\sigma(Q^r)$, to which it is adapted by a suitable variation of $\alpha^s$. Thus, an interpretation in terms of opportunity costs seems preferable: $d\mu/d\sigma\,|_{U(\mu,\sigma)}$ measures the increase in expected end-of-period wealth brought about by a structure-maintaining widening of the total risk portfolio that increases the standard deviation by one unit.

To find out what (22) implies for the size of $\alpha_k^r$, the single items are explicitly calculated from (14) and (15):

$$\frac{\partial \mu}{\partial \alpha^s} = a q^s, \tag{23}$$

[14] The fact that this slope depends on the preference of *all* wealth owners does not contradict the assumption that it is considered as given in the individual optimization problem.

$$\frac{\partial \mu}{\partial(\alpha^r \alpha_k^r)} = aE(Q_k^r), \tag{24}$$

$$\frac{\partial \sigma}{\partial(\alpha^r \alpha_k^r)} = \frac{a^2/2}{\sigma(a\alpha^r Q^r)} 2 \sum_{i=1}^{n} \alpha^r \alpha_i^r \varrho_{ik} \sigma(Q_k^r)\sigma(Q_i^r) \tag{25}$$

$$= a \frac{\sum_{\substack{i=1 \\ i \neq k}}^{n} \alpha_i^r \varrho_{ik} \sigma(Q_k^r)\sigma(Q_i^r) + \alpha_k^r \sigma^2(Q_k^r)}{\sigma(Q^r)}.$$

If these values are inserted into (22) the following expression results:

$$\alpha_k^r = \frac{\dfrac{E(Q_k^r) - q^s}{\sigma(Q_k^r)}}{\left.\dfrac{d\mu}{d\sigma}\right|_{U(\mu,\sigma)}} \frac{\sigma(Q^r)}{\sigma(Q_k^r)} - \frac{\sum_{\substack{i=1 \\ i \neq k}}^{n} \alpha_i^r \varrho_{ik} \sigma(Q_k^r)\sigma(Q_i^r)}{\sigma^2(Q_k^r)} \quad \forall k = 1, \ldots, n. \tag{26}$$

Although it does not give an explicit solution for $\alpha_k^r$, this expression nevertheless can be meaningfully interpreted[15].

If we set all coefficients of correlation with $i \neq k$ equal to zero, the second term on the right-hand side of the equation disappears, so that only the first matters. Within this term, the first quotient relates a 'price of risk' $E(Q_k^r - q^s)/\sigma(Q_k^r)$, specific to asset $k$, to the average 'price of risk' $d\mu/d\sigma \mid_{U(\mu,\sigma)}$ of the total portfolio. Assuming that, because of $d\mu/d\sigma \mid_{U(\mu,\sigma)} > 0$, risk assets are held at all and taking into account that, by definition, $\sigma(Q^r)/\sigma(Q_k^r) > 0$ we find $E(Q_k^r - q^s)/\sigma(Q_k^r) > 0$ to be a necessary and sufficient condition for $\alpha_k^r > 0$. This is the most important result of Markowitzian portfolio theory. It implies that it is not only the asset with the highest expected return that enters the risk portfolio. On the contrary, *all* risky assets that promise a higher expected return than the safe asset are included.

The result changes drastically if coefficients of correlation other than zero are allowed. If the second term on the right-hand side of (26), which measures the correlation between asset $k$ and the remaining assets, is strictly positive, then the expected return factor $E(Q_k^r)$ must exceed the return factor $q^s$ of the safe asset by a sufficient amount if asset $k$ is to be included in the portfolio. But if the $k$th asset has the rare property of being negatively correlated with the rest of the portfolio,

---

[15] Equation (26) does not give an explicit solution since $\sigma(Q^r)$ and $d\mu/d\sigma \mid_{U(\mu,\sigma)}$ depend on $\alpha_k^r$.

then, because it counteracts the other risks, this asset may be demanded even if $E(Q_k^r) < q^s$. Thus we can conclude that, under fairly general conditions, a well-diversified portfolio is held. Therefore here too, putting all your eggs in one basket, which is wise under risk neutrality, turns out to be suboptimal.

### 3.2. *The Age Dependence of the Optimal Portfolio Structure*

As we know, according to our preference hypothesis, with the passage of time, the degree of relative risk aversion ($\varepsilon$) relevant for current decision making approaches the value of unity, provided the decision maker's propensity to consume out of wealth is strictly positive. The implications of such a change in risk aversion on the optimal portfolio structure can be illustrated in a comparative static analysis where a given opportunity set of end-of-period wealth distributions is assumed.

Equations (III A 52) and (III A 53) show that on a given ray through the origin the indifference-curve slope rises with $\varepsilon$:

$$\left. \frac{d \left. \frac{d\mu}{d\sigma} \right|_{U(\mu,\sigma)}}{d\varepsilon} \right|_{\frac{\mu}{\sigma}} > 0.$$

Thus, an increase in risk aversion shifts the point of tangency between an indifference curve and the opportunity set to the left and hence induces a process of restructuring the portfolio towards the safe asset and away from the risky assets. A decrease in risk aversion has the opposite effect. In Figure 2 both cases are illustrated, account being taken of the fact that the direction of a change in risk aversion depends on its absolute level. The point of tangency between the indifference curve labelled $\varepsilon = 1$ and the efficiency frontier $AA'$ of the opportunity set is a trough. If initially the point of tangency is below this then, because of $\varepsilon > 1$, with the passage of time it moves upward, and if it is above, it moves downward. This is illustrated by arrows. Thus a person, who, compared to point $P$, initially invests a high proportion of his wealth in the risk-bearing assets, over time restructures his portfolio in favor of the safe asset and a person, who starts off with a high proportion of the safe asset, increases his demand for risk-bearing assets as he grows older.

A superficial interpretation of this result would be that a portfolio owner who holds a proportion of the safe asset greater than this asset's share in the market portfolio will, with the passage of time, be inclined to reduce this proportion. Such an interpretation, however, would only be admissible if $\varepsilon = 1$ indicated a normal or average level of risk aver-

sion. But there is no reason to suppose this. On the contrary, a relative risk aversion increasing with time seems to be the normal case. Thus, the capital market equilibrium will be such that the market portfolio occurs at a point to the left of point $P$, that is, in the range where indifference curves with $\varepsilon<1$ are tangent to the efficiency frontier. Even investors who demand a proportion $\alpha^s$ of the safe asset that is below the share of safe assets in the market portfolio may then tend to increase this proportion as they grow older.

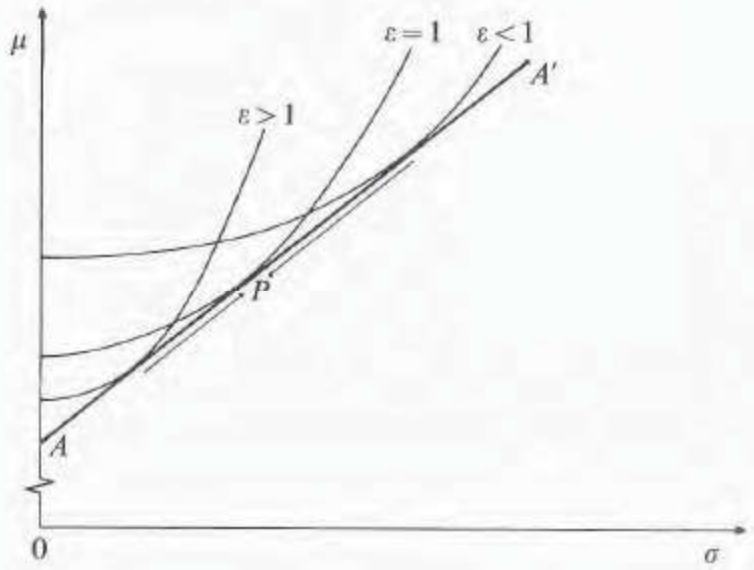


*Figure* 2

The age dependence of the optimal portfolio structure has important implications for the capital-market equilibrium when the age structure of wealth owners is changing. For example, an increase in the average life span will increase the demand for safe assets relative to risky assets and will therefore imply a relative fall in share prices to ensure that the market portfolio is willingly held by the public.

This will not be without implications for productive decisions on the part of firms. Firms maximizing the market value of their shares will try to avoid this fall in share prices by engaging in less risky production decisions. Unfortunately, however, this means accepting a reduction in excepted profits since production possibilities that promise both a lower risk and a higher or even the same level of expected return are not available. If they were available the firms would surely have chosen them *before* the fall in share prices caused by the change in age structure. Thus, a general decrease in labor and capital productivities seems unavoidable. The underlying cause is that, with a change in the age structure, households supply less of a factor of production. Despite its being

neglected in recent economic theory, this factor seems to be of great practical importance – we refer to the factor called 'risk'[17].

### 3.3. *The Wealth Independence of the Optimal Portfolio Structure*

A peculiarity of the Weber functions, that we used continually in the above multiperiod analysis[18], is the *separation property* detected by PYE (1967):

$$\max E[U(aQ)] \sim \max E[U(Q)]\,|_a. \tag{27}$$

It implies that the optimal standard risk project can be found independently of the amount of capital invested. Given the wealth owner's goal as described by (6), this means that the optimal portfolio structure is independent of his wealth. The result holds for arbitrary shapes of the gross or balance sheet distributions of wealth. Neither the assumption $Q_j^r \geq 0$ nor the conditions for an application of the $(\mu, \sigma)$ approach are needed.

If we are allowed to use the $(\mu, \sigma)$ approach, however, then the separation property can easily be illustrated as in Figure 3. There, the efficiency frontiers of two opportunity sets, brought about by two different levels $a$ and $a'$, $a' > a$, of capital invested, are plotted. Since in (14) and (15) the level of capital appears as a factor of proportionality, one efficiency frontier can be produced from the other by means of a projection through the origin. The indifference curves following from Weber's law can also be constructed in such a way. Thus the points of tangency $P$ and $P'$ have to lie on a ray through the origin that partitions the efficiency frontiers in the proportion $\alpha^r : \alpha^s$ so that the proportion of wealth invested in the risk portfolio is independent of wealth. Of course, also the structure of the risk portfolio itself is unchanged, for this structure is in any case independent of personal preferences and would be maintained even if $P$ and $P'$ were not situated on the same ray through the origin.

The wealth independence of the portfolio structure is a very plausible result. For want of something better, it provided HICKS (1967, p. 114) with an incentive for postulating a homothetic indifference-curve system. Nevertheless, it seems to be at variance with an obvious empirical fact that we now want to consider.

---

[17] For the concept of risk as a factor of production see, e.g., HICKS (1931) and PIGOU (1932, pp. 771 ff.).

[18] Cf. chapter IV B 1.2 and 2.2.

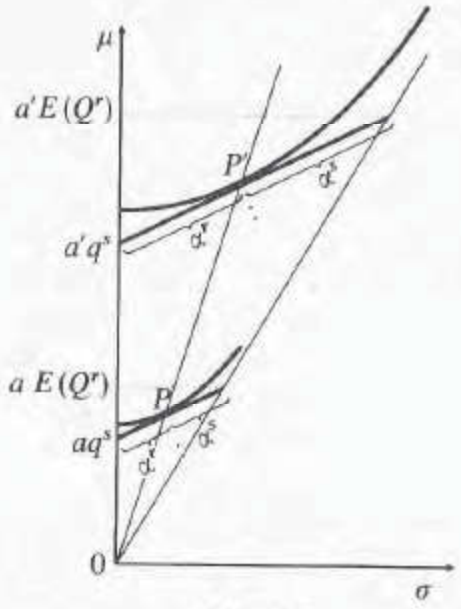


*Figure* 3

### 3.3.1. The Apparent Rejection of the Relativity Axiom by the Observation of a Decreasing Velocity of Money Circulation

Suppose the safe asset considered above is money so that, with constant commodity prices, we have to set $q^s = 1$. Then, the wealth independence of the portfolio structure implies that the partial elasticity of money holding is unity.

This implication seems to contradict reality, at least this is what ARROW (1965, p. 44; 1970, pp. 103 f.) believes. He interprets the empirical investigations into the demand for money in the United States carried out by SELDEN (1956), FRIEDMAN (1959), LATANÉ (1960), and MELTZER (1963) by saying that they 'agree in finding a wealth elasticity of demand for cash balances of at least 1' and he maintains that this evidence strongly supports his own hypothesis of *increasing* relative risk aversion[19]. Although from a theoretical point of view Arrow's hypothesis did not seem to be particularly convincing, it now appears as if, after all, it is better on empirical grounds than our hypothesis of *constant* relative risk aversion. However, the appearance is deceptive, for a closer look reveals that there is not very much to the empirical evidence. The proof of this contention can easily be given.

First, it must be mentioned that only in one of the above empirical investigations (Meltzer) is the *wealth* elasticity of cash demand measured. In all the others, cash demand is assumed to depend on

[19] ARROW (1965, pp. 37–44, and 1970, pp. 98–104) explains the demand for money by a direct use of the expected-utility approach and assumes there are two assets, one safe and one risk-bearing. Since in this case, all distributions of the opportunity set belong to the same linear class, the solution in the $(\mu, \sigma)$ diagram that was first derived by TOBIN (1958) is identical with the direct solution found by Arrow. Cf. chapter II D 2.3.

*income*, although this is intended to stand in for wealth. The assumptions are only identical if wealth and income grow proportionally, that is, if the capital-output ratio in the economy does not change. If, in the process of economic growth, this ratio is increasing then we would already have an explanation for the income elasticity of money demand being greater than unity, i.e., for the falling velocity of money circulation[20].

Let us, however, put these problems aside and take for granted the fact that there has been a secular rise in the money-wealth ratio. Can we then conclude that the empirical evidence supports Arrow's hypothesis? The answer is no.

Unlike Selden and Friedman, Meltzer and Latané consider interest rates as well as wealth as explanatory variables for the demand for cash. Their findings clearly suggest that a secular fall in interest rates is the true explanation of the rise in the money wealth ratio. The *partial* income or wealth elasticity of cash demand found by Meltzer is slightly above unity but does not, according to him, significantly deviate from this value. Latané even interprets his results as clearly supporting a partial elasticity of unity. Meanwhile, a value of unity has also been found for countries other than the United States. For West Germany, for example, KÖNIG (1968), WOLL (1969), WESTPHAL (1970, pp. 51–77), and MATTFELDT (1973, pp. 128–154) all provided evidence for the hypothesis of a unitary wealth elasticity, although they found a much lower interest elasticity than Latané and Meltzer did.

While these findings seem to support our hypothesis rather than Arrows's, Arrow suggests yet another argument in his favor. It refers to the fact that, apart from the portfolio motive, there is a transactions motive for money holding. The inventory-theoretic approach to money demand as formulated by BAUMOL (1952) and TOBIN (1956) implies that the elasticity of cash demand with respect to the transactions volume is 1/2 if the cost of exchanging money and interest-bearing assets is constant[21]. If the transactions volume is in fixed proportion to wealth, then this result seems to indicate that the part of the wealth elasticity of cash demand explained by the portfolio motive must be above 1 to ensure

---

[20] It is true, there does not seem to be any clear-cut empirical evidence for a rise in the capital-output ratio. On the other hand, as shown in SINN (1975, pp. 683–690), if the government sector is increasing relative to the private sector, if there is a process of vertical integration of private firms, and/or if the economy's rate of growth is slowing down, then there are clear effects that imply that the rise in the *measured* capital-output ratio understates the actual rise.

[21] VIKAS (1975) even contends that there is a negative partial wealth elasticity of cash demand. The unrealistic assumption that the interest payments on wealth come just at that point in time when transactions are to be carried out seems to be crucial for this result.

that, on balance, an elasticity of unity is brought about. But this, too, is not a particularly strong conclusion. The constancy over time of the cost of exchanging money for interest-bearing assets can be questioned. If, in a situation of balanced growth with a constant population, the wage rate rises in strict proportion to labor productivity and if the cost of exchange is purely a labor cost, for example the trip to the bank, then the Baumol-Tobin model implies that the cash-wealth ratio stays constant over time.

Suppose, once more, that we ignore this objection and accept Arrow's argument. Does it then follow that relative risk aversion is an increasing function of wealth? It does not. At best, we can conclude from an increase in the money-wealth ratio that relative risk aversion has been increasing, but the reason for its increase cannot be inferred. Arrow thinks the rise in wealth was responsible. The time dependence of the optimal portfolio structure analyzed in the last section suggests instead a different explanation. Since life expectancy has increased significantly during this century, the degree of relative risk aversion and hence the relative demand for safe assets must have been rising, provided, as was argued before, people in general have a relative risk aversion below unity.

Up to now, the explanation of money holding as being an attempt to reduce portfolio risk has not been questioned. Following an objection raised by STIGLITZ (1969b), WESTPHAL (1970, p. 18), SHELL (1972), and others, we now make good this omission.

Assume that, beside money, there is an interest bearing safe asset available to the decision maker. Let $\alpha_1^s$, $\alpha_2^s$, and $\alpha^r$ denote the proportions in the portfolio of money, the interest bearing asset, and the risk portfolio respectively, and let $q_1^s$, $q_2^s$, and $Q^r$ be the corresponding return factors where $1 = q_1^s < q_2^s < E(Q^r)$. Then, analogously to (14) and (15), the parameters of the attainable end-of-period wealth distributions are

$$E(V) = a[\alpha_1^s + \alpha_2^s q_2^s + \alpha^r E(Q^r)], \tag{28}$$

$$\sigma(V) = a\alpha^r \sigma(Q^r). \tag{29}$$

In these equations the structure of the risk portfolio does not show up explicitly. In the reasoning that follows any arbitrary, but fixed, structure can be assumed, including the optimal one. To determine the optimal proportions $\alpha_1^s$, $\alpha_2^s$, and $\alpha^r$, consider the triangular opportunity set depicted in Figure 4. As indicated in this figure, this set can be constructed by forming linear combinations of the coordinates

$$\begin{array}{llll} E(V) = a, & \sigma(V) = 0, & \text{if} & \alpha_1^s = 1, \\ E(V) = aq_2^s, & \sigma(V) = 0, & \text{if} & \alpha_2^s = 1, \\ E(V) = aE(Q^r), & \sigma(V) = a\sigma(Q^r), & \text{if} & \alpha^r = 1. \end{array}$$

Since the upper boundary of this opportunity set is an efficiency frontier, it turns out that in the optimum $\alpha_1^s = 0$. No cash is demanded for portfolio purposes.

The result holds for any arbitrarily given structure of the risk portfolio. Since, given this structure, all attainable distributions $V = \alpha_1^s a + \alpha_2^s q_2^s a + \alpha^r Q^r a$ belong to the same linear class, the $(\mu, \sigma)$ approach perfectly represents the expected-utility rule[22] provided that the indifference-curve system which belongs to the corresponding class is consulted. The conclusion is that, even when the structure of the risk portfolio is subject to choice, it is never optimal to hold money for portfolio purposes. Thus Arrow's attempt to find an empirical basis for his hypothesis that relative risk aversion decreases with a rise in wealth must be considered to be a failure.

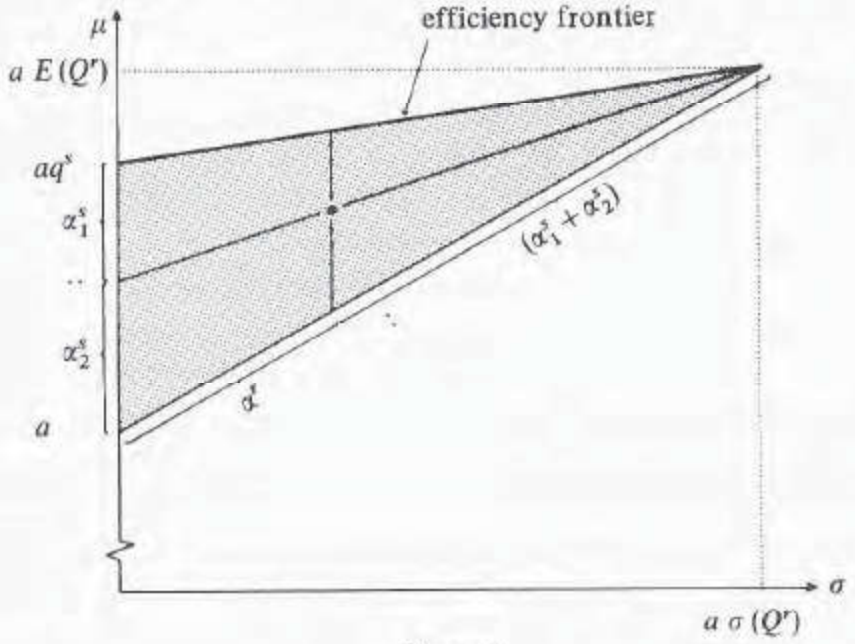


*Figure* 4

### 3.3.2. A Risk-Theoretic Wealth Effect of a Government Budget Deficit

In the monetarist-fiscalist debate on the efficacy of anticyclical budgetary policy, the impact of a change in wealth on private commodity demand is of crucial importance. For example, a question that has been vigorously discussed is whether the policy of a *deficit without spending*[23], financed by taking up credit in the private market, has

[22] Because of $Q^r \geq 0$, the distributions of this linear class are bounded to the left at $v_{\min} = E(V) - \underset{\sim}{k}\sigma(V)$ which, because of $v_{\min} = a[\alpha_1^s + \alpha_2^s q_2^s]$, implies that

$$a[\alpha_1^s + \alpha_2^s q_2^s] = a[\alpha_1^s + \alpha_2^s q_2^s + \alpha^r E(Q^r)] - \underset{\sim}{k} a \alpha^r \sigma(Q^r)$$

and hence $\underset{\sim}{k} = E(Q^r)/\sigma(Q^r)$.

[23] A budget deficit brought about by a reduction in taxes.

expansionary effects on private demand. The question could also be phrased 'what happens if government donates bonds to the public', for this is what the described policy really is.

Suppose, as seems realistic, the public suffers from fiscal illusion, which means that those who get the bonds feel richer while the future tax payers do not worry about the increase in their liabilities. Then, according to the multiperiod model of chapter IV B 2, people will increase the consumption levels associated with any given rate of interest and hence economic activity will be stimulated. The monetarist objection to this argument is that the increase in wealth raises the portfolio demand for cash so that contractive forces are brought into operation by way of an increase in the market rate of interest. This objection, however, is not a valid one. First, it is deficient because it uses the portfolio motive to explain money demand. Second, there seems to be an important effect, very similar to the wealth effect on money demand, which operates in the opposite direction.

If we consider goverment bonds as safe assets, then the wealth independence of the optimal portfolio structure, as implied by Weber's law, suggests that people are not willing to absorb all the additional bonds into their portfolios as long as the structure of returns is unchanged. Rather, they try to exchange some of them for other assets, not, as monetarists are wont to contend, for money, but for risk-bearing assets. Thus, for any given rate of interest $q^s - 1$ of the safe asset, share prices go up and hence the cost of raising equity capital goes down. This will increase private investment and hence stimulate economic activity.

The result is by no means self-evident but hinges crucially on what kind of hypothesis concerning people's risk preferences is made. This is illustrated in Figure 5. This figure is similar to Figure 4, but refers to the

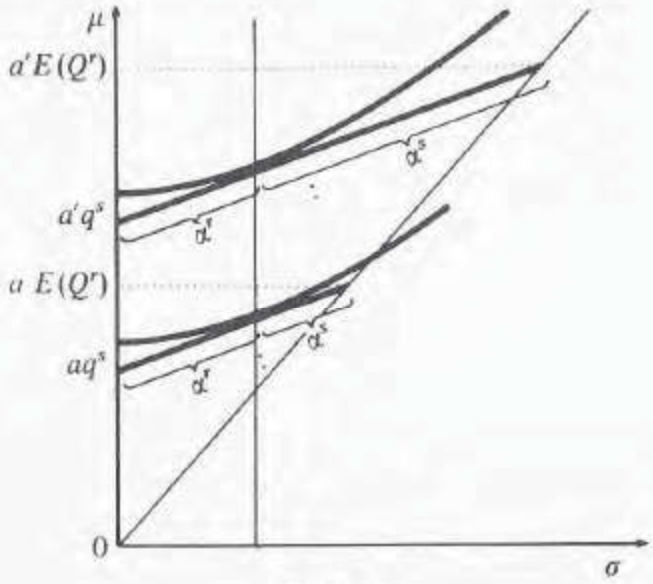


*Figure 5*

indifference-curve system generated by the hypothesis of constant absolute rather than that of constant relative risk aversion.

Since in this indifference-curve system the single curves can be produced from one another by vertical shifts, an increase in wealth obviously implies that the point of tangency between an indifference curve and the efficiency frontier shifts up vertically. This means that, independently of wealth, a given amount $\alpha' a$ is invested in the risk portfolio. The gift of government bonds is fully absorbed into the portfolios without a change in the return structure, and hence the policy of deficit without spending does not directly affect private investment. Weber's law permits this case to be dismissed as irrelevant.

### 4. *Summary*

It was shown that the decision problem of the portfolio optimizer can be integrated perfectly into the previously developed multiperiod approach. Then an attempt was made to find the conditions under which the tool box of the Markowitz-Tobin portfolio theory can be used to calculate approximately the optimal portfolio as indicated by this multiperiod approach. Apart from the standard results of portfolio theory reported for the sake of completeness, the analysis concentrated on the implications that follow from the particular preference hypothesis established in this book.

First an interesting age dependence of the portfolio structure was found. In the normal case of a degree of relative risk aversion below unity, this age dependence implies that an increase in the average age of wealth owners leads to a decrease in the demand for risk-bearing assets which itself is likely to induce a general decline in productivity.

Another important aspect was found to be that the optimal portfolio structure is independent of wealth. This aspect implies that an increase in private wealth leads to a reduction in the cost of equity capital and thus stimulates investment demand. The significance of this effect was demonstrated with reference to the monetarist-fiscalist debate on the efficacy of a policy of deficit without spending.

We discussed fairly thoroughly the empirical evidence of a secularly rising money-wealth ratio that Arrow believed supported his hypothesis of increasing relative risk aversion. It turned out that, even if Arrow's approach to an explanation of money demand is accepted, the empirical results favor our preference hypothesis rather than his. But actually Arrow's theory of money demand itself is not very convincing. As soon as a short-run interest-bearing asset is introduced, this theory fails to explain money holding.

# Section B
# The Theory of Currency Speculation

In this section the theory of economic decision making under uncertainty developed above is applied to the choice problem of the currency speculator. In contrast to portfolio theory, the analysis is only an exercise in positive theory. Since speculators are important people it is worth-while trying to understand their behavior, but the aim of speculation theory is not to show them how to increase their profits.

The analysis draws partly on the approaches of GRUBEL (1966), FELDSTEIN(1968), LELAND (1971), and HOCHGESAND (1974). However, in so far as this study integrates the problem with a multiperiod approach and explores the particular implications of Weber's law and the BLOOS rule, it goes beyond them.

## 1. *The Basic Problems of the Spot and Forward Speculators*

The decision problem of the currency speculators is analyzed in a highly idealized model of currency markets with perfectly flexible exchange rates. It is assumed that all transactions are carried out at fixed dates between which nothing happens. At each date there is a spot and a forward market. In the latter the conditions for an exchange in currency on the subsequent transactions day are settled[1]. The analysis is confined to the two-country case. The domestic country is the United States, the foreign country Germany; accordingly the currencies are \$ and DM. The exchange rate is the dollar price of one Deutschmark.

At the decision date, speculators know the current spot rate $w_0^K$ and the current forward rate $w_0^T$, but the rates that will obtain after one period, $W_1^K$ and $W_1^T$, as well as all other future rates, are unknown. Decisions therefore have to be made on the basis of equivalent objective probabilities.

### 1.1. *Forward Speculation*

The forward speculator buys or sells forward currency, planning to carry out a compensating transaction in the spot market when the delivery date comes and he has to meet his obligations. Consider first the case where he buys, i.e., where he has a long position in currency futures. If he buys $h$ DM 'today' in the forward market then 'to-

[1] The possibility of multiple forward markets, each concerned with a different time in the future, is excluded. A model with multiple forward markets was developed by SOHMEN (1966 and 1973).

morrow' he has to spend \$ $h\ w_o^T$ to meet his obligations, but after exchanging the $h$ DM received through the forward contract he has a dollar revenue of $h\ W_1^K$. His expected profit in dollars after one period therefore is

$$X = h(W_1^K - w_0^T). \tag{1}$$

Next consider the case of a short position. If the speculator 'today' sells $k$ DM in the forward market then he 'tomorrow' receives \$ $k\ w_0^T$, but to meet his obligations he has to spend \$ $k\ W_1^K$ for a purchase of Deutschmarks in the spot market. Hence his profit is \$ $k(w_0^K - W_1^K)$. If we set $k = -h$ so that a forward sale of Deutschmarks is interpreted as a negative purchase, then his profit is again given by equation (1).

Forward speculators, in principle, do not need capital. It could, therefore, be conjectured that a speculator can make his commitment $h$ as large as he wishes. But this is not so. In practice an institutional rule has developed that limits his commitments, and we shall see that there are good reasons for this rule. The banking companies carrying out the forward contracts for their customers usually require a safety margin of between 10 and 20 percent of the dollar value of the forward commitment, since they are liable to the foreign trading partners. The level of the safety margin in general seems to be independent of whether speculators sell short or buy long. As interest is paid by the banks on the value of the safety margin this rule does not involve costs to the speculator but merely links the maximum speculative commitment with his personal wealth[2].

Given this information, the opportunity set of end-of-period wealth distributions ($V$) attainable by the speculator can easily be described. For simplicity we assume that, apart from the speculative profit, the speculator does not have further random income flows. As usual, non-random flows are admissible, however. It is assumed that their present value is a part of wealth and can be used to provide the safety capital required by the banks. Let $\beta$ denote the proportion of safety capital in the dollar volume of the commitment, $a$ the speculator's wealth after subtracting period consumption, and $q-1$ the safe market rate of interest. Then the opportunity set sought obviously is given by

$$V = aq + h(W_1^K - w_0^T), \quad |h| \leq \frac{aq}{\beta w_0^T}. \tag{2}$$

[2] The information was given on July 27, 1977, by the Deutsche Bank, Mannheim. GRUBEL (1965, p. 252) reports a 10% safety margin, but his information refers to the time where the exchange rate could only fluctuate within the narrow official band.

### 1.2. *Spot Market Speculation*

Next, consider the case of speculation in the spot market. In contrast to forward speculation, a speculation in the spot market is associated with an international movement of capital. Spot-market speculators import or export capital without protecting themselves in the forward market. Suppose a spot-market speculator plans, for one period, to buy foreign fixed-interest bonds for $h^*$ dollars and to invest the remainder of his wealth, $a-h^*$, in fixed-interest domestic bonds. Then at the end of the speculation period his wealth is

$$V=(a-h^*)q^I+h^*\frac{W_1^K}{w_0^K}q^A \tag{3}$$

$$=aq^I+\frac{h^*q^A}{w_0^K}\left[W_1^K-\frac{q^I w_0^K}{q^A}\right]$$

where $q^I-1$ is the domestic and $q^A-1$ the foreign rate of interest. If, however, the speculator plans to borrow $k^*$ dollars from foreigners in order to invest them in the domestic country then his end-of-period wealth is $aq^I+k^*q^I-q^A W_1^K k^*/w_0^K$ or, if with $k^*=-h$ we interpret a capital import as a negative capital export, it is again given by equation (3).

### 1.3. *Interest Arbitrage as the Link between Spot and Forward Speculation*

A comparison of the end-of-period wealth distributions (2) and (3) of the forward-market and the spot-market speculators may, at first sight, give the impression that both kinds of speculation are significantly different from each other. This impression is, however, wrong. This can be seen very clearly when the role of interest arbitrage is considered in addition to speculation[3]. Like spot-market speculators, interest arbitragers are capital importers or exporters. The difference is simply that arbitragers protect themselves in the forward market while speculators do not.

Suppose, from the viewpoint of the arbitragers, domestic and foreign fixed-interest bonds are perfect substitutes. Then we must have[4]

---

[3] Interest arbitragers and speculators are not necessarily different people, the latter can participate in arbitrage with that part of their wealth which is safely invested.

[4] With equation (4) we assume an infinitely elastic 'arbitrage schedule' of the kind assumed in the classical theory of interest parity. GRUBEL (1966, pp. 18–21) questioned this assumption, pointing out that even arbitrage is subject to political risk. Cf. also SCHRÖDER (1969, pp. 30–32).

$$q \equiv q^I = q^A \frac{w_0^T}{w_0^K} \tag{4}$$

for, if $q^I > q^A w_0^T / w_0^K$, there would be an enormous inflow of risk free capital which would reduce $w_0^K$ and/or raise $w_0^T$, and if $q^I < q^A w_0^T / w_0^K$, a corresponding capital outflow would induce the opposite adjustments.

By using the arbitrage equation (4), equation (3) now can be written as

$$V = aq + \frac{h^* q^A}{w_0^K}(W_1^K - w_0^T). \tag{5}$$

This equation already resembles equation (2). The only difference is that in (2) the speculative commitment $h$ was measured in foreign currency (DM) while in (5) the commitment is denoted by $h^*$ which is the dollar value of the capital to be exported. This difference, however, does not matter. If we measure the commitment of the spot-market speculator by the Deutschmark value of the redemption by setting $h \equiv q^A h^* / w_0^K$, then (5) takes on exactly the same form as (2). This surprising result originates from TSIANG (1959) who showed that spot-market speculation is, in economic terms, the same as a combination of forward speculation and pure interest arbitrage.

Thus it seems that from now on we only have to consider the forward speculator. However, the constraint $|h| \leq aq/(\beta w_0^T)$ has yet to be examined. If the capital exporting spot-market speculator uses his total wealth to buy foreign bonds or if the capital importing spot-market speculator takes on a debt up to the value of his wealth, then we have $|h^*| = a$, i.e., $|h| = aq/w_0^T$. It is not very likely that a speculator could succeed in making an even higher commitment by taking on additional debt for, in this case, some creditors would have to lend without security. Thus, for values of $\beta$ that realistically are significantly below unity, the opportunity set of the forward speculator who makes use of banks specializing in speculation is larger than that of the spot-market speculator. Since, however, the latter can always become a forward speculator, expression (2) can be used quite meaningfully to describe the opportunity set of any type of currency speculator.

### 1.4. *Integrating the Speculation Problem into the Basic Multiperiod Approach*

Provided with this attractive result, an attempt is now made to integrate the decision problem of the speculator into the basic model of stochastic multiperiod planning. First we check whether the opportunity

set described by (2) satisfies the requirement of stochastic constant returns to scale and write it in the form

$$(6) \qquad V = aQ$$

$$\text{where } Q \equiv q + \gamma\left(\frac{W_1^K - w_0^T}{w_0^T}\right), \quad \left|\gamma \equiv \frac{hw_o^T}{a}\right| \leq \frac{q}{\beta}.$$

This expression isolates an opportunity set of standard risk projects. Stochastic constant returns to scale prevail if this opportunity set is independent of the level of wealth, $a$. Since the admissible range for $\gamma$ is obviously independent of wealth, the condition is satisfied only if, in addition, the level of the speculative commitment has no influence on the current forward rate, $w_0^T$, and the probability distribution of the future spot rate, $W_0^K$. We ensure this by the assumption of a competitive market structure.

A second assumption in the basic model is that the distributions $Q$ at different points in time are stochastically independent of each other. If it is assumed that the speculators understand the operation of the market sufficiently well to take account of the arbitrage equation (4), then, in (6), $w_0^T$ can be replaced by $w_0^K q^I/q^A$. As in the case of portfolio analysis, our assumption therefore implies that the speculators expect stochastically independent growth rates, that is, a random walk in the exchange rate[5]. The assumption implies that, after an increase in the current exchange rate, the speculators do not expect either that there will be a relatively smaller rise in the exchange rate than they conjectured before this increase or that the observed change is simply a sign of even greater relative changes in the future. In short, a unitary expectation elasticity is assumed[6].

Another condition required for the multiperiod planning model was that the opportunity set should contain at least one element that avoids with certainty the loss of all wealth. This condition is clearly satisfied, since each $h$ in the range $0 \leq h < aq/w_0^T$ gives the desired protection.

With this, the integration of a wide class of speculation problems into our basic model is almost complete. We have only to assume additionally that, at each transactions date, the decision maker, after completing

---

[5] If the present approach is interpreted as referring to speculation in commodities futures then an assumption concerning the kind of price movement is unnecessary since there is no connection between forward and spot prices similar to (4).

[6] We thus decide for an intermediate solution between two extreme assumptions that have been favored in the literature. Cf. FRIEDMAN (1953, p. 175), ALIBER (1970, esp. pp. 304–306), and Nurkse, R., International currency experience. Princeton 1944. The last is cited according to ALIBER (1970, p. 304) and SOHMEN (1973, p. 73) since it was not available in the West German library system.

the previous contracts, thinks not only about his new commitment but also about the level of withdrawals for current consumption and that he attempts to maximize the multiperiod preference functional derived from the laws of Weber and Fechner. Then his implicit short-run aim is

$$\max_h E[U(aq + h(W_1^K - w_0^T))], \quad |h| \leq \frac{aq}{\beta w_0^T}. \tag{7}$$

where $U(.)$ is one of the time-dependent Weber functions. The implications of this aim will be discussed in the following sections.

## 2. *Optimal Speculation in the Ideal Case*

### 2.1. *The Two-Sided* $(\mu, \sigma)$ *Diagram*

To solve the maximization problem (7), the $(\mu, \sigma)$ diagram is considered again. Thus, from (2) the needed distribution parameters

$$E(V) = aq + h[E(W_1^K) - w_0^T] \tag{8}$$

and[7]

$$\sigma(V) = h \operatorname{sgn} h \, \sigma(W_1^K). \tag{9}$$

are calculated.

As is known, for an exact representation of the choice problem in a $(\mu, \sigma)$ diagram it is necessary for all distributions in the opportunity set to belong to the same linear class. To check this condition, calculate the standardized random variable $Z = [V - E(V)]/\sigma(V)$:

$$Z = \operatorname{sgn} h \frac{W_1^K - E(W_1^K)}{\sigma(W_1^K)}. \tag{10}$$

Equation (10) shows that, in general, there are *two* linear distribution classes rather than one. If a long position is taken ($\operatorname{sgn} h = +1$), then the distribution class to which the future spot rate ($W_1^K$) belongs applies, but if a short position is taken ($\operatorname{sgn} h = -1$), the distribution class defined by the 'mirror image' of the future spot rate applies. Only if $W_1^K$ is symmetrically distributed, will these two distribution classes coincide.

---

[7] For the definition of the 'sign' function cf. footnote 36 in chapter II D.

But there is no reason to expect symmetry. Rather, the exchange rate distribution seems to be right skewed. This is already suggested by the fact that, at the level of zero, the exchange rate has a lower bound while there is no obvious upper bound. If a symmetry assumption is suitable at all, then it should refer to the logarithm of the exchange rate so that $W_1^K$ and $1/W_1^K$ are equally distributed. Hence, the usual $(\mu, \sigma)$ diagram cannot be used to find a solution. But what about considering two diagrams?

This is done in Figure 6, where two indifference-curve systems of the kind depicted in Figure 7 in chapter III A are put together in an appropriate way. In this figure, it is assumed that the wealth distributions are bounded to the left, which, according to (10), implies that $W_1^K$ is bounded from above and from below[8]. For the time being, the ranges of abnormal indifference curves where the BLOOS rule comes into operation are left out. Accordingly, it is temporarily assumed that the opportunity locus does not intersect with these ranges. In section B 3 other possibilities are considered in detail.

The right-hand section of the indifference-curve system refers to a long position ($h>0$), and the left, that is the mirror image of the normal representation, refers to a short position ($h<0$). Because of the asymptotic efficiency of the variance[9], the indifference curves are nearly symmetrical with respect to the ordinate when the coefficients of variation are small. But the higher the standard deviation for any given mean, the greater the effect the difference between the two distribution classes has on the indifference-curve shapes. When the distribution of $W_1^K$ exhibits the described asymmetry, the end-of-period wealth distribution is right skewed in the case of a long position and left skewed in the case of a short position. In connection with the preference for right skewed distributions, suggested by Weber's law[10], this implies that the indifference curves are more curved in the left section of the diagram than in the right[11].

Since the points where the indifference curves enter the ordinate indicate the corresponding certainty equivalents, the indifference curves of

---

[8] With $W_1^K$ being unbounded from above, in the case of strong risk aversion ($\varepsilon \geq 1$) there would be lexicographic pseudo indifference curves in the left section of the diagram. Under weak risk aversion ($0<\varepsilon<1$), however, even in the case of an unbounded distribution of $W_1^K$, in the neighborhood of the ordinate there is always a range where the indifference curves have the normal shapes provided that, for $w_1^K \to \infty$, the density converges at least as fast as that of a normal distribution. Cf. the analysis towards the end of section III B 1.2.

[9] Cf. chapter II D 2.2.1.

[10] Cf. the corresponding remarks in the last third of section III A 2.3.2.

[11] That long and short speculation cannot be treated symmetrically was recognized by KENEN (1966, pp. 151 and 166).

both sections of the diagram that enter the ordinate at the same point can be considered as single indifference curves extending over both sections. Whenever the points representing end-of-period wealth distributions are situated on the same connected indifference curve, they are evaluated as being equal, no matter whether they are in the right or the left sections of the diagram.

Strictly speaking two arrows should be shown in the diagram, which are labelled $\sigma(V)$, start at the origin of the abscissa, and go in opposite directions. However, to obtain a scale that goes in one direction over the whole abscissa, the left-hand part is indicated as $-\sigma(V)$ and the right-hand part as $+\sigma(V)$, that is, in general as $\operatorname{sgn} h\,\sigma(V)$. This way, the indifference curves define a preference structure over the distribution parameters $E(V)$ and $\operatorname{sgn} h\,\sigma(V)$ that is identical with the one implied by the expected-utility criterion. Thus, for arbitrary distribution classes of $W_1^K$, the goal function (7) can be replaced by

$$\max_h U[E(V), \operatorname{sgn} h\,\sigma(V)], \quad |h| \leq \frac{aq}{\beta w_0^T}. \tag{11}$$

By using the $(\mu, \operatorname{sgn} h\,\sigma)$ diagram the optimal speculative commitment

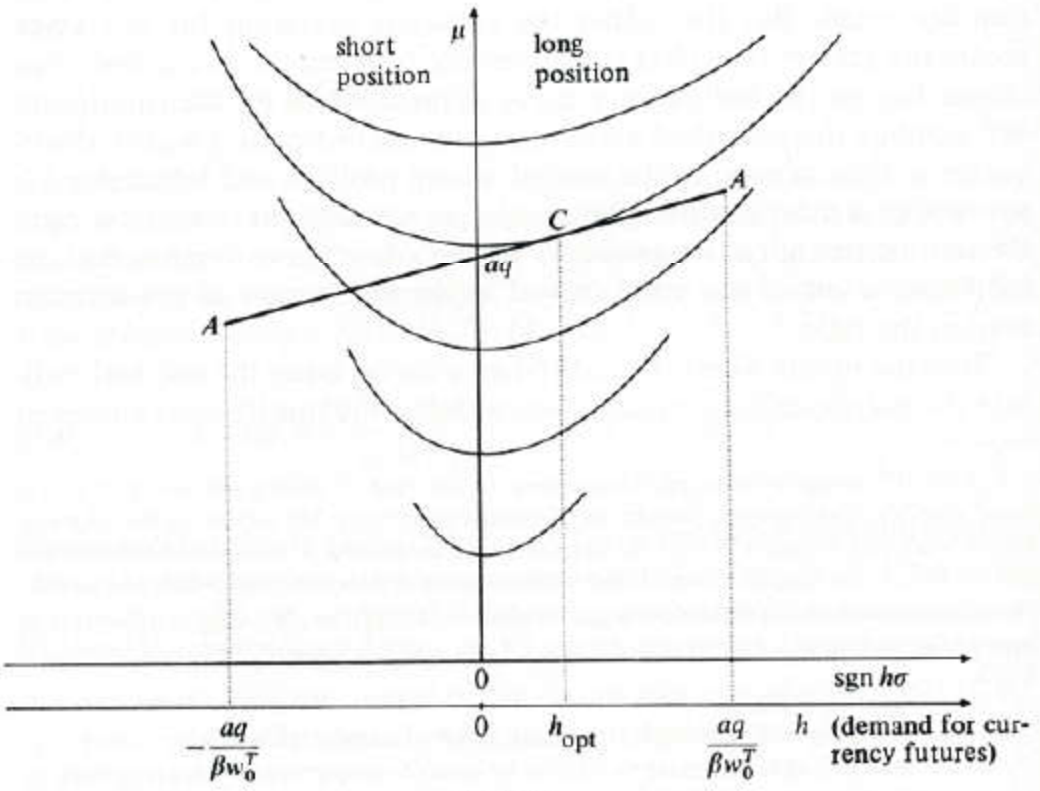


*Figure 6*

can easily be determined if the opportunity set is known. This set will now be considered. Calculating

$$h = \frac{\operatorname{sgn} h\,\sigma(V)}{\sigma(W_1^K)} \tag{12}$$

from (9) and inserting this expression into (8) we find that the opportunity locus is given by a straight line,

$$\mu = E(V) = aq + \operatorname{sgn} h\,\sigma(V)\frac{E(W_1^K) - w_0^T}{\sigma(W_1^K)}, \tag{13}$$

where account has to be taken of the constraint

$$\sigma(V) \leq \frac{aq}{\beta}\frac{\sigma(W_1^K)}{w_0^T} \tag{14}$$

that corresponds to the constraint $|h| \leq aq/(\beta w_0^T)$.

As an example for the case $E(W_1^K) > w_0^T$, the opportunity locus is represented by the line $AA$ of Figure 6. Since, by assumption, this line does not enter the range of abnormal indifference curves, there are two possibilities.

Either, as in the diagram, the optimal point is determined by a tangency solution or it coincides with the right-hand end of the 'opportunity line'. In the latter case the speculator buys as many Deutschmarks in the forward market as the bank allows. In general, the number of Deutschmarks bought is shown by the ray parallel to the abscissa which was constructed by using the proportionality between $\operatorname{sgn} h\,\sigma(V)$ and $h$ as given by (12).

### 2.2. *The Reaction of the Demand for Forward Currency to Changes in Expectations*

The influence speculators have on the spot and forward exchange rates is crucially determined by what they expect the future spot rate to be. For an evaluation and possible regulation of speculative trade, it is therefore useful on the one hand to have a theory concerning the formation of expectations and, on the other, to have a theory explaining how speculators react to a change in expectations. The former is beyond the scope of this book. The latter, however, is implicit in the approach developed above.

2.2.1. Changes in the Expected Spot Rate

It has already been shown in Figure 6 that a long position is advantageous to the speculator if $E(W_1^K) > w_0^T$. The kind of commitment that is chosen in the cases $E(W_1^K) < w_0^T$ and $E(W_1^K) = w_0^T$ can easily be determined.

According to (13) and (14) a reduction of $E(W_1^K)$ turns on the line $AA$ in a clockwise direction, while at the same time its length changes, since the ends of the line move vertically. If $E(W_1^K) = w_0^T$, the line is horizontal. Because the connected indifference curves have a slope of zero at the ordinate, in this case the point of tangency coincides with the ordinate and hence $h_{opt} = 0$. If the expected spot rate is below the forward rate then the line $AA$ slopes downwards to the right and the point of tangency $C$ is in the left section of the diagram: a short position is advantageous ($h_{opt} < 0$). The result is summarized in the following expression

$$h \left\{ \begin{matrix} > \\ = \\ < \end{matrix} \right\} 0 \Leftrightarrow E(W_1^K) \left\{ \begin{matrix} > \\ = \\ < \end{matrix} \right\} w_0^T. \tag{15}$$

Although (15) shows that the demand for forward Deutschmarks *globally* is a rising function of its expected spot rate, we do not know whether this function is monotonic. FELDSTEIN (1968, pp. 186 f.) pointed out that there may be counteracting income and substitution effects of a change in $E(W_1^K)$ so that there is a possibility that the demand for forward Deutschmarks is not everywhere a rising function of the expected Deutschmark spot rate[12].

For a general evaluation of speculation, the question is of great importance regardless of whether the speculators' abilities in forecasting the proper spot rate are estimated optimistically or pessimistically. The pessimist would stress that the expectation of a speculator as described by $E(W_1^K)$ is usually wrong and is subject to large fluctuations so that, from his point of view, it would be desirable if $\partial h / \partial E(W_1^K) = 0$, for then the transmission mechanism between expectations and forward rates is interrupted. The optimist, on the other hand, believes that speculators link changes in the forward rate with changes in the actual future spot rate, which requires $\partial h / \partial E(W_1^K) > 0$ if speculators are well-informed.

The reason for the indeterminateness mentioned by Feldstein is the generality of the preference hypothesis he used, which required nothing more than risk aversion. Fortunately, Weber's law provides us with additional information that gets rid of the indeterminateness: the pessimist's hopes are dashed and the optimist's hopes are confirmed.

To see why, consider Figure 7. There the original opportunity line $AA$ with the point of tangency $C$ moves counterclockwise towards BB since

[12] Cf. also LELAND (1971, pp. 260 f.) and HOCHGESAND (1974, pp. 116 f. and 128 f.).

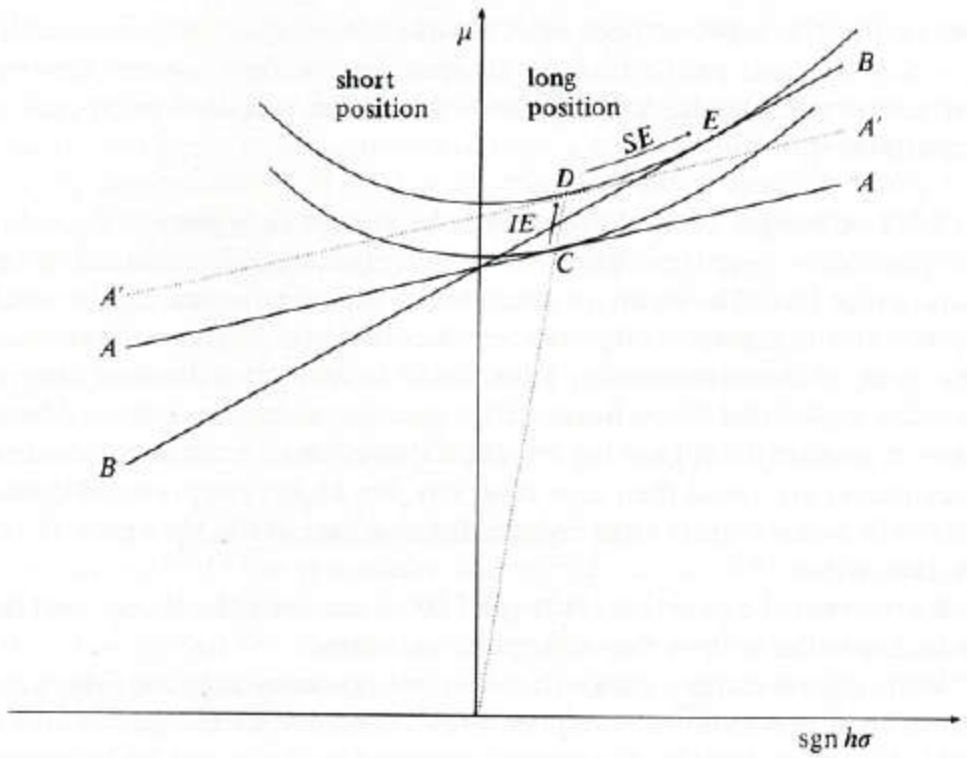


*Figure* 7

an increase in the expected spot rate is assumed. The new point of tangency is $E$. The movement from $C$ to $E$ can be divided into an income and a substitution effect. The income effect (IE) is represented by a parallel shift of the line $AA$ to the position $AA'$ and by the corresponding shift in the point of tangency from $C$ to $D$. Because of the homotheticity of the indifference-curve system as implied by Weber's law, point $D$ is to the right of point $C$: the income effect is positive. The substitution effect is represented by a movement of the opportunity line from position $A'A'$ to $BB$ along a given indifference curve. The point of tangency accordingly moves from $D$ to $E$. Because of the convexity of the indifference curves, $E$ clearly is to the right of $D$. Hence the income and substitution effects reinforce each other and so we have $\partial h/\partial E(W_1^K) > 0$.

In an analogous way, this reasoning can be used for the case of a short position. If the supply of forward currency is interpreted as a negative demand, the unambiguous result emerges that, concerning tangency solutions of the kind $C$ and $E$, the demand for forward currency is a strictly monotonically increasing function of the expected spot rate. By inspection of (12) and (14), it is easy to see that this result cannot be maintained in the case of a corner solution. Speculators who have committed as much as their banks allow do not react to marginal

changes in the expected spot rate. On the other hand these speculators are only a small proportion of all speculators; they cannot have an influence on the qualitative aspects of the market reactions to changes in expectations.

2.2.2. Changes in the Variance of the Future Spot Rate

Speculators base their behavior on a conjectured distribution of the future spot rate. The standard deviation of this distribution can be interpreted as a measure of their degree of confidence in the estimation of the mean of this distribution. Thus, from an allocative point of view, it is to be hoped that the influence that speculators have on the exchange rates is smaller the higher the standard deviation, for the less confident speculators are about their own forecasts, the higher the probability that they will create, rather than reduce, fluctuations in the time path of the exchange rate.

By reference to equations (12) and (13), it can easily be shown that the model speculator does not disappoint this hope.

With regard to the change in the initial tangency solution, there are two effects that, thanks to Weber's law, reinforce each other[13]. One is that, according to (13), the opportunity line in the $(\mu, \operatorname{sgn} h\sigma)$ diagram gets flatter when $\sigma(W_1^K)$ rises. It takes place regardless of whether the point of tangency was initially in the left or right section of the diagram. As is known from the previous section, as a reaction to this movement in the opportunity line, the point of tangency moves unambigously towards the ordinate. Thus $\sigma(V)$ is getting smaller. This effect, that, by itself, reduces the optimal commitment $h$, is reinforced by a second effect: according to (12), after an increase in $\sigma(W_1^K)$, each value of $\sigma(V)$ is associated with an absolutely lower value of $h$ than before.

If, initially, the solution point is situated at an end of the opportunity line, then, for marginal changes in $\sigma(W_1^K)$, the speculative commitment is not affected. Independently of $\sigma(W_1^K)$ we then have $h = aq/\beta w_o^T$ or $h = -aq/\beta w_0^T$.

Since a corner solution, as a rule, does not occur for all speculators, we again find a clear-cut conclusion for the aggregate. If all speculators expect higher variances of the future spot rate the commitments in the aggregate are reduced regardless of whether these are long or short.

## 2.3. *The Wealth Effect and the Stability Problem*

From the application of our basic model to the portfolio problem (A 3.3.), we know that the optimal portfolio structure is independent of

[13] Cf., however, FELDSTEIN (1968, p. 187).

the decision maker's wealth. In the case of speculation a similar result can be found.

In Figure 6, an increase in the level of wealth ($a$) available after a subtraction of period consumption implies a parallel shift and lengthening of the opportunity line so that its ends ($A$) move upwards along rays through the origin. This can be checked directly by inspecting (13) and (14). Because of the homotheticity of the indifference-curve system, this also means that the point of tangency $C$ moves upward along a ray through the origin. Hence the *demand* for Deutschmark futures rises in strict proportion to wealth. If initially, in contrast to Figure 6, $h_{opt}<0$ had been the case, then analogous reasoning would have shown that the optimal *supply* of Deutschmark futures ($-h_{opt}$) rises in proportion to wealth.

This wealth effect has some relevance for the stabilizing effects of speculation[14]. If the speculator was already committed in the previous period, then at the beginning of the period his wealth before consumption depends on the current spot rate. Because of the constancy of the marginal propensity to consume out of wealth, this means that the level of the funds to be reinvested also depends on this rate. With a long position it rises, with a short position it falls. Hence the wealth effect implies that the current demand for currency futures depends on the current spot rate.

Suppose, before the decision point in time 0, the speculator expects $E(W_1^K)>w_0^T$. In this case, he plans to take a long position and, because of the assumption that the expectation elasticity is one, he sticks to this plan regardless of what the variates of the current spot and forward rates, $w_0^K$ and $w_0^T$, happen to be. If the speculator's previous commitment was long, then, at point in time 0, his demand for currency futures is a rising function of the spot rate and, because of the arbitrage condition (4), also of the forward rate. Obviously, in this case, the wealth effect is destabilizing. If, however, the speculator was previously in a short position, then the reverse is the case. At point in time 0, his demand for Deutschmark futures is a falling function of the forward rate. The wealth effect is stabilizing.

Analogous reasoning can be applied to the case where, at point in time 0, the speculator decides to sell short. Thus we reach the general conclusion that the wealth effect has a stabilizing influence on the exchange market if speculators switch between long and short positions, and has a destabilizing influence if they stay with a given type of speculation.

[14] As far as is known, the wealth effect has been disregarded in the extensive literature on the problem of whether or not speculation is stabilizing. Reviews of the literature are given by HOCHGESAND (1974) and STEINMANN (1970).

### 3. *On the Possibility of an Excessively Short Position*

In this section, an aspect of the speculator's decision problem is studied that may bring about a particular preference for short positions. It is the BLOOS rule[15] that is responsible for this preference for it allows the speculator to shift part of the speculation risk on to the shoulders of others.

As we know, the BLOOS rule comes into operation only if the gross wealth distribution extends partly over the negative half of the wealth axis. Then, in the usual $(\mu, \sigma)$ diagram, the distribution is represented by a point below the border line[16] $\mu = \underline{k}\,\sigma$, where $-\underline{k}$ is the highest lower bound to the standardized end-of-period wealth distribution. We therefore need to think about where this border line is located in the two-sided $(\mu, \sigma)$ diagram and what shape the indifference curves have beyond it.

Since, in the present case, there are *two* standardized end-of-period wealth distributions according to whether the speculator holds a long or a short position, two lower bounds, $\underline{k}_L$ and $\underline{k}_S$, have to be distinguished. By using (10), these bounds can be derived from the distribution of the future spot rate $W_1^K$. Assume, to take a plausible[17] example, logarithmically symmetrical bounds:

$$\frac{E(W_1^K)}{1+\lambda} \leq W_1^K \leq (1+\lambda)E(W_1^K), \quad 0 < \lambda \leq \infty. \tag{16}$$

Then

$$\underline{k}_L = \frac{E(W_1^K)}{\sigma(W_1^K)} \frac{\lambda}{1+\lambda} \tag{17}$$

and

$$\underline{k}_S = \frac{E(W_1^K)}{\sigma(W_1^K)} \lambda, \tag{18}$$

so that we find the following border lines in the $(\mu, \operatorname{sgn} h\,\sigma)$ diagram:

$$E(V) = \underline{k}_L \operatorname{sgn} h\,\sigma(V), \tag{19}$$

$$E(V) = -\underline{k}_S \operatorname{sgn} h\,\sigma(V). \tag{20}$$

---

[15] Cf. chapter III B.

[16] Cf. expression (III A 44).

[17] Cf. the remarks in section 2.1.

They are both depicted in Figures 8 and 9. Beyond these border lines two types of indifference curves are possible[18]. In the case of strong risk aversion ($\varepsilon \geq 1$; Figure 8), there are pseudo indifference curves in the form of straight lines through the origin. In the case of weak risk aversion ($\varepsilon < 1$; Figure 9) there are genuine indifference curves that become concave at some stage, change the signs of their slopes, and eventually intersect the abscissa.

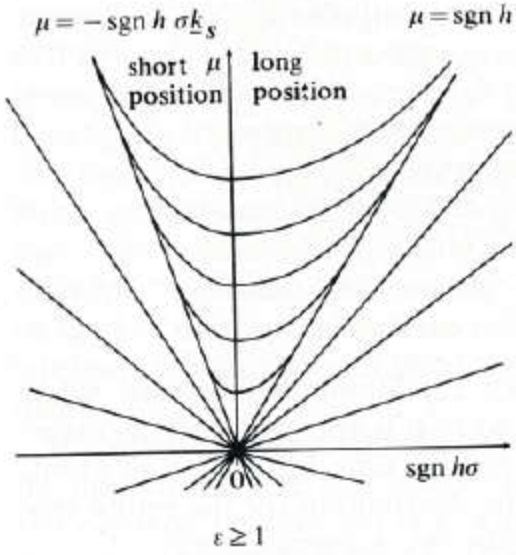


*Figure* 8

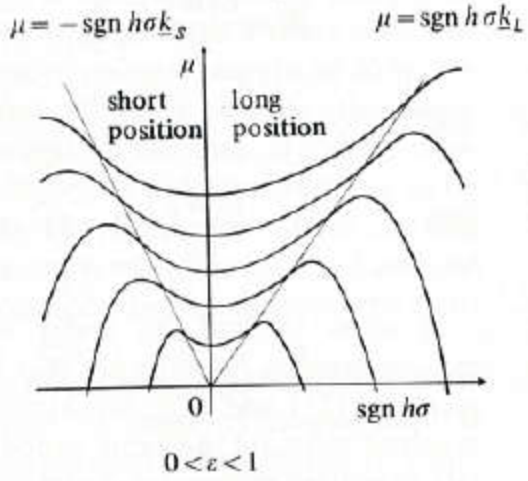


*Figure* 9

The question now is, under which conditions does the opportunity set available to the decision maker contain end-of-period wealth distributions that map beyond the border lines (19) and (20). According to (13) and (14), the safety margin parameter $\beta$ required by the banks is of crucial importance for this question. If use is made of this parameter, the question can be posed more precisely by asking which values of $\beta$ define critical levels below which the opportunity line given by (13) and (14) goes beyond the right (19) or left (20) border lines. To calculate these levels, call them $\beta_L^*$ and $\beta_S^*$, first combine (14) with (19) or (20), so that

$$\frac{E(V)}{\underline{k}} = \frac{aq}{\beta^*} \frac{\sigma(W_1^K)}{w_0^T}, \quad \underline{k} = \underline{k}_L, \underline{k}_S; \; \beta^* = \beta_L^*, \beta_S^*.$$

[18] Cf. Figures 10 and 12 in chapter III B. In Figure 8 above it is assumed that the funnel edges are tangent to the indifference curves. As we know from the discussion of Figure 12 in chapter III B this is a particular property which does not hold for all types of distribution if $1 < \varepsilon < 2$. The property is irrelevant for the present discussion.

Now replace $E(V)$ according to (8) and substitute

$$h = \operatorname{sgn} h \frac{aq}{\beta^* w_0^T}, \qquad \beta^* = \beta_L^*, \beta_S^*.$$

This is possible because the constraint $|h| < aq/(\beta w_0^T)$ from (7) is equivalent to (14). Finally, specify $\underset{\sim}{k}$ by using the values given alternatively by (17) and (18). Then the result is

$$\beta_L^* = \frac{E(W_1^K)}{w_0^T} \frac{\lambda}{1+\lambda} \tag{21}$$

and

$$\beta_S^* = \frac{E(W_1^K)}{w_0^T} (1+\lambda) - 1. \tag{22}$$

In order to find out under which conditions the required safety margin satisfies its purpose, that is, when it is above the critical values given by (21) and (22), we assume for the time being that all agents involved estimate the same probability distribution for the future spot rate (idealized uncertainty) and consider two *extreme* cases.

1. Suppose $E(W_1^K) > w_0^T$. Then $\beta_L^* \leq 0$ if $\lambda$ is small enough to render $E(W_1^K)/(1+\lambda) \geq w_0^T$. In this case, no safety margin need to be required by the banks, since the smallest possible spot rate exceeds the forward rate. It is true, since $\beta_S^* > 0$, a safety margin would be necessary to exclude negative gross wealth in the case of a short position. However, even if the banks did not require a safety margin, no one would engage in this type of speculation, for, at best, wealth would be maintained at the same level while the most likely outcome would be that it is reduced. This result holds irrespective of the fact that the speculator can shift some of the risk on to his bank's shoulders. An analogous argument can be given in the case $E(W_1^K) < w_0^T$ and $E(W_1^K)(1+\lambda) \leq w_0^T$ so that $\beta_S^* \leq 0$. Here, too, we find that for $\lambda$ sufficiently small no safety margin is required.

2. Another extreme case is $\lambda \to \infty$. For a long position a safety margin of 100% ($\beta_L^* = 1$) will then be needed. If, in the worst of all cases, the value of the foreign currency falls to zero, the wealth of even the most courageous speculator would be just enough to buy, as contracted, the devalued foreign currency and to throw it into the waste paper basket. The situation is very different in the case of a short position. Here, in

the limiting case, an infinitely high safety margin is needed; this means that the bank should not allow short speculation at all.

By their very nature, both of the extreme situations illustrated do not reflect normal expectations about changes in the exchange rate. Nevertheless, there are more realistic examples, which do confirm the observation that, in the case of a short position, it is hardly possible to avoid transferring some of the speculator's risk to other's shoulders. Suppose that a doubling or a halving of the spot rate are considered to be the most extreme possibilities ($\lambda = 1$) and assume, for simplicity, $E(W_1^K) \approx w_0^T$. Then for a long position a minimum safety margin of about 50% is needed while about 100% is needed for a short position. Comparing this with the, in practice, more realistic margin of 20%, we find from $|h| \leq aq(\beta w_0^T)$ that, in the case of a long position, the speculative commitment can be 2 1/2 times and, in the case of a short position, 5 times as large as it would have to be if a risk transfer were to be excluded. Formally this means that the opportunity line in the $(\mu, \operatorname{sgn} h\,\sigma)$ diagram exceeds the right-hand border line by 2 1/2 and the left-hand border line by 5 times the corresponding distance to the ordinate!

In the light of this dramatic change in the previous assumption that the opportunity line does not go beyond the range of normal indifference curves, the previous results definitely need to be rechecked. Little happens, when there is strong risk aversion ($\varepsilon \geq 1$; cf. Figure 9). Since all pseudo indifference curves outside the funnel are subordinate to those inside, a solution is only possible within the funnel, in the extreme case at the edges (cf. footnote 18). Equation (15) will then continue to be true. Speculators in this case are so afraid of losing their wealth that, being able to avoid some of their obligation in the case of a total disaster, has no appeal for them. Unfortunately, it was this very hypothesis of strong risk aversion that was shown to be rather unrealistic[19]. Under weak risk aversion ($0 < \varepsilon < 1$), optimal solutions outside the funnel are clearly possible.

Figure 10 shows a particularly curious situation. There, $E(W_1^K) > w_0^T$, so that a long position with the point of tangency $C$ could be expected to be optimal. But, in fact, the opportunity line reaches the highest indifference curve at its left end, at point $P$. Not a moderate long position but a short position of the highest possible extent is optimal. This is the case mentioned in the introduction.

The reason for this result is the assumption of logarithmically symmetrical bounds to the probability distribution of the spot rate. It implies

[19] Cf. section III B 2, towards the middle, and section IV B 2.3.2.

the limiting case, an infinitely high safety margin is needed; this means that the bank should not allow short speculation at all.

By their very nature, both of the extreme situations illustrated do not reflect normal expectations about changes in the exchange rate. Nevertheless, there are more realistic examples, which do confirm the observation that, in the case of a short position, it is hardly possible to avoid transferring some of the speculator's risk to other's shoulders. Suppose that a doubling or a halving of the spot rate are considered to be the most extreme possibilities ($\lambda = 1$) and assume, for simplicity, $E(W_1^K) \approx w_0^T$. Then for a long position a minimum safety margin of about 50% is needed while about 100% is needed for a short position. Comparing this with the, in practice, more realistic margin of 20%, we find from $|h| \leq aq(\beta w_0^T)$ that, in the case of a long position, the speculative commitment can be 2 1/2 times and, in the case of a short position, 5 times as large as it would have to be if a risk transfer were to be excluded. Formally this means that the opportunity line in the $(\mu, \operatorname{sgn} h\sigma)$ diagram exceeds the right-hand border line by 2 1/2 and the left-hand border line by 5 times the corresponding distance to the ordinate!

In the light of this dramatic change in the previous assumption that the opportunity line does not go beyond the range of normal indifference curves, the previous results definitely need to be rechecked. Little happens, when there is strong risk aversion ($\varepsilon \geq 1$; cf. Figure 9). Since all pseudo indifference curves outside the funnel are subordinate to those inside, a solution is only possible within the funnel, in the extreme case at the edges (cf. footnote 18). Equation (15) will then continue to be true. Speculators in this case are so afraid of losing their wealth that, being able to avoid some of their obligation in the case of a total disaster, has no appeal for them. Unfortunately, it was this very hypothesis of strong risk aversion that was shown to be rather unrealistic[19]. Under weak risk aversion ($0 < \varepsilon < 1$), optimal solutions outside the funnel are clearly possible.

Figure 10 shows a particularly curious situation. There, $E(W_1^K) > w_0^T$, so that a long position with the point of tangency $C$ could be expected to be optimal. But, in fact, the opportunity line reaches the highest indifference curve at its left end, at point $P$. Not a moderate long position but a short position of the highest possible extent is optimal. This is the case mentioned in the introduction.

The reason for this result is the assumption of logarithmically symmetrical bounds to the probability distribution of the spot rate. It implies

[19] Cf. section III B 2, towards the middle, and section IV B 2.3.2.

that, in the case of a short position, the gross wealth distribution is strongly left skewed. As we know, this property is a disadvantage if the gross and the net distributions coincide, an aspect that is represented by the stronger curvature of the indifference curves in the left section of the figure[20]. However, if the gross distribution can take on negative values, then, because of the BLOOS rule, this disadvantage can be compensated or even overcompensated.

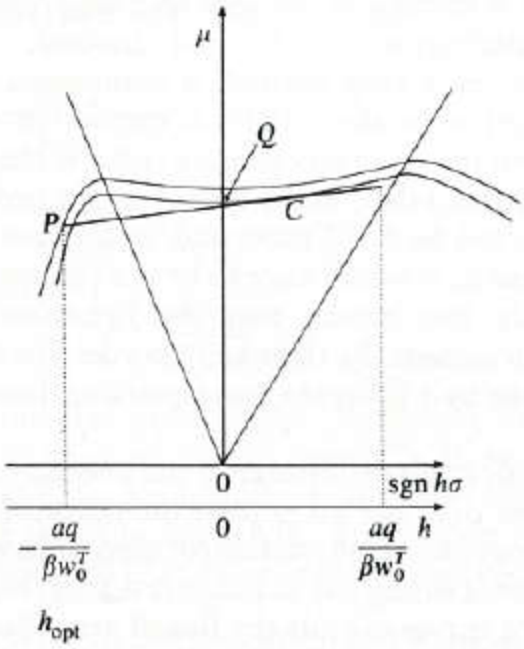


*Figure* 10

That there is the possibility of an overcompensation can be shown by a thought experiment. Consider, as a first step, the case of a risk neutral decision maker ($\varepsilon = 0$). For him, the indifference curves within the funnel are horizontal and outside they bend downwards, since the kink in the utility curve, brought about by the BLOOS rule, clearly implies risk loving behavior[21]. Assume that the opportunity line the decision maker faces is also horizontal, $E(W_1^K) = w_0^T$, and that, at its right-hand side, it just ends at the edge of the funnel ($\beta = \beta_L^*$). Then, since (21) and (22) give $\beta_S^* = \lambda > \lambda/(1+\lambda) = \beta_L^* = \beta$, the opportunity line has to go beyond the left-hand edge of the funnel. The situation is illustrated in Figure 11. Obviously, the optimal point is at the left end of the opportunity line, i.e., at point $P$. It is clearly better than, for example, point $Q$ which is on the ordinate.

---

[20] Cf. section 2.1 above.

[21] To see that there is a negative slope consult equation (III B 5).

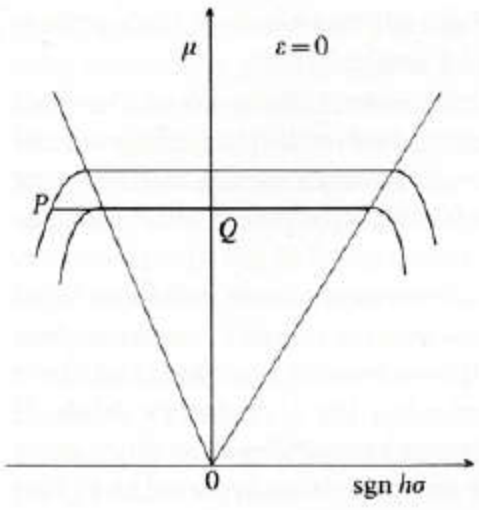


*Figure* 11

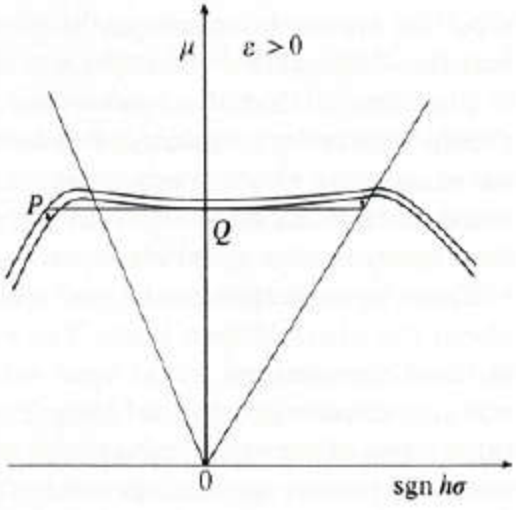


*Figure* 12

Now consider the second step. Rather than setting $\varepsilon = 0$, we assume $\varepsilon$ is slightly above 0. Then, for any given ratio $\mu/(\operatorname{sgn} h\sigma)$, in the right-hand section of the diagram, the indifference-curve slope is higher and, in the left-hand section, it is lower than before, while at the ordinate it remains zero. With a strong increase in $\varepsilon$, it could happen that $P$ is situated on the same, or even on a lower, indifference curve than $Q$. But, for a sufficiently small increase in $\varepsilon$, the indifference curve passing through $Q$ will, as in Figure 11, still be below $P$. Figure (12) represents the new situation.

In the third and last step, change the position of the opportunity line by reducing the initial wealth $a$ and increasing the expected spot rate $(E(W_1^K) > w_0^T)$. Via the movements of the ends of the opportunity line, as indicated by the arrows in Figure 12, we are then indeed able to reproduce the situation depicted in Figure 10.

This raises considerable doubt concerning the functioning of the market in currency futures. If speculators choose a short position simply because they hope they will not have to meet their liabilities in certain cases, they cannot be expected to properly link the forward rate with the future spot rate. But, in fact, the problem is even more complex.

The suspicion of a malfunctioning of the market seems to vanish if we take into account the behavior of banking firms. Banks will have a strong interest in avoiding being burdened with the speculator's insolvency risk, so they require a safety margin just high enough to prevent this from happening. Actually, the safety parameter $\beta$ could be considered as a measure of the most extreme changes in the spot rate that the banks think possible. Unfortunately, however, there are two reasons

why this argument cannot really eliminate the doubts about the performance of the market raised by the above analysis.

The first is that the banks themselves, particularly when they are facing insolvency problems, may have an incentive to take advantage of an excessively short speculation. The recent bankruptcy of the West German *Herstatt Bank* that had bad luck in an excessively short speculation seems to be a good example.

The second is that banks and speculators might have different ideas about the possible spot rates. The safety margins banks require at best indicate the changes in the spot rate they consider possible, but these margins do not provide information on what the speculators think. If both types of agents calculate with different probability distributions, it may well be that speculators believe the situation is as depicted in Figure 10 while banks believe that they are secure.

There does not seem to be a straight-forward way of evaluating this possibility from a welfare point of view, even if we confine ourselves to the Pareto criterion. From an extremely subjectivist position, an excessively short speculation is not a disadvantage, either from the viewpoint of the speculator or from the viewpoint of the bank; otherwise a contract would not have been made. The Pareto criterion may, however, also be used in a more objective and less tautological form. Suppose, after an exchange of information between bank and speculator, both parties agree on the probability distribution of the spot rate. Then, the decision in favor of a speculative commitment that, for the speculator and the bank together, brings about an expected net loss must indicate an objective deterioration for at least one of them. If the exchange of information really took place the excessively short speculation would no longer occur. But it seems hard to imagine such an exchange of information taking place, for the speculator who plans an excessively short commitment has no incentive to reveal his information.

### 4. *Summary*

With the theory of currency spot and forward speculation another area has been investigated to which the previously developed approach can be meaningfully applied. For this application it seemed useful to derive a two-sided $(\mu, \sigma)$ diagram by means of which the optimal speculative commitment, whether long or short, can easily be found without assuming a particular distribution class for the spot rate estimated by speculators. A number of results were achieved that, although formulated with respect to forward speculation, are equally relevant for the behavior of spot market speculators, since these may be interpreted as forward speculators engaged in interest arbitrage.

When sufficiently high safety margins are required by the banks, a long position is advantageous to the speculator if the expected spot rate exceeds the observed forward rate, and a short position is preferable in the reverse case. Because of the preference structure implied by Weber's law, the demand for forward currency reacts normally to a change in the expected spot rate, that is, it rises when the expected spot rate increases. An increase in the assumed variance around the expected spot rate leads to a reduction in the speculative commitment, irrespective of whether the speculator takes a long position or a short position.

If the safety margin required by the bank is not high enough to make negative variates of the speculator's gross wealth distribution impossible, or if the bank itself engages in speculation, then these results may not hold. The speculator may well prefer to risk an excessively short commitment although his expectation of the future spot rate is above the forward rate and although his preferences are characterized by a concave von Neumann–Morgenstern utility function. This is another implication of the BLOOS rule.

The speculator's demand for forward currency *ceteris paribus* is proportional to his wealth. This wealth dependence is important for the question of whether speculation has a stabilizing or a destabilizing effect on the spot and forward rates, since, in the case of repetitive speculation, current wealth is determined by the current level of the spot rate. The wealth effect stabilizes if the speculation changes between long and short positions. It destabilizes when a particular type of commitment is maintained.

## Section C
## Theory of Insurance Demand

One of the most important and most obvious accomplishments of risk theory is to explain why the insurance business is rewarding for both the insurance purchaser and the insurance company. For this reason, the situation of the insurance purchaser has frequently been used in this book to illuminate the discussion[1]. Now an attempt is made to give a more systematic and comprehensive analysis of insurance demand. Section C 1 considers the determinants of insurance demand for given risks and section C 2 extends the analysis to the case of endogenous risks where the household can decide not only, as usual, the optimal rate of consumption, but also can choose between alternative loss prevention policies and insurance contracts.

---

[1] Cf. chapter II C 1.2, II C 1.3, and III B 1.1.

## 1. *Insurance Demand for Given Risks*

### 1.1. *The Basic Calculus of the Insurance Purchaser*

The analysis of insurance demand for given risks begins by once again considering BARROIS's (1834) problem of determining the maximum willingness to pay for a full-coverage contract. Then, the determinants of the optimal degree of loss coverage are studied, a problem first considered by BORCH (1961) and MOSSIN (1968b)[2,3]. The present contribution to these problems is to extend the analysis to liability insurance and to integrate it into the multiperiod approach.

The choice problem underlying both types of problems can be formulated by a common approach. As before[4], let $aq$ denote end-of-period wealth if the decision maker does not buy insurance and is lucky enough to avoid any loss and let $C$, $C \geq 0$, denote the possible loss which is a random variable with finite density for all variates. Then the decision maker's end-of-period wealth distribution without insurance is $V = aq - C$. The proportion $\theta$ of losses is underwritten by the company at the cost of a premium, payable at the end of the period. The premium loading factor is $\bar{g}$. Insurance brings about an improvement in the end-of-period wealth distribution by the amount $\theta C$ and a worsening by $\bar{g}\,E(\theta C)$ so that, in general, this distribution can be written as

$$V = aq - C(1-\theta) - \bar{g}\theta E(C). \tag{1}$$

The first question, which is obviously identical with the question about the intensity of insurance demand ($g$) defined in chapter II 1.3, now is: suppose the decision maker can choose between $\theta = 0$ and $\theta = 1$. How large is the loading factor $\bar{g} = g$, that he is just willing to accept? And the second question is: which degree of coverage $\theta$ is chosen by the insurance purchaser if a continuum of alternatives $0 \leq \theta \leq 1$ is available[5]

---

[2] The problem was also discussed by EHRLICH and BECKER (1972, pp. 625–633) and, within a growth-optimum model, by HOFFLANDER, RENSHAW, and RENSHAW (1971). RAZIN (1976) offered a minimax-regret solution, which is very different.

[3] Besides the ones considered below various other questions have been discussed. HAMBURG and MATLACK (1968) and SMITH (1968) studied the problem of casuality loss insurance. ARROW (1963, pp. 969–973), PASHIGIAN, SCHKADE, and MENEFEE (1966), MOSSIN (1968b, pp. 561–563), GOULD (1969), HAEHLING VON LANZENAUER (1971), and HAEHLING VON LANZENAUER and WRIGHT (1975) investigated the optimal level of deductibles. A very extensive and general analysis covering various kinds of insurance is provided by ARROW (1974b).

[4] Cf. chapter II C 1.2.

[5] The range $0 \leq \theta \leq 1$ corresponds to the normal concept of insurance, for $\theta < 0$ means that the insurance 'purchaser' increases his risk and $\theta > 1$ that he turns into a gambler. In practical insurance problems the possibility $\theta > 1$ cannot always be avoided. The moral-hazard problem arising from this possibility in the case of manipulable risks is discussed in section C 2.1.3.

and the insurance company predetermines the loading factor? The questions have in common the fact that they lead to an insurance demand function, the first to a discontinuous function

$$\theta = \begin{cases} 1, & \text{if } \tilde{g} < g \\ 0, & \text{if } \tilde{g} > g \end{cases}$$

and the second to a possibly continuous function $\theta = \theta(\tilde{g})$.

To integrate the insurance problem into the multiperiod approach[6] we have to think about how the opportunity locus described by (1) changes over time. First, of course, the fact has to be taken into account that the decision maker's wealth, which gives an upper bound to the size of his effective loss, equals the end-of-period wealth of the previous period minus current consumption. Hence, wealth depends on the degree of coverage chosen in the previous period as well as on the variate the loss variable took on before. Next we must note that the wealth so determined will itself have some influence on the loss distribution. In fact, often there seems to be a very close relationship between a person's property and the size of his possible losses. To depict this relationship in a simple idealized form, it is assumed that the loss can be expressed as the product of the decision maker's wealth $a$ with an, arbitrarily distributed, random 'loss factor' $F$:

$$C = aF, \quad F \leq 1. \tag{2}$$

Equation (1) then can be written in the form

$$V = aQ, \quad Q \equiv q - F(1-\theta) - \tilde{g}\theta E(F), \tag{3}$$

and it turns out that (2) implies stochastic constant returns to scale.

Thus, important conditions underlying the multiperiod approach are satisfied. For a full applicability of the multiperiod model, two further conditions are needed however. The first is the stochastic intertemporal independence of the standard risk projects $Q$. It is satisfied if the loss factors $F$ are stochastically independent over time, as is assumed. The second condition is that the opportunity set contains at least one alternative that with certainty avoids a complete loss of wealth. This condition is satisfied if full coverage insurance contracts that cost less than the insurance purchaser's wealth are available. This obviously weak condition is also assumed to hold.

The integration of the insurance purchaser's decision problem into

[6] Cf. chapter IV B 2.

our multiperiod approach is thereby accomplished. Accordingly, from a long-run perspective, it is optimal to evaluate a wealth distribution attainable at the end of the current period by referring to the preference functional

$$E\{U[aq - C(1-\theta) - \tilde{g}\theta E(C)]\} \tag{4}$$

where $U(.)$ is one of the time-dependent Weber functions defined by the measure of relative risk aversion $\varepsilon$. However, as in the treatment of the portfolio and the speculation problems, we again prefer to replace the preference functional (4) by an equivalent[7] preference functional

$$U[E(V), \sigma(V)] \tag{5}$$

where

$$\begin{aligned} E(V) &= aq - E(C)(1-\theta) - \tilde{g}\theta E(C) \\ &= aq - E(C) - \theta E(C)(\tilde{g} - 1) \end{aligned} \tag{6}$$

and

$$\sigma(V) = (1-\theta)\sigma(C). \tag{7}$$

The properties of this preference functional were investigated in chapter III B 1.2 and B 2. To ensure that the replacement does not imply any loss of accuracy, it must be required that the end-of-period wealth distributions, attainable through a choice of $\theta$ in the open unit interval, all belong to the same, arbitrarily choosable, linear distribution class. As can be seen from calculating the standardized variable

$$\begin{aligned} Z &= \frac{V - E(V)}{\sigma(V)} \\ &= \frac{[aq - C(1-\theta) - \tilde{g}\theta E(C)] - [aq - E(C)(1-\theta) - \tilde{g}\theta E(C)]}{(1-\theta)\sigma(C)} \\ &= \frac{E(C) - C}{\sigma(C)}, \end{aligned} \tag{8}$$

this requirement is met. Hence, with (5)–(7) we have a basis for finding answers to the questions initially posed.

---

[7] The preference functionals (4) and (5) are identical up to a strictly increasing monotonic transformation.

### 1.2. *The Maximum Willingness to Pay for a Full-Coverage Insurance Contract*

The maximum willingness to pay for a full-coverage insurance contract divided by the expected loss is what we called the intensity of insurance demand, $g$. It is implicitly given by the equation

$$U[aq - gE(C), 0] = U[aq - E(C), \sigma(C)] \tag{9}$$

which is obtained by setting alternately $\theta = 0$ and $\theta = 1$ in (5).

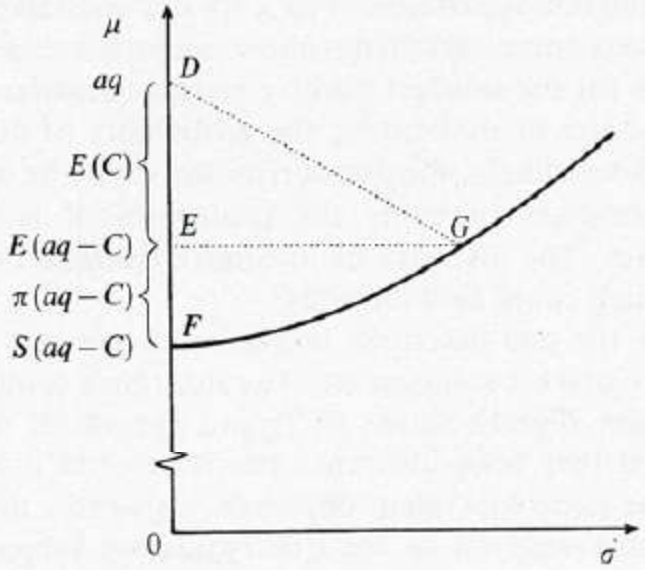


*Figure* 13

The size of $g$ can be graphically determined by inspecting Figure 13. There, the end-of-period wealth distribution without insurance, $aq - C$, is represented by point $G$. The corresponding level of end-of-period wealth without loss, $aq$, is shown by point $D$, the expected level of end-of-period wealth, $E(aq - C)$, by point $E$, and the certainty equivalent of the end-of-period wealth distribution, $S(aq - C)$, by point $F$ where the indifference curve passing through $G$ reaches the ordinate. The distance $\overline{DE}$ measures the expected loss $E(C)$ and the distance $\overline{EF}$ the subjective price of risk $\pi(aq - C)$. The intensity of insurance demand therefore is[8]

$$g = \frac{\pi(aq - C) + E(C)}{E(C)} = \frac{\overline{FD}}{\overline{ED}}. \tag{10}$$

In the figure, $g > 1$, since convex indifference curves were assumed. According to the results of chapter III (A 2.2, B 1.2, B 2) such a shape is

[8] Cf. chapter II C 1.3.

ensured if the loss cannot exceed the decision maker's initial wealth, that is, if the BLOOS rule is not operative. This is necessarily the case with property risks, but it may also occur with liability risks when the loss distribution is bounded from above.

Convex indifference curves, however, also occur in the case of an unbounded loss distribution if the standard deviation of losses is sufficiently small, risk aversion is weak ($\varepsilon > 1$), and the density at the upper tail of the loss distribution converges at least as fast as that of a normal distribution; such a constellation is possible for some types of small-scale liability insurance.

If the decision maker is extremely risk averse ($\varepsilon \geq 1$), then, in the case of loss distributions unbounded from above, there is a different picture. In this case, even for the smallest positive levels of standard deviations, the lexicographic aim of minimizing the probability of disaster comes into operation. Accordingly, the decision maker would be willing to give nearly all he possesses to enjoy the protection of a full-coverage insurance contract. The intensity of insurance demand then is almost $g = aq/E(C)$, which could be enormous[9].

If this case or the one described in Figure 13 prevails then a competitive insurance market can operate. The insurance companies would require a premium slightly above $E(C)$ and almost all risk could be insured, provided that both insurance purchaser and insurance company estimate the same equivalent objective probability distribution of losses. (Even if this were not so, the strictly positive subjective price of risk $\pi$ would still provide some scope for mutually beneficial insurance contracts.) Unfortunately however, in the case of liability risks, other constellations are also possible. These will now be analyzed.

Let us assume, for this purpose, that the loss distribution $C$ is bounded from above and see how the intensity of insurance demand develops if, starting from a situation of the kind depicted in Figure 13, the loss distribution is subject to proportional extension or compression given the level of *normal wealth* $aq$. An inspection of (6) and (7) shows that a proportional extension reduces $E(V)$ and increases $\sigma(V)$. To see this more precisely write (6) as

$$E(V) = aq - \frac{E(C)}{\sigma(C)}\sigma(V), \tag{11}$$

where $\sigma(C) = \sigma(V)$. Note that the upper boundary $\bar{k}$ of the standardized

---

[9] Precisely speaking, the intensity of insurance demand is not defined in this case, since the purchaser would be willing to pay any amount that is even minutely below his total level of wealth, but not an amount equal to this level.

distribution (8) is reached just where $C$ takes on its lowest variate 0. This implies

$$\bar{k} = \frac{E(C)}{\sigma(C)}. \tag{12}$$

Hence, (11) changes to

$$E(V) = aq - \bar{k}\sigma(V). \tag{13}$$

By graphing this equation it is easy to read the maximum willingness to pay from the indifference-curve system appropriate to the level of risk aversion we wish to assume.

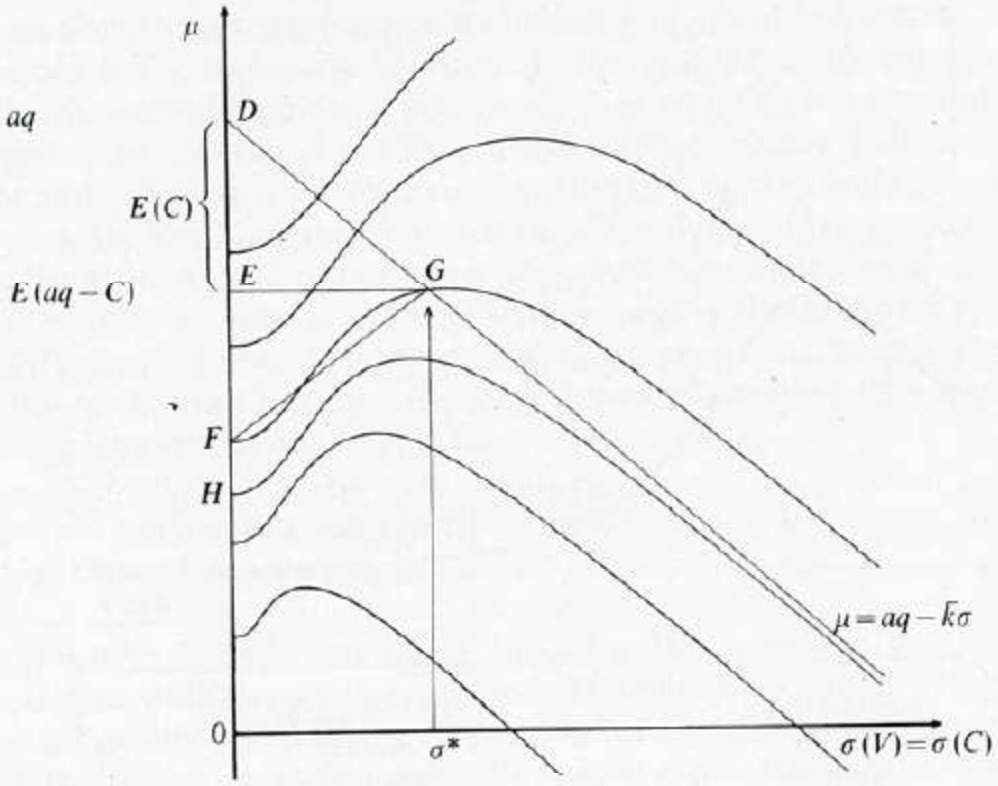


*Figure* 14

Consider first the indifference-curve system of Figure 11 in chapter III B, which depicts the case of weak risk aversion ($0 < \varepsilon < 1$). In Figure 14 this indifference-curve system and the straight line described by (13) are plotted. As before, the maximum willingness to pay can be discovered by finding on the straight line the point $G$ which belongs to a given $\sigma(C) = \sigma^*$ and by measuring the distance between the intercept $D$ of the line with the ordinate and the starting point $F$ of the indifference

curve passing through $G$. Since the vertical distance $\overline{ED}$ between $D$ and $G$ is the expected loss, the intensity of insurance demand can be seen directly from the diagram. Obviously it is given by the quotient $\overline{FD}/\overline{ED}$. By carrying out this procedure for alternative values of $\sigma$ a functional relationship between the expected loss $E(C) = \bar{k}\sigma(C)$ and the intensity of insurance demand $g$ can be established. A graphical illustration of this relationship is given by the unbroken line depicted in Figure 15. Because of the homotheticity of the indifference-curve system as implied by Weber's law, there is not only a stable functional relationship between $g$ and $E(C)$ for any given $aq$, but also between $g$ and the standardized expectation of the loss, $E(C)/aq$. The labelling of the abscissa takes account of this fact.

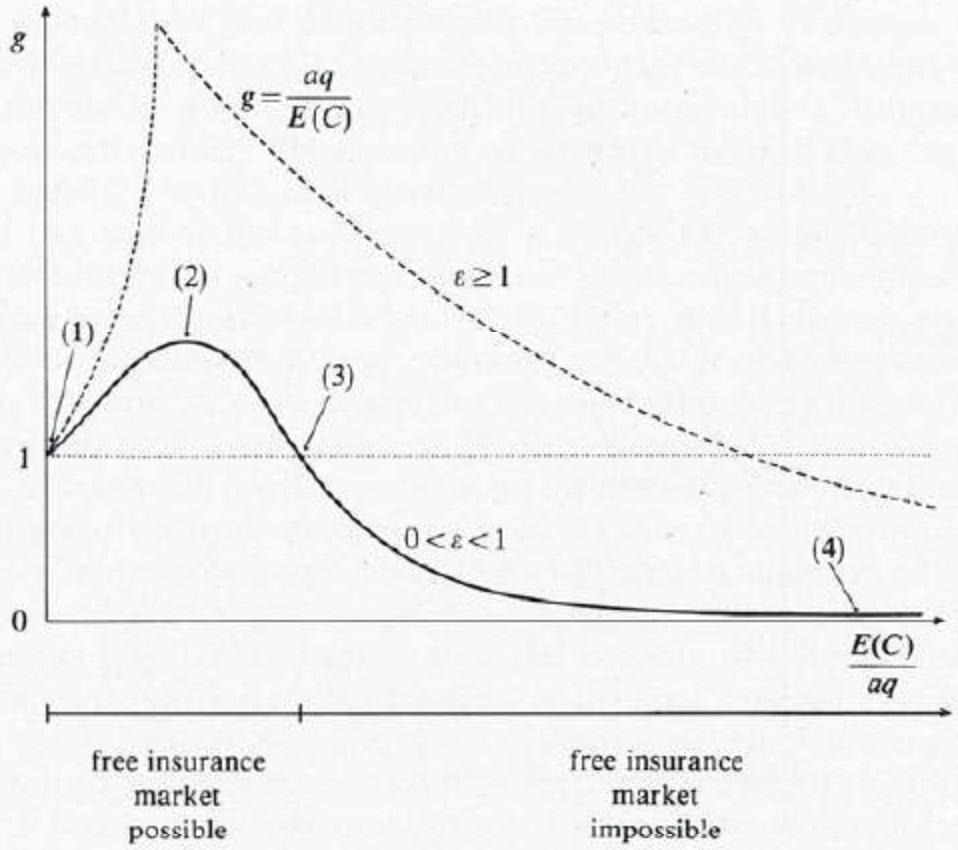


*Figure* 15

The following properties of the function $g = g[E(C)/aq]$ which are also labelled in Figure 15 hold in general:

(1) $g \to 1$ for $E(C)/aq \to 0$.
(2) $g$ is maximal if $E(C)$ determines point $G$ on line (13) such that it is situated on an imaginary line connecting the points of inflexion of the indifference curves. This is the case when the highest possible

loss is already greater than the normal wealth $aq$ so that the BLOOS rule is operative.

To prove these properties note that

$$g = \frac{\tan \sphericalangle FGE + \tan \sphericalangle EGD}{\tan \sphericalangle EGD},$$

where $\tan \sphericalangle EGD = \bar{k} = \text{const.}$ and

$$\tan \sphericalangle FGE = \frac{\int_0^{\sigma^*} \left. \frac{dE(V)}{d\sigma(V)} \right|_{U = U(\overline{OF}, 0)} d\sigma(V)}{\sigma^*}.$$

According to this expression, the intensity of insurance demand is a monotonically increasing function of the average indifference-curve slope between the points $F$ and $G$ which is equal to the slope of the straight line between $F$ and $G$. (1) then follows directly from the fact that all indifference curves enter the ordinate[10] perpendicularly, which implies risk neutrality in the evaluation of small risks. (2) is explained by the fact that, with an increase in $\sigma^*$ the average indifference-curve slope is increasing as long as point $G$ remains within the range of normal indifference curves. This range, as is known, ends at a value of $E(V)/\sigma(V)$ where the gross wealth distribution already partly overlaps the negative half of the wealth axis.

(3) $g < 1$ obtains at a value of $E(C)/aq < 1$.
(4) $g \to 0$ for $E(C)/aq \to \infty$.

The reason for (3) and (4) is that, for the insurance purchaser's maximum willingness to pay, there is an upper limit, smaller than the normal wealth, which can never be exceeded regardless of the size of the standard deviation and of the mathematical expectation of the loss distribution. In Figure 14 this upper limit is given by the distance $\overline{DH}$ for, at $H$, the indifference curve, to which the straight line (13) is an asymptote, enters the ordinate. The existence of an indifference curve of this type follows from the facts that 1. the indifference-curve slope is everywhere greater than $-\bar{k}$ (cf. chapter III B 1.1.2) but approaches this value asymptotically as $\sigma(V) \to \infty$, that 2. the straight line (13) has the slope $(-\bar{k})$, and that 3. below the line (13) there are indifference curves entering the positive half of the ordinate.

[10] Cf. expression (II D 62).

With this analysis of Figure 15, for an arbitrary linear distribution class[11] the proof has been given that the intensity of insurance demand may be insufficient to allow for a contract profitable from the viewpoint of the insurance company when there are large liability risks. A free market for the insurance of large liability risks therefore cannot be expected to operate, a result that is confirmed in practice. A good example is automobile liability insurance in states where insurance is not compulsory. There, many people are wont to buy comprehensive insurance, covering damage to themselves, rather than liability insurance, covering damage to others.

The result summarized in Figure 15 was derived for the probably realistic[12] case of weak aversion $\varepsilon < 1$. In a similar way we can now study the implications of the indifference-curve system that prevails in the case $\varepsilon \geq 1$. For classes of wealth distributions bounded to the left such an indifference-curve system is depicted in Figure 12 of chapter III B. The result is a shape of the function $g[E(C)/aq]$ as shown by the broken line in Figure 15. Suppose the unbroken line and the broken line refer to the same linear distribution class. Then the broken line also starts at $g = 1$, but for $E(C) > 0$ it is always above the unbroken line. Provided that the highest possible loss is less than the normal wealth, this follows from the fact that, on given rays through the origin, the indifference-curve slope is a rising function of the measure of relative risk aversion $\varepsilon$ (cf. equations (III A 53) and (III B 18)). If, however, the highest possible loss can exceed normal wealth, then the reason is the lexicographic preference ordering between $v \leq 0$ and $v > 0$ implying that, if necessary, the decision maker is willing to sacrifice nearly all his wealth to obtain insurance protection[13]. The intensity of insurance demand then would be almost[14] $g = aq/E(C)$. Of course the intensity of insurance demand even here will eventually fall short of unity when $\sigma(C)$ is sufficiently large, that is, when $E(C) > aq$, but this does not seem to be a very relevant case.

### 1.3. *The Optimal Degree of Coverage*

Suppose the decision maker has the opportunity of choosing the optimal degree of coverage $\theta$, $0 \leq \theta \leq 1$, given the loading factor $\tilde{g}$. Then the

[11] For the special case of a linear class of binary distributions the proof was given in chapter III B 1.1.

[12] Cf. chapter III B 2 (middle), chapter IV B 2.3.3, and section 1.4 that follows below.

[13] The broken line in Figure 15 is assumed to make no jumps. This, however, is not a general property as is evident from equation (III B 21). Instead, in the case $1 \leq \varepsilon < 2$, the intensity of insurance demand has to jump to the value $aq/E(C)$ if $\mu/\sigma = \underline{k}$, i.e., if $\sigma = \mu/\underline{k}$, and if the wealth density function is truncated at the left such that $f(-\underline{k}+) > f(-\underline{k}-) = 0$.

[14] Cf. footnote 9 above.

decision problem can be illustrated as in Figure 16 which was constructed in a similar way to Figure 13. Figure 16 refers to the case of convex indifference curves which is typical for property insurance but may also be relevant for small-scale liability insurance.

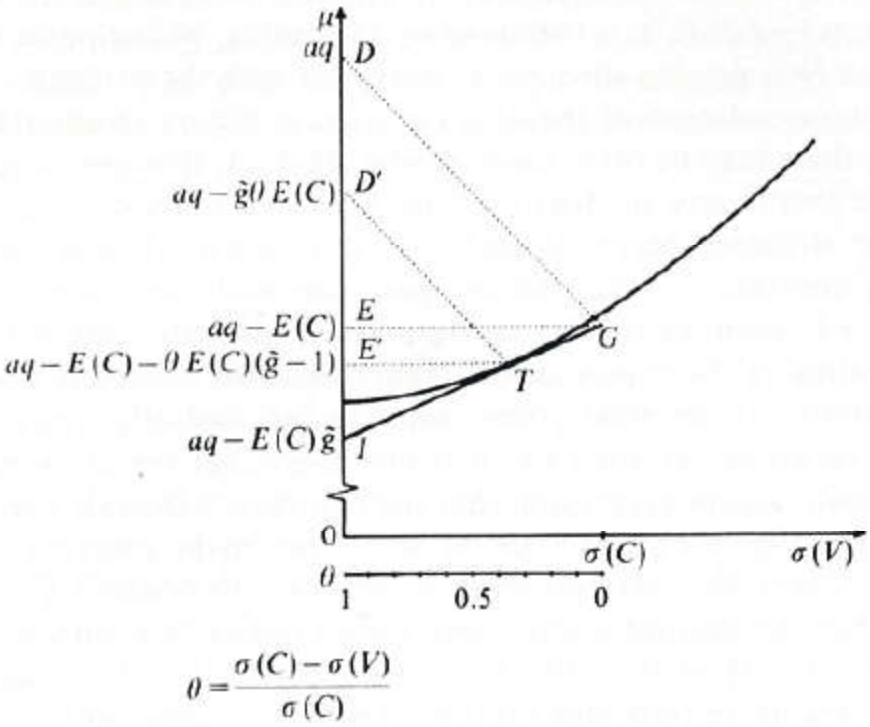


*Figure* 16

If the decision maker chooses $\theta = 0$, then, as before, the end-of-period wealth distribution is represented by point $G$. But if he chooses a higher degree of coverage, he achieves a distribution with a lower dispersion. This distribution is mapped somewhere on the line $\overline{GI}$. The properties of this line, call it *insurance line*, can be found by calculating $\theta$ in equation (7) and inserting the result into (6):

$$E(V) = [aq - E(C)] - \left[\frac{\sigma(C) - \sigma(V)}{\sigma(C)} E(C)(\tilde{g} - 1)\right]. \tag{14}$$

The slope of the insurance line is positive if the insurance company requires a price above the expected loss ($\tilde{g} - 1 > 0$). An increase in the degree of coverage $\theta$ in this realistic case thus not only brings the advantage of a lower standard deviation, but also the disadvantage of a reduction in the expected level of end-of-period wealth. The insurance line ends at the ordinate for, if the highest possible degree of coverage ($\theta = 1$) is chosen, the end-of-period wealth distribution degenerates to the non-random value $aq - \tilde{g}E(C)$. In the usual way, the optimal distri-

bution is given by that point on the insurance line which is situated on the highest indifference curve. In Figure 16 this is the point of tangency $T$. The corresponding degree of coverage can be read from the scale plotted parallel to the abscissa, which embodies the relationship $\theta = 1 - \sigma(V)/\sigma(C)$ from equation (7). The size of wealth in the absence of losses, $aq - \tilde{g}\theta E(C)$, is indicated by $D'$, that is, by the point where a parallel to $\overline{DG}$, passing through $T$, intersects with the ordinate.

An interior solution of the kind depicted in Figure 16 should be the rule, but there may be other cases as well. If $\tilde{g} \leq 1$, then the slope of the insurance line is zero or negative and, because the indifference curves enter the ordinate perpendicularly, it is optimal to demand a full coverage contract[15]. Since the company normally requires a loading factor $\tilde{g} > 1$, such a case can be observed in practice only if the purchaser estimates the expected loss higher than the company does. For the possibility of the other corner solution, no similarly simple condition can be given. At any rate, it is not likely that the insurance purchaser would accept everything that the company offers. It was shown that the indifference-curve slope at any point in the diagram is lower than the slope of the corresponding ray through the origin[16]. This aspect implies that the insurance purchaser would not bother with insurance protection if $\tilde{g}$ is sufficiently close to $aq/E(C)$, for, if $\tilde{g} = aq/E(C)$, point $I$ in Figure 16 coincides with the origin of the diagram, that is, the insurance line would be a straight line through the origin. So much for the optimal degree of coverage for the range of convex indifference curves.

An extension of the analysis to the total range is first carried out for weak risk aversion ($0 < \varepsilon < 1$), where the indifference curves become concave and take on negative slopes provided that the loss distribution extends far enough beyond the level of normal wealth $aq$. Figure 17 illustrates an example of the decision situation where the end-of-period wealth distribution without insurance is depicted in the abnormal indifference-curve range.

In Figure 17 three alternative loading factors $\tilde{g}$ have been assumed, and accordingly there are three different insurance lines, $\overline{GI'}$, $\overline{GI''}$, $\overline{GI'''}$. As in Figure 16, in all three cases, there are tangency solutions indicating

---

[15] This result was found by Mossin (1968b).

[16] In the limiting case $\mu = \underset{\sim}{k}\sigma$ when the lower bound of the wealth distribution coincides with $v = 0$, then, under strong risk aversion ($\varepsilon \geq 1$), the slopes *may* be equal, i.e., we may have

$$\lim_{\mu/\sigma \to \underset{\sim}{k}} \left.\frac{d\mu}{d\sigma}\right|_U = \frac{\mu}{\sigma} = \underset{\sim}{k}.$$

Cf. equations (III A 49) and (III B 21).

local utility maxima. Compared to the normal case however, the special aspect here is that the local maxima are not necessarily global maxima. The point of tangency $T'$ does represent a global maximum. But the maximum at point $T''$ is not unique because this point is situated on the same indifference curve as point $G$. If the loading factor $\tilde{g}$ is increased even beyond the value corresponding to $T''$, then the point of tangency $T'''$ is obtained. This point is on a lower indifference curve than point $G$ and hence the local maximum is not a global one: the optimal degree of coverage 'jumps' to the value of zero.

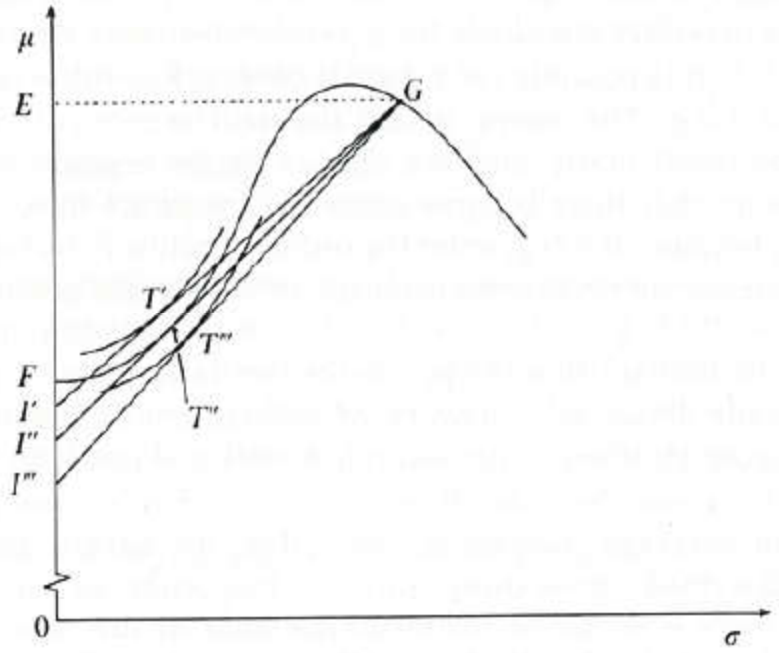


*Figure* 17

The reason the local maximum does not necessarily coincide with the global one is that, starting from zero, an increase in the degree of coverage initially implies a decrease rather than an increase in utility. With an increase in $\theta$ the premium to be paid to the company always increases in strict proportion. However, the additional protection bought with each further unit of coverage at best is only partly beneficial to the purchaser. Some part of the protection is useless to him since it simply brings about an absolute decrease in the negative variates of the gross wealth distribution. This decrease benefits those who, in the absence of insurance, are endangered by the non-redeemable part of the liability risk, but, because of the BLOOS rule, it does not improve the insurance purchaser's net wealth distribution. Only when the degree of coverage is high enough to prevent gross wealth from becoming negative, will an additional unit of coverage fully benefit the purchaser and, despite the

increase in the premium, increase his expected utility. Whether this increase can offset the initial decrease is an open question.

Some information helpful in providing an answer can be obtained from a comparison between the loading factor $\tilde{g}$ required by the company and the intensity of insurance demand $g$. As with the all-or-nothing supply on the part of the company, $\tilde{g} < g$ is once more a sufficient condition for making the purchaser prefer insurance. The reason is simply that an extension of the opportunity set by the inclusion of partial coverage contracts cannot prevent full coverage from being more attractive than no coverage. However, unlike the previous case, $\tilde{g} \leq g$ is no longer a necessary condition for a positive insurance demand, for, in the case $g > 1$, it is possible for a partial coverage contract to be attractive despite $\tilde{g} > g$. The reason is that the indifference curves enter the ordinate perpendicularly and are convex in its neighborhood. Both aspects ensure that there is some scope for insurance lines of the type $\overline{I'G}$ which, because of $\tilde{g} > g$, enter the ordinate below $F$, but which intersect indifference curves that are situated above the one passing through $F$ and $G$.

It could be thought that the possibility of the company's requiring a loading factor above the intensity of insurance demand reduces the range in Figure 15 where a free insurance market is possible. But unfortunately this is not the case. When $g \leq 1$, so that a mutually advantageous full coverage contract is impossible, the advantage of partial coverage described above disappears. By inspection of Figure 14 it is clear that there is no scope for insurance lines of the type $\overline{I'G}$ in this case. It is true that, with $\tilde{g} > 1$, partial coverage is still better than full coverage, but, at the same time, no insurance at all is better than the best possible partial contract. Thus the pessimistic impression that Figure 15 gives concerning the possibility of a free insurance market for large liability risks is not dispelled when partial coverage contracts are possible. So much for the case of weak risk aversion.

If, with $\varepsilon \geq 1$, there is strong risk aversion, then a completely different picture appears. Because of $\lim_{v \to 0} U(v) = -\infty$, the decision maker tries in this case to avoid the total loss in wealth regardless of the price. For a loss distribution unbounded from above, this aim requires a coverage of 100%, but if we realistically assume that the loss distibutions are bounded, smaller degrees of coverage will be sufficient. Say that the standardized loss distribution is bounded from above at $\underline{k}$ and hence the standardized end-of-period wealth distribution is bounded from below at $-\underline{k}$. Then all degrees of coverage that, in the $(\mu, \sigma)$ diagram, lead to points to the left of the line $E(V) = \underline{k}\sigma(V)$ satisfy the decision maker's lexicographic aim. From (6) and (7) we can therefore calculate a lower boundary to the degree of coverage:

$$\theta^* = \frac{\underline{k}\sigma(C) + E(C) - aq}{\underline{k}\sigma(C) + E(C) - \tilde{g}E(C)}. \tag{15}$$

The degree of coverage $\theta$ that should be chosen in the range $\theta^* \leq \theta \leq 1$ can easily be determined. If $\tilde{g} \leq 1$, it is at any rate optimal to buy full-coverage insurance[17], but, in the realistic case $\tilde{g} > 1$, the degree of coverage must always be in the range $\theta^* \leq \theta < 1$. It is worth noting that, contrary to what is suggested by the normal shape of a demand curve, $\theta$ is not a monotonically falling function of the loading factor $\tilde{g}$. It is true, the convexity of the indifference curves and their zero slope at the ordinate imply that, when $\tilde{g}$ increases slightly beyond 1, there is a fall from $\theta = 1$ to $\theta < 1$. But, since (15) gives

$$\lim_{\tilde{g} \to aq/E(C)} \theta^* = 1, \tag{16}$$

the degree of coverage ultimately has to increase again.

If people were usually as risk averse ($\varepsilon \geq 1$) as was just assumed then we need not fear that a lack of demand will impede the liability insurance market. But probably the case of strong risk aversion is not a very relevant one. Why that is so is shown in the following section.

### 1.4. *The Age Dependence of Insurance Demand*

The basic multiperiod model established in chapter IV B implies that risk aversion depends on age. Equations (III A 53) and (III B 18) say that, with an increase in the degree of relative risk aversion $\varepsilon$, the slopes of the non-pseudo indifference curves rise at each point in the $(\mu, \sigma)$ diagram[18,19]. It is straightforward to interpret these pieces of information for the insurance problem.

If $\varepsilon > 1$ so that risk aversion decreases with age then, with the passage of time, there is a decrease

- in the intensity of demand for an insurance of property risks
- in the optimal degree of coverage if initially there was a tangency solution with $\theta > 0$.

The intensity of insurance demand for liability risks which are large enough to include the possibility of negative gross wealth is unaffected,

[17] Provided that $E(C) < aq$. Otherwise, in the multiperiod approach, the optimal degree of coverage may not exist.

[18] Except at the ordinate.

[19] We forgo the proofs for the results reported in what follows.

for, as long as $\varepsilon \geq 1$, the decision maker is always willing to give away almost all his wealth in exchange for insurance protection.

If, however, the case $\varepsilon < 1$ prevails, where relative risk aversion increases over time, then there is a gradual increase

- in the intensity of demand for an insurance of all kinds of loss distrition
- in the optimal degree of coverage if initially there was a tangency solution with $\theta < 1$.

For completeness it should be mentioned that, in addition to the possibility of a continuous increase in the degree of coverage for interior solutions, it is also possible for $\theta$ to jump from zero to some positive level. For this jump to occur, a local maximum must change into a global one when initially there is an insurance line like $\overline{I'''G}$ in Figure 17.

According to everyday experience, and also according to a poll carried out by GREENE (1964, cf. exp. p. 36), risk aversion increasing with age seems to be the normal case. Thus, we must consider weak risk aversion ($\varepsilon < 1$) to be the standard case and unfortunately state that the doubts concerning the workability of the liability insurance market gain additional weight[20].

## 2. *Insurance and the Size of Risk*

With an analysis of insurance demand for given risks only one aspect of the economic meaning of insurance has been elucidated. The other one, which is probably even more important, is the insurance-induced change in people's behavior that alters the sizes of the risks underwritten by the companies[21].

It is well known that, after buying an insurance contract, people tend to become very careless, sometimes even going so far as to destroy the insured object deliberately. Insurance may however also induce people to stop undertaking risky activities. We shall see that this possibility arises if compulsory insurance is introduced for liability risks. In what follows, such changes in behavior will be studied. Section C 2.1 analyzes the allocative effects of insurance under ideal conditions, section C 2.2 considers what is called reproachfully *moral hazard,* and section C 2.3 examines the allocative implications of insurance in the case of liability risks when the BLOOS rule is operative.

---

[20] Cf. Figure 15 above.

[21] The present section draws heavily on previous work by the author. See SINN (1977, 1978). However, by analyzing liability risks this study goes further.

### 2.1. *The Insurance-Induced Substitution Effect under Ideal Conditions*

If insurance causes carelessness in dealing with risks, misallocation would appear to be present, but appearances can be deceptive. In fact, the appearance of misallocation arises from a misunderstanding that we now attempt to remove[22].

We consider simultaneous decisions about the optimal degree of insurance coverage and about activities that influence the size of the risk to be insured. Such activities include the installation of sprinklers or burglar alarms, the use of fire-proof materials, the purchase of safes, and many others. The activities have two things in common, first, they bring about costs ($b$) which reduce the level of end-of-period wealth even if the decision maker is lucky enough to avoid any damage and, second, they reduce the levels of loss for given probabilities and reduce the probabilities for given loss variates[23]. The set of all end-of-period wealth distributions attainable through such manipulations is denoted the *original opportunity set* ($M$). Assuming this set contains only distributions which belong to the same linear class, we can meaningfully depict it in a $(\mu, \sigma)$ diagram. Figure 18 shows an example.

For each point in $M$ there is a downward sloping ray like $\overline{DG}$ connecting this point with a point on the ordinate which indicates the corresponding level of normal wealth net of prevention costs, $(a-b)q$. The significance of this ray is the same as that of the identically labelled rays in Figures 13, 14, and 16. Formally the ray is described by the equation

$$E(V) = (a-b)q - \bar{k}\sigma(V) \tag{17}$$

which is similar to equation (13) above. By the assumption of a linear distribution class there is the same $\bar{k}$ for all distributions from $M$ and hence all rays are parallel to one another. This property permits an interesting comparison of those distributions that are plotted on the upper boundary of the opportunity set. Let $P$ denote that point where the highest possible ray is tangent to the opportunity set and let the point where this ray reaches the ordinate indicate the level of normal wealth $(a-b)q$ achieved with $b=0$. Then, left of $P$, a movement to the right reduces loss prevention costs $b$, but raises the expected loss $E(C)$

[22] It seems that, e.g., GRUBEL's (1971) study is not completely free from this misunderstanding.

[23] EHRLICH and BECKER (1972) call the former 'self-insurance' and the latter 'self-protection'. Although these possibilities have a clear meaning for the binary distributions considered by these authors they may be indistinguishable in the case of multivariate distributions.

and the level of risk $\sigma(V)$. Such a movement, therefore, represents the intuition behind the above examples. At point $P$, a further reduction in loss prevention activities is impossible. To reach the region to the right of $P$, it would be necessary for the decison maker to bear costs if he tried to enlarge the loss distribution beyond its 'natural' shape.

The opportunity set depicted in Figure 18 refers to the current choice problem of the decision maker that we are going to analyze. Similar opportunity sets are available at all later points in time up to the horizon. In line with our basic multiperiod approach, it is assumed that these opportunity sets satisfy the requirement of stochastic constant returns to scale. To illustrate this requirement suppose that, by chance, at the beginning of the next period wealth net of consumption turns out to be half as large as in this period. Suppose also that in the next period a particular type of prevention policy as represented by a point in $M$ is chosen. Then the prevention cost $b$ as well as the loss associated with any variate of the loss factor $F$ from (2) is half as large as it would have been had wealth not changed.

For the time being it is assumed that, at least in the current period, the opportunity set does not intersect with the range of abnormally sloped indifference curves where the BLOOS rule is in operation. This assumption will be removed in section 2.3.

Without insurance the decision maker's optimal choice is point $S$ in Figure 18. The question is how this choice is affected if insurance protection is available for all elements of the original opportunity set.

To find an answer we first have to find out how the original opportunity set is enlarged when insurance is possible. Assume that partial coverage at a degree $\theta$ in the range $0 \leq \theta \leq 1$ is allowed and that the company practices *equivalence rating*. It monitors all loss prevention activities of the purchaser, reckons with the same probability distribution of losses as he does, and sets a premium loading factor $\tilde{g} > 1$ to ensure that the premium required is above the expected loss underwritten. Under these conditions, for each point in the opportunity set $M$, an insurance line of the kind introduced in Figure 16 can be constructed.

There is then a twofold decision problem for the potential purchaser. First he has to determine the best choice from $M$, thus deciding on a particular one of the possible alternative insurance lines. Then, in the usual way, he has to determine the best point on this insurance line by choosing an appropriate value of $\theta$. Analytically, the first part of the problem can easily be solved. Replace in equation (14) $aq$ by $(a-b)q$ and calculate the slope

$$\frac{dE(V)}{d\sigma(V)} = \frac{E(C)}{\sigma(C)}(\tilde{g}-1), \tag{18}$$

of an insurance line. Since the prevalence of a linear distribution class, as with (12), implies $E(C)/\sigma(C) = \bar{k} = \text{const.}$ it turns out that all achievable insurance lines are parallel. This fact ensures that there is one insurance line that is unambiguously the best. Obviously it is the highest one.

In Figure 18 this line is labelled $\overline{IG}$. It is tangent to the opportunity set $M$ at point $G$ and terminates at this point. On this line, the best point which indicates the end-of-period wealth distribution generating the highest level of expected utility is the point of tangency $T$ with an indifference curve. It corresponds to point $T$ in Figure 16.

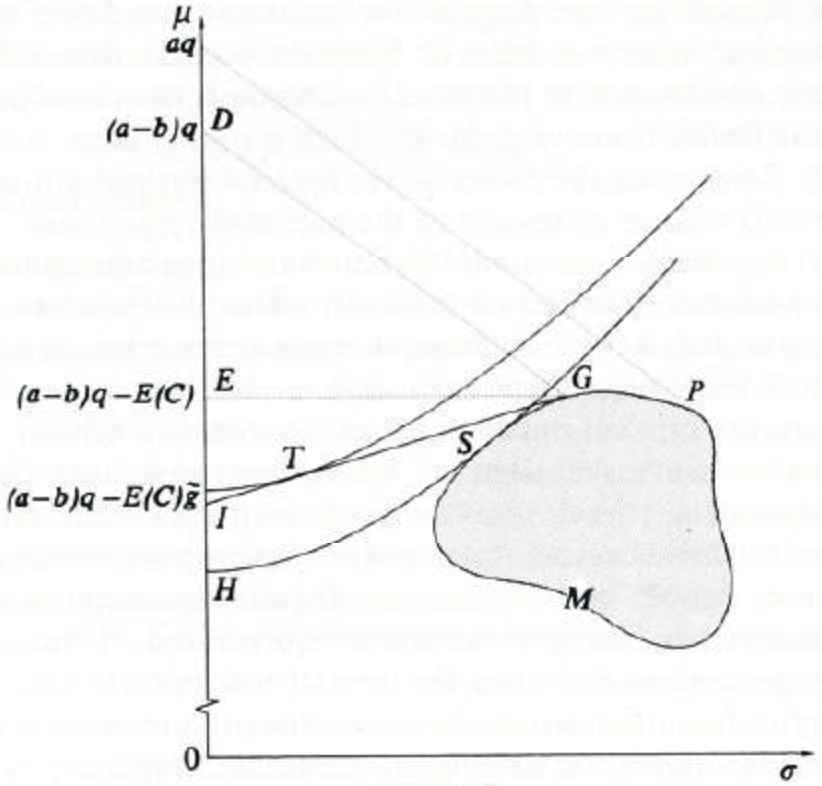


*Figure* 18

An interesting aspect of this optimization procedure is that, as with Tobin's Separation Theorem for portfolio analysis, the choice of the best distribution from $M$, i.e., the choice of the optimal loss prevention policy, can, to a large extent, be made regardless of the decision maker's preferences[24]. The preferences merely determine whether or not insurance is demanded at all. Obviously, a positive demand occurs if, and only if, the slope of the insurance line is smaller than the indifference-curve slope at point $S$. Provided a decision for insurance

[24] With a different approach a similar result was achieved by EHRLICH and BECKER (1972, pp. 636 f.).

demand has been made, the only 'task' of the preferences is to determine the optimal degree of coverage. They do not affect the position of point $G$ and hence do not affect the loss prevention policy of the insurance purchaser.

Up to now, it was assumed that partial coverage is possible. It may, however, also be of interest to consider the possibility that the insurance company makes an all-or-nothing offer ($\theta = 0, \theta = 1$). Although in this case only the end points of the insurance lines are relevant, the highest insurance line clearly remains the best one. Thus, as with optimal partial coverage, the loss prevention policy represented by point $G$ in Figure 18 is chosen if insurance is bought. Whether insurance is bought, however, no longer depends on the slope of the insurance line being below the indifference-curve slope at point $S$. Instead a positive demand requires the stronger condition that the best, i.e., highest, insurance line enters the ordinate (point I) above point $H$ which is on the same indifference curve as $S$. Concerning the choice of the optimal loss prevention policy, this is the only change compared to the partial-coverage case.

Initially a promise was made to remove a misunderstanding concerning the allocative effects of insurance. The time has now come to keep it. It was shown that, with the purchase of insurance, a movement from $S$ to $G$ takes place, that is, a shift to the right along the upper boundary of the original opportunity set. Indeed, this type of shift is a general feature since the slope of the optimal insurance line is, of necessity, lower than the slope of the upper boundary at the initial point $S$. The result clearly confirms the initial conjecture that, with a purchase of insurance, people become careless about preventing losses. Loss prevention cost $b$ is falling with the consequence that there is a rise in both the expected loss $E(C)$ and the level of risk $\sigma(V)$.

Contrary to first impressions there is nothing in the result that allows us to blame insurance for causing misallocation. Suppose, because of government controls or of competition, the premium loading factor required by the company is just sufficient to compensate for the burden of the indemnification claims. Then, with the insurance purchaser's choice, a Pareto optimal situation is reached, given the kinds of contracts described. The reason is that by assumption the company is indifferent to this choice while the purchaser cannot reach a higher indifference curve than the one resulting from his own decision. In particular, the decision maker would be comparatively worse off if he did not change his loss prevention policy and merely chose the best point available on the insurance line passing through point[25] $S$. From an allocative

[25] This line has not been plotted in Figure 18.

point of view, the reduction in care is a desirable outcome of insurance. To reduce the risk by pooling is cheaper than to reduce it through expensive protection measures[26].

It is not difficult to find examples for the favorable allocative effects of insurance. One very striking example was the development of insurance in medieval Venice[27]. A Venetian merchant who sent a ship to foreign harbors was engaged in a risky business, since he often lost both ship and cargo. Thus, for a long time, the risk was prohibitive and the journeys went no further than the neighboring coasts. But at some point it proved advantageous to shift the burden of risk on to the shoulders of speculators who could consolidate risks by employing the Law of Large Numbers[28]. This increased the Venetian risk bearing capacity to such an extent that merchants were able to venture much further and the Venetian fleet came to dominate the whole Mediterranean.

### 2.2. *Moral Hazard*

The set of problems that lie behind the concept of moral hazard has been the subject of intensive discussions in insurance theory. What the concept means is an insurance-induced change in the purchaser's behavior that is to the disadvantage of the company and ultimately also to himself. The behavioral change may be to exploit the contract beyond what was intended or may involve insurance fraud as, for example, in the case of deliberate destruction. At any rate, the concept suggests what is ordinarily regarded as dishonest behavior[29].

Rather than just pointing a moral finger, we should try to find an economic explanation for the observable fact that, when insurance purchasers make rational calculations, allocation patterns emerge that are different from the one described in such favorable terms in the previous section. In principle there seem to be three categories of economically motivated moral hazard: deliberate destruction of the insured object, too much demand brought about by the cost-compensation principle, and excessive carelessness associated with community rating. All three

---

[26] Although this result by no means coincides with the usual view of the allocative effects of insurance, it was anticipated by MAHR (1951, esp. pp. 88 f. and 91 f.) and ARROW (1970, pp. 137 f.). Implicitly, it is also contained in the article by EHRLICH and BECKER (1972, esp. pp. 636 f.). For a contrary view typical among insurance brokers see SLANEC (1972, pp. 16).

[27] Cf. PERDIKAS (1966).

[28] Cf. chapter IV A.

[29] An overview of the moral-hazard literature with particular emphasis on the definition of moral hazard was given by MAHR (1972). Perhaps the first analysis of the moral hazard problem was provided by HAYNES (1895, pp. 445 f.).

categories will be treated in what follows. The first two are only indirectly related to a risk problem. They are, nevertheless, considered here because they can help in sorting out and understanding the third problem as well as the insurance-induced behavior change considered in the previous section.

2.2.1. Deliberate Destruction of the Object Insured

This type of moral hazard is the most obvious one. The insurance buyer can profit from a deliberate destruction of the object insured or by worsening existing damage since the payment from the company overcompensates for the loss. Arson is a well-known example. By the end of the last century HAYNES (1895, p. 445) reported that, in the United States, 35–50% of fire insurance damages were due to arson. Today also, the increase in the damages caused by fire during recessions can be explained this way. Of course this effect indicates a clear misallocation. If all the insured behaved this way, each would obtain a compensation which could not exceed his premium, but the objects insured would be destroyed. Fortunately, however, there is a very simple way of avoiding this misallocation. The company only has to take care that no loss is covered by more than[30] 100%. This might be difficult in practice, but it does not seem to be a theoretical problem.

2.2.2. The Excess Burden of the Cost-Compensation Principle

In this section we study an insurance-induced behavior change which, unlike insurance fraud, is of great practical importance. Its essence is that there is an inflated demand for repairs in the case of damage, or for medical attention in the case of illness[31]. The example of the insured car

---

[30] Cf. HAYNES (1895, pp. 445 f.), FISHER (1906, pp. 294 f.), and ARROW (1970, pp. 142 and 148). Another type of insurance where there is an incentive for deliberate destruction is the one where a person makes a contract that provides him with indemnification payments when someone else is damaged. Here the insured person can make a gain of the difference between the insurance value and the premium, if he causes the damage to the third party. FISHER (1906, pp. 294–295), for example, reports the so-called 'graveyard insurance' which was possible in the United States. Basically it was to take out an insurance contract on the lives of other people. It is not difficult to imagine the gruesome form insurance fraud took on with this kind of insurance. The judgement of its allocative value is of course also obvious. Fortunately, this type of insurance has no practical importance today since it is forbidden by law in most countries.

[31] Theoretically this seems to have been analyzed first by PAULY (1968). Cf. also ZECKHAUSER (1970), SPENCE and ZECKHAUSER (1971), GRUBEL (1971), FELDSTEIN (1973), and ROSETT and HUANG (1973). As far as these authors are concerned with the welfare loss of the cost-compensation principle, their approaches can be criticized since they calculate the loss in terms of consumer rents. The indifference-curve analysis utilized here avoids the strong assumptions necessary for such a calculation.

owner who has his whole car sprayed at the company's expense, although there is only a small scratch, and the housewife who every month must have her cosy chat with the doctor are certainly familiar. Less familiar to many readers may be the fact that in West Germany there are more than 250 spas that depend for their existence on the generous support for rest cures provided by the insurance system[32]. In all these cases, the reason for an excessive demand is that the insurance companies do not pay unconditional money compensation as was implicitly assumed up to now, but make the compensation dependent on the costs of repairs or recovery, even sometimes paying the whole of these costs.

The decision problem of the person insured in the case of damage can be illustrated with the aid of Figure 19, which shows an indifference-curve system for goods $x$ and $y$, where $x$ measures the number of repairs units and $y$ the person's wealth, which is a quantity index of all the other

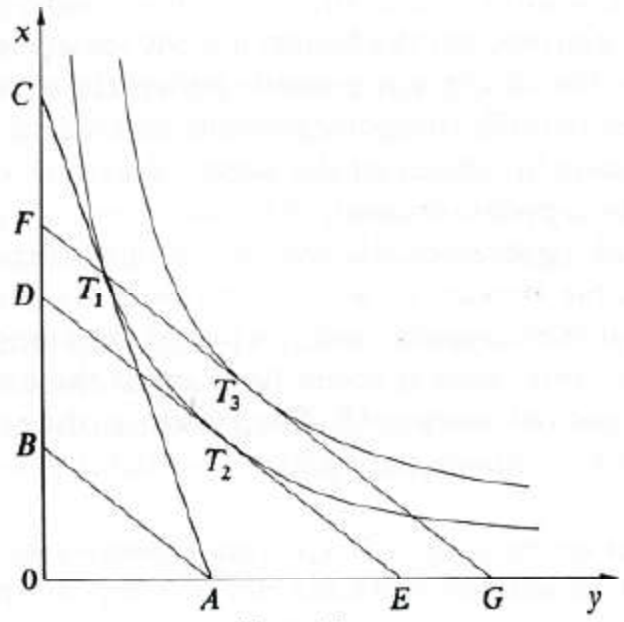


*Figure* 19

[32] Here, too, an idea of the importance of moral hazard can be gained by considering the relationship between the number of insurance claims and the level of employment. For example, in West Germany, from winter 1974 to winter 1975 (recession) the number of rest cures in the famous government spas Oeynhausen and Meinberg fell by 21% and 30%, respectively. The reduction was probably only caused by members of the private work-force, for the number of government employees applying for financial support (Beihilfe) for rest cures during the same time increased by 4%. A further indicator of the importance of moral hazard is that the demand of government employees was not only more stable than that of private employees, but was also higher. For example, in 1974, 11% of the employees in the Ministry of the Interior took their rest cures, but only 4% of all members of the social insurance system including government employees did so. Cf. PIEL (1976).

goods he consumes now or in the future. The indifference-curve system is contingent on a particular insurance damage having occurred. As usual the indifference curves are assumed to be convex. Before having any repairs made and before receiving compensation from the company, the decision maker is at point $A$. If he has repairs made without being compensated for any of the cost, then he can move along budget line $\overline{BA}$, the position of which is determined by the constant competitive price of a repairs unit[33]

$$P_r = \overline{OA}/\overline{OB} \tag{19}$$

and wealth $\overline{OA}$. Things are different if the company pays back the share $\theta$ of the documented costs. In this case, the cost compensation principle reduces the net price ($P_{rn}$) for the insured to

$$P_{rn} = (1-\theta)P_r = \overline{OA}/\overline{OC} \tag{20}$$

with $\theta = \overline{BC}/\overline{OC}$, so that the budget line moves to the new position $\overline{AC}$. Now, on this line the person insured chooses the commodity bundle $T_1$, which implies that the company pays the amount of money $\overline{AG}$[34].

Unfortunately the choice of the bundle $T_1$, which is optimal from the viewpoint of the person insured, indicates a clear misallocation. Had the company paid the amount $\overline{AG}$ unconditionally, then the insured would have chosen the bundle $T_3$ which represents a higher utility than $T_1$. Conversely, if the company had not linked the compensation payment to the repair costs, then it could have saved the amount $\overline{EG}$ without making the insured worse off. The reason is that the person insured would then have chosen the bundle $T_2$ which he likes as much as the bundle $T_1$.

The useless excess costs $\overline{EG}$ are undoubtedly a burden on the insurance market. In general, when the BLOOS rule is not in operation, a risk averse insurance purchaser should be willing to pay a premium beyond the expected monetary indemnification payment from the company, provided that this payment is unconditional. In order to make this result applicable if the person insured is indemnified according to the cost-compensation principle, we can transform the probability distribution of the company's payments into an equivalent distribution of uncondi-

---

[33] It is assumed that the transformation curve between the repairs commodity and the bundle of other commodities can be linearly approximated in the relevant case.

[34] Note that $\overline{OG}$, where $G$ is the intersection with the abscissa of a budget-line through $T_1$, parallel to $\overline{BA}$, measures the monetary value of the bundle represented by $T_1$, but that, after the damage, the insurance purchaser's wealth was only $\overline{OA}$.

tional payments. Then, of course, the fact that the insurance purchaser would be willing to pay a premium in excess of the mean of the latter distribution, by no means indicates that he would also be willing to pay more than the mean of the former distribution, as the company would require.

The allocative evaluation of the cost compensation principle is obvious, since the excess cost $\overline{EG}$ induced by this principle is nothing but a pointless waste of resources. Whether insurance, despite this excess burden, causes a net welfare gain is a question which can only be answered by the insurance market itself. If a free insurance market exists, there must be a net gain for at least one of the parties. In the light of this, public compulsory insurance must be regarded sceptically. When it uses the cost-compensation principle, it may well bring about a net loss in welfare.

Where it is practiced the cost-compensation principle should be abandoned, except where repairs are a merit good or produce positive external effects, as might be the case with health insurance. One must, however, be careful to see that, in this case, the decision maker is no better off after the damage than before since this would induce insurance fraud as discussed above. So, for example, in the case of full-coverage insurance a change in the compensation principle has to be accompanied by a reduction in the indemnification payments by the company, which, of course, makes it possible to reduce premiums, too.

### 2.2.3. The Optimal Loss Prevention Policy under Community Rating

The third type of moral hazard is more subtle than the other two, but is nevertheless of great importance. Only this type is inseparably connected with insurance and creates insurmountable barriers. We thus need to consider it in detail.

The problem of community rating leads us back to the framework of section C 2.1 where the substitution effect of insurance under ideal conditions was studied. An assumption underlying that analysis was that the insurance company exercises equivalence rating. It can monitor the loss prevention policy of the purchaser and can adjust the premium it requires accordingly. Obviously, with this assumption, the quality of the company's information was seen in a too favorable light. In reality, the purchaser always has some scope for manipulating his risk without the company's knowledge. This scope is the third source of moral hazard[35].

[35] Related approaches have been chosen by SPENCE and ZECKHAUSER (1971, p. 383) and PAULY (1974). Cf. also EHRLICH and BECKER (1972, pp. 642 f.), SEIDL (1972), HELPMAN and LAFFONT (1975), MARSHALL (1976), and EISEN (1976). A practically oriented study is given by MENGES (1970, esp. pp. 109 f.). For the related problem of the workability of an insurance market with non-homogeneous risks cf. also PAULY (1970) and AKERLOF (1970).

To see the problem clearly, let us first assume that the insurance company has no possibility of monitoring the loss prevention policy of the purchaser at all. Thus, independently of the distribution the decision maker chooses from the original opportunity set $M$ (see Figure 18), the company has to determine a premium $\tilde{g}$ per unit of coverage $\theta$. It practices *community rating*. To illustrate this case graphically would not be very illuminating. So, an algebraic treatment is used. For this purpose the following notation is introduced:

$\mu_M, \sigma_M$ expectation and standard deviation of a particular end-of-period wealth distribution from $M$,

$\mu_C, \sigma_C$ expectation and standard deviation of the loss distribution belonging to $\mu_M$ and $\sigma_M$,

$\hat{\mu}_C$ the insurance company's estimation of $\mu_C$,

$\mu_V, \sigma_V$ expectation and standard deviation of the end-of-period wealth distribution the insurance purchaser faces when choosing $\mu_M$ and $\sigma_M$ and buying the proportion $\theta$ of insurance coverage.

The definitional equations of the distribution parameters in which the insurance purchaser is ultimately interested are given by (cf. (6) and (7))

$$\mu_V = \mu_M + \theta(\mu_C - \tilde{g}\hat{\mu}_C) \tag{21}$$

and

$$\sigma_V = (1-\theta)\sigma_C. \tag{22}$$

It is worth-while making some transformations in both the equations. Note first that $\sigma_C = \sigma_M$ and, because of (12), $\mu_C = \bar{k}\sigma_C$. Now, define a function $\bar{\mu}_M(\sigma_M)$ that represents the shape of the upper boundary of $M$ in Figure 18; for simplicity it is assumed that $\bar{\mu}_M(\sigma_M)$ is twice differentiable with $\bar{\mu}''_M(\sigma_M) < 0$. Moreover, define a function $\hat{\mu}_C = \hat{\mu}_C(\mu_C)$ which expresses the relationship between the expected loss as calculated by the purchaser and as estimated by the company. With $\hat{\mu}'_C(\mu_C) = 1$ this allows the ideal case considered above to be depicted. With $\hat{\mu}'_C(\mu_C) = 0$ it reflects the case of community rating considered here. Then (21) and (22) can be written as

$$\mu_V = \bar{\mu}_M(\sigma_M) + \theta[\bar{k}\sigma_M - \tilde{g}\hat{\mu}_C(\bar{k}\sigma_M)] \tag{23}$$

and

$$\sigma_V = (1-\theta)\sigma_M. \tag{24}$$

These equations demonstrate that the parameters $\mu_V$ and $\sigma_V$ of the decision maker's end-of-period wealth distribution $V$ depend on two control variables: on the degree of insurance coverage $\theta$ and on $\sigma_M$ which indicates the loss prevention policy. That $\sigma_M$ determines the loss prevention policy, although, given $\sigma_M$, the opportunity set allows for alternative $\mu_M$ values, is caused by the possibility of eliminating as inefficient all distributions that map below the upper boundary of $M$. Suppose for a moment that $\bar{\mu}_M(\sigma_M)$ is not a function but a correspondence that associates $\sigma_M$ with a set of alternative values of $\mu_M$ and consider equations (23) and (24). Since, given $\sigma_M$ and given $\theta$, there is a given value of $\sigma_V$, the highest value of $\mu_M$ is obviously the best one; whatever value is given to $\theta$, the highest value of $\mu_M$ leads to the highest possible value of $\mu_V$.

The aim of the decision maker is to optimize his end-of-period wealth distribution by a suitable choice of $\theta$ and $\sigma_M$:

$$\max_{\{\theta,\sigma_M\}} U(\mu_V, \sigma_V) \quad \begin{cases} \theta = 0, \theta = 1 & \text{(all-or-nothing supply)} \\ 0 \le \theta \le 1 & \text{(partial coverage allowed).} \end{cases} \tag{25}$$

Utilizing the relationship

$$-\frac{\partial U(.)/\partial \sigma_V}{\partial U(.)/\partial \mu_V} = \left.\frac{d\mu_V}{d\sigma_V}\right|_{U(.)}$$

we can calculate from (23)–(25) the necessary conditions for a maximum. From $\partial U/\partial \sigma_M = 0$, we have

$$\bar{\mu}'_M(\sigma_M) + \theta\bar{k}\left(1 - \tilde{g}\frac{\partial \hat{\mu}_C}{\partial \mu_C}\right) = \left.\frac{d\mu_V}{d\sigma_V}\right|_{U(.)}(1-\theta) \tag{26}$$

and, from $\partial U(.)\partial\theta = 0$,

$$\bar{k}\left(\tilde{g}\frac{\hat{\mu}_C}{\mu_C} - 1\right) = \left.\frac{d\mu_V}{d\sigma_V}\right|_{U(.)}. \tag{27}$$

While both conditions have to be satisfied in the case of partial coverage, in the case of an all-or-nothing offer of course only the first is relevant. After inserting (27) into (26) and carrying out some elementary manipulations we combine both conditions to

$$\bar{\mu}'_M(\sigma_M) = \theta\bar{k}\left(\tilde{g}\frac{\partial \hat{\mu}_C}{\partial \mu_C} - 1\right) + (1-\theta)\bar{k}\left(\tilde{g}\frac{\hat{\mu}_C}{\mu_C} - 1\right), \quad \theta > 0. \tag{28}$$

By construction, this condition must be satisfied in the case of partial coverage. A comparison with (26), however, reveals that, in the special case $\theta = 1$, it coincides with the condition (26) of optimal behavior in the all-or-nothing case. Hence (28) contains all the information needed to find out which loss prevention policy the insurance purchaser chooses under the kinds of contracts we consider.

Assume first $\hat{\mu}_M/\mu_M = \partial\hat{\mu}_M/\partial\mu_M = 1$ which is the case of *equivalence rating* that was considered in section 2.1. Then, from (28), we get the condition

$$\bar{\mu}'_M(\sigma_M) = \bar{k}(\tilde{g} - 1). \tag{29}$$

Referring to Figure 18 in connection with equation (18) we found above that the optimal loss prevention policy is given by the point of tangency between the upper boundary of the opportunity set $M$ and an insurance line. This result is confirmed by (29).

The question now is which change in the loss prevention policy is brought about by switching to the hypothesis $\partial\hat{\mu}_C/\partial\mu_C = 0$. The answer partially depends on whether, with $\hat{\mu}_C/\mu_C > 1$, the insurance purchaser takes less care than the company expects, whether, with $\hat{\mu}_C/\mu_C < 1$, the company is too optimistic, or whether the intermediate case $\hat{\mu}_C/\mu_C = 1$ prevails. The latter seems to be a particularly interesting case.

Let there be a large number of insurance purchasers endowed with identical preferences and identical time-invariant opportunty sets of standard risk projects $Q$, but not necessarily identical wealth. Assume that, in addition to $\theta$, the company is able to monitor the decision maker's wealth so that it knows the opportunity set $M$ from which a particular purchaser chooses. Then, with stochastic independence between the risks of the single purchasers, the company will be able to calculate from the observed sum of losses a correct estimation $\hat{\mu}_C = \mu_C$. Contrary to what might be supposed at first, this does not mean that, in (28), we have to set $\partial\hat{\mu}_C/\partial\mu_C = 1$. If the *single* insurance purchaser decides for himself, independently of others, to change $\sigma_M$, then, because his risk is negligible in the total portfolio of the company, there is no way that he can alter the company's rating system. Hence $\partial\hat{\mu}_C/\partial\mu_C = 0$ has to be maintained in (28) despite employing the *equilibrium condition* $\hat{\mu}_C = \mu_C$. Instead of (29) we therefore have

$$\bar{\mu}'_M(\sigma_M) = \bar{k}(\tilde{g} - 1) - \theta\bar{k}\tilde{g}. \tag{30}$$

A comparison between this expression and (29) reveals the outcome of *community rating*. If insurance is demanded at all, then the decision maker chooses a point on the efficiency boundary $\bar{\mu}_M(\sigma_M)$ of the

original opportunity set $M$ where this boundary has a lower slope than at the point chosen under equivalence rating. The difference in slopes is greater, the larger the degree of coverage, regardless of whether the buyer could choose this degree or whether the company dictated it.

In section C 2.1 the allocation described by (29) was identified as being Pareto optimal. This already implies that the different allocation as given by (30) must bring about a welfare loss. This loss is illustrated in Figure 20. There, two insurance lines are shown. The upper one, $\overline{GI}$, which is tangent to the original opportunity set, is chosen in the case of equivalence rating, the lower one, $\overline{G'I'}$, under community rating. Both lines have the same slope since, for the reason given above, even in the absence of direct control the company knows the size of[36] $\mu_C$. Assume, as in the discussion of the substitution effect of ideal insurance, that, independently of the insurance purchaser's choice, the company is just compensated for the burden it takes on and is therefore indifferent to the insurance buyer's action. Then, the utility loss illustrated in the figure, which the purchasers suffer when the allocation changes from $T$ to $T'$, is a clear deterioration with regard to the Pareto criterion.

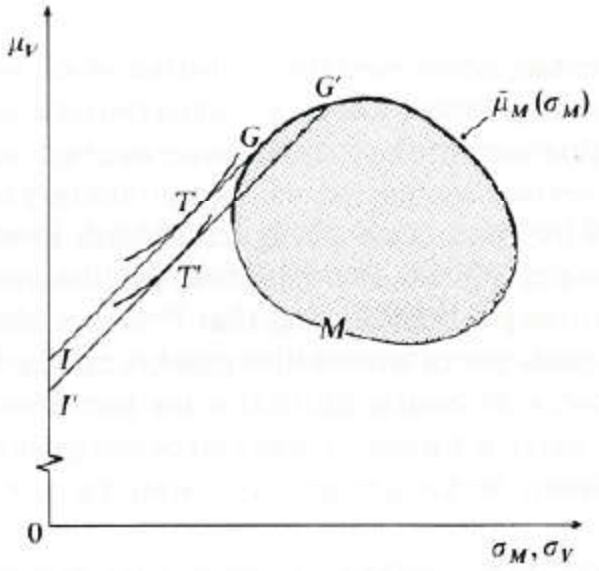


*Figure* 20

The shift from $G$ to $G'$ in Figure 20 is a shift towards a higher level of expected end-of-period wealth. To denigrate this as suboptimal may not sound very convincing. It could well be argued that, from a macroeconomic point of view, a maximization of expected end-of-period

[36] The equality of slopes follows from equation (14) but may also be calculated from (23) if $\theta$ is replaced according to (24) and account is taken of $\bar{\mu}_C = \mu_C = \bar{k}\sigma_M$.

wealth is optimal, a solution approached more by $G'$ than by $G$. This argument could be supported by the supposition that, at least theoretically, all risks could be pooled so that, for the welfare of the community, it would be best to base the choice on the mean-value criterion. The weakness of this argument, however, is that it neglects the possibly prohibitive administration costs of such a solution and that it lacks an explanation of why a choice similar to that under perfect pooling should be made if a consolidation of risks is not actually carried out[37].

Apart from this, the doubts concerning our result do not have a solid basis since an increase in the expected value $\mu_V$ is not a general feature of the movement from $G$ to $G'$. Assume, for example, that, because of the absence of administration costs, the company only requires a unitary loading factor, $\bar{g} = 1$. Then, according to (29), $G$ is at the maximum of the original opportunity set and the switch from equivalence to community rating cannot bring about an increase in $\mu_V$. As is known, in the case $\bar{g} = 1$ the insurance line is horizontal so that the optimal degree of coverage is 100%. Because of (30) this implies

$$\tilde{\mu}'_M(\sigma_M) = -\bar{k}. \tag{31}$$

Hence, indeed, $G'$ must be to the right of and below $G$. In this particular case community rating brings about a comparatively smaller expectation of end-of-period wealth than equivalence rating does.

Condition (31) defines a point on the upper bundary of $M$ where the slope is $-\bar{k}$. From the discussion of Figure 18, it is known that such a slope characterizes point $P$ where the highest auxiliary line of type $\overline{DG}$ reaches the opportunity set, indicating that $P$ brings about the highest level of normal wealth net of prevention costs, i.e., the highest level of $(a-b)q$. This aspect very clearly illustrates the fact of misallocation: if the insurance purchaser is forced to take full coverage insurance or if he chooses it himself then, under community rating, he *completely* neglects loss prevention.

Up to now, only the question of which loss prevention policy is chosen *if* insurance is bought has been examined. Whether insurance is worth-while at all for the purchaser has not been considered. A simple answer to this question does not seem to be available, but a comparison with the case of equivalence rating shows that community rating reduces the scope for gains from contracting significantly. Consider Figure 20 once more. There, the question of whether, for $\sigma$ sufficiently high, the indifference curve which is tangent at $T'$ enters the opportunity set was

---

[37] In a similar form the problem was discussed in the famous welfare-theory debate on compensation criteria by Kaldor, Scitovsky, and Samuelson.

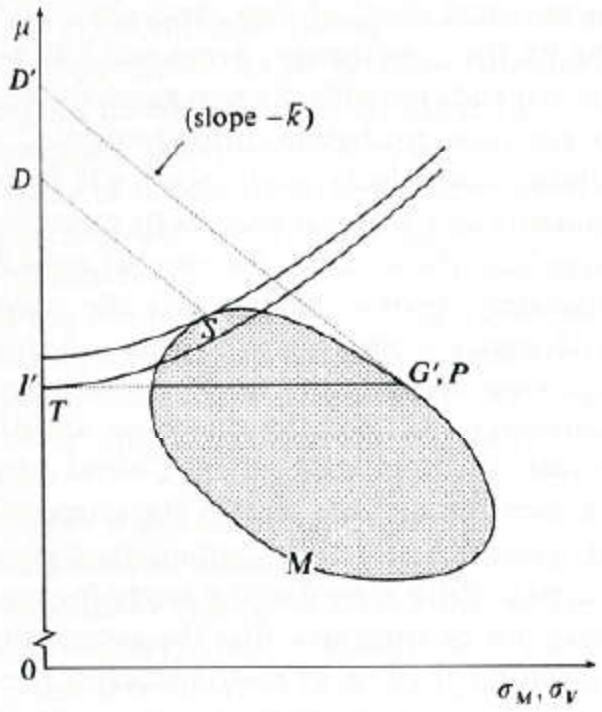


*Figure* 21

left open. If it does, insurance is not worth the price. It would be cheaper for the decision maker to 'insure' himself by choosing that distribution where the highest indifference curve is tangent to the original opportunity set.

The result sheds doubt on a proposal by ARROW (1963) that risks not covered by private companies should be covered by compulsorary government insurance[38,39]. If, as we would expect, the government is unable to control the individual's actions completely, then, even if fair premiums are required, there is no guarantee whatsoever that insurance brings about a net gain for the people as a whole.

In Figure 21 a situation, not necessarily unrealistic, that could be called *the dilemma of the welfare state,* is shown. It is assumed that the members of a community of identical individuals are obliged to buy full coverage insurance and that the government demands a premium which is just enough to cover the observed average loss. Under these assumptions, everyone chooses the distribution $P$ which is characterized by a complete absence of loss prevention efforts. So, the individual has done

[38] With reference to the moral-hazard effect caused by the cost-compensation principle, Arrow's proposal was criticized by PAULY (1968). Cf. also LEES and RICE (1965), who pointed out that government cannot avoid the administration costs that may prevent market solutions coming into existence.

[39] Arrow recognizes the general problem of moral hazard, but does not find it particularly important. A similar view is expressed by MALINVAUD (1969, p. 239).

the best from his personal point of view, but, since all losses eventually have to be borne by the community, premiums, or taxes, must be so high that each citizen ends up with the non-random wealth $\overline{OI'}$. This he clearly likes less than the probability distribution $S$, which he would have chosen without compulsory insurance[40]. Of course, it would be best for the community as a whole if each of its members were to choose a loss prevention policy which makes his original expected wealth maximal, but, unfortunately, people seem to lack the collective rationality which would be necessary if they were to behave in this way.

This pessimistic view of insurance under community rating is just a possibility. Of course it is also *possible* that there will be a net gain from insurance in the case of community rating. *Ceteris paribus,* the greater the degree of risk aversion and the smaller the scope for manipulations in end-of-period wealth distributions remaining unnoticed, the more likely this is to happen. With regard to the scope for making such manipulations, it should not be forgotten that the assumption that the company has no information at all on its customer's loss prevention policy is as extreme as the assumption that it can perfectly monitor the customer's actions. The truth will be somewhere in between, that is, the company has some, but by no means complete, information on the prevention policy that the individual chooses. The rating groups it sets up using the available information can formally be integrated into the above approach by dividing $M$ into subsets that may be, but are not necessarily, disjunct. The misallocation is then constrained to the possibility that within such a subset the 'wrong' distribution is chosen. The selection of the subset itself, which the purchaser makes when he decides on loss prevention, is a choice that is to be welcomed from an allocative point of view.

## 2.3. *The Allocation of Liability Risks*

### 2.3.1. The Incentive to Shift Risk

In the above analysis of moral hazard, misallocation mechanisms were found to prevail, which are caused by insurance. Now we consider the case where misallocation exists without insurance and is rectified when insurance is bought.

To present the problem in its simplest form, we leave the case of moral hazard and return to the equivalence-rating model of section

[40] For the sake of information, with $\overline{OD}$ and $\overline{OD'}$ the figure illustrates the levels of normal wealth net of prevention costs, $(a-b)q$, for points $S$ and $G'$. The level $\overline{OD'}$ refers to the fictitious case that $G'$ is chosen although no insurance is bought. With compulsory insurance, normal wealth falls short of $\overline{OD'}$ by the amount $\overline{I'D'}$.

C 2.1. There will be one thing different however. While above it was assumed that the opportunity set $M$ of end-of-period wealth distributions does not extend beyond the normal range of convex indifference curves, now the opposite is assumed. We consider the choice from a set of liability risks some of which are large enough to bring about negative variates of gross wealth with positive probability. Because the assumption of strong risk aversion[41] ($\varepsilon \geq 1$) is, in practice, not significant and because it would imply the trivial result that the decision maker under all circumstances tries to avoid these distributions, we confine our attention to the case of weak risk aversion. As is known, this case implies that, for $\sigma$ sufficiently large, the indifference curves are negatively sloped.

The optimal choice with indifference curves of this type is shown in Figure 22. When there is no possibility of buying insurance, a point like $S$ may be optimal. Usually this point is regarded as being inefficient for risk averse decision makers, since, to the left of it, there are other points with equal $E(V)$, but smaller $\sigma(V)$. The reason $S$ is nevertheless attractive is that, in case of damage, the BLOOS rule makes it possible to shift part of the loss on to other people[42]. An *external effect* can, therefore, be made responsible for the choice of $S$.

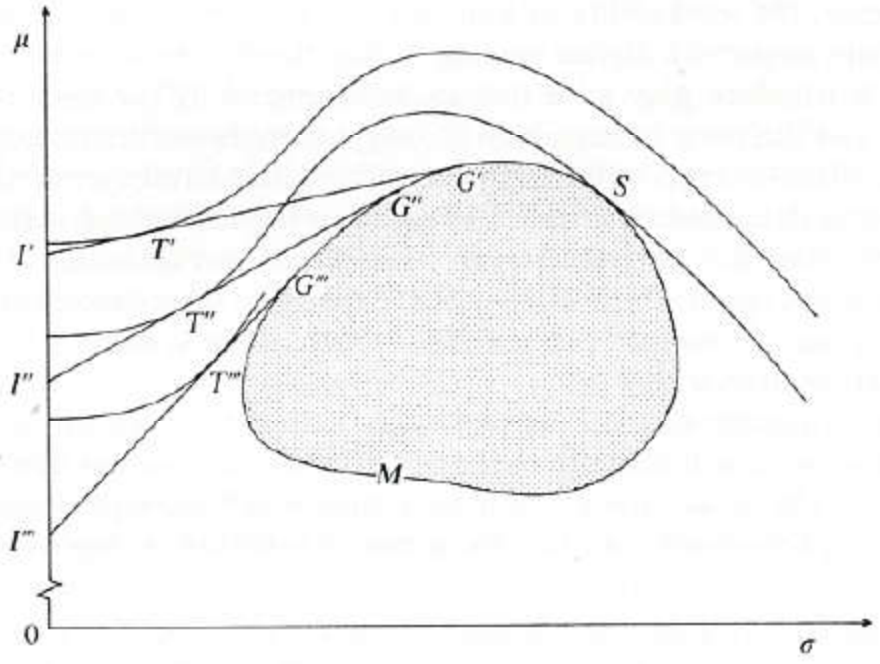


*Figure 22*

[41] Cf. section III B 2 (towards the middle), section IV B 2.3.2, and section V C 1.4.
[42] Cf. chapter III B.

Let us consider how the choice is affected if insurance is offered. Depending on the size of the loading factor $\tilde{g}$ required by the company, in the usual way an insurance line can be constructed for each point in $M$. It is possible that, for $\tilde{g}$ slightly above unity, there is an insurance line like $\overline{I'G'}$ in Figure 22 which has the property of being tangent at point $T'$ to an indifference curve above the one on which $S$ is situated. In this case, a point on the upper boundary of $M$ is chosen which is to the left of the maximum. Contrary to the previous situation, insurance now implies an increase in the amount of care devoted to loss prevention[43].

For the same reason, the misallocation that existed without insurance is reduced or removed with the purchase of insurance. The latter is the case if the point of tangency $T'$ is left of the line $E(V) = \underline{k}\sigma(V)$ which is the border between the range of those distributions extending partly over the negative half of the wealth axis and those confined to the positive half[44]. If the *whole* of the end-of-period wealth distribution is situated on the positive half of the wealth axis, then there is no external effect and hence no misallocation. Liabilities in the case of damage are paid by the insurance company, and the latter is compensated by receiving the purchaser's insurance premium.

Insurance must, therefore, be welcomed in the case of liability risks. However, the workability of a private market is not guaranteed. If the company requires a higher loading factor than has been assumed up to now, a situation may arise that is characterized by the insurance line $\overline{I''G''}$ and the point of tangency $T''$ where there is neither an advantage nor a disadvantage in buying insurance. Alternatively, even the constellation described by $\overline{I'''G'''}$ may occur, where insurance is definitely sub-optimal. But nevertheless, if, as assumed in Figure 22, the maximum of the opportunity set $M$ is above the point where the indifference curve passing through $S$ enters the ordinate there is some scope for a market solution with $\tilde{g} > 1$.

Unfortunately there is no particular reason for such an optimistic assumption. The maximum of the opportunity set is just as likely to fall short of the point where the indifference curve through $S$ enters the ordinate. Obviously, in this case a market solution is impossible. The

---

[43] Note that the indifference-curve slope is always $> -\bar{k}$. Thus, from $S$, a movement to the right along the upper boundary implies lower prevention costs and a higher level of normal wealth. A movement to the left implies the reverse.

[44] For the role of the line cf. chapter III A 2.2 and III B. The possibility of a point of tangency to the right of this line cannot, in principle, be excluded since the indifference curves may be convex in the right-hand neighborhood of the line. The case described in the text can be produced by choosing $\tilde{g}$ sufficiently close to unity. The line $E(V) = \underline{k}\sigma(V)$ was not plotted in Figure 22 so as to leave open the question of which case prevails.

insurance purchaser would only buy insurance if the loading factor $\bar{g}$ were sufficiently far below unity, but for such a value the company would not be willing to supply protection. In a free insurance market the misallocation described could not be avoided.

There are many widely discussed examples of misallocation caused by the BLOOS rule. We only have to think of the catastrophies, reported from various parts of the world, caused by selling pharmaceutical products before they have been properly tested, or of the danger arising from nuclear power plants and chemical plants. Soveso and Harrisburg are two names that can be mentioned here. There is also the case of smaller airlines that are frequently accused of not taking sufficient safety precautions. There is no reason to blame them; it simply is not worth their while spending money on preventing indemnification claims that, because liable capital is insufficient, would not have to be completely met. A similar comment can be made on the cable-car accident that happened some years ago in Northern Italy. Because the cable-car company did not spend the money necessary for a regular check of the cable, many people lost their lives. The reason for the neglect of safety again seems to have been the BLOOS rule, for, after the accident, the fact emerged that the liable capital was sufficient to cover only a very small fraction of the indemnification claims made by the relatives of the dead.

### 2.3.2. The Role of the Coase Theorem

The accusation of misallocation was based on the prevalence of external effects. In this connection we come up against the criticism of the traditional view of external effects formulated by COASE (1960).

Suppose the activities of one economic decision maker affect another decision maker adversely. Then, Coase maintains, independently of whether the one who causes the damage is liable or not, a negotiation between the two persons will produce a Pareto optimal level of the activity in question. According to the traditional view expressed by PIGOU (1932, esp. pp. 134 f., 174 f., 183–188) such a possibility does not exist. Only in the case of liability is there sufficient incentive for trying to prevent damage. The simple argument by which Coase rejects the Pigovian view is that, in the absence of liability, the person facing the possibility of being damaged will try to bribe the other to reduce or stop his activities. Thus a satisfactory allocation will be achieved independently of the liability rule[45].

[45] The Coase Theorem refers to the activities of both parties involved. In the present context we are only concerned with changes in loss prevention policy.

The validity of the Coase Theorem has frequently been questioned. Among the problems discussed is whether the theorem is valid in the case of separable cost functions. Cf. DAVIS and WHINSTON (1962), MARCHAND and RUSSEL (1973, 1975), COEHLO (1975), GIFFORD and STONE (1975), and GREENWOOD, INGENE, and HORSFIELD (1975). Moreover, the invariance with respect to changes in property rights contended by Coase must be doubted in the case where one of the parties has market power. See SINN and SCHMOLTZI (1981).

For the above examples, however, a market solution of the kind described by Coase is not available in reality. The question therefore arises of why this is so. There seem to be two reasons in particular that will account for it.

Consider first the examples referring to the danger from nuclear power plants and chemical plants, as well as the example of automobile liability risks to which reference has been made from time to time in this book. What these examples have in common is that the risk is dispersed over a great many people and cannot be divided up among them. The agent causing the risks is not obliged by law to pay compensation *ex ante,* but must compensate for any damage *ex post.* Yet, because of the BLOOS rule, he cannot.

In such a case according to the Coase Theorem, the people endangered could be expected to bribe the one causing the damage in order to induce him to change his behavior. The reason that this does not happen in reality relates to the public-goods aspect of the problem. When bribes are to be collected to induce the desired behavior change on the part of the agent causing the risk, people prefer to have their neighbors pay. This free-rider problem was seen, in principle, by COASE (1960, pp. 17 f.) and has been stressed by many others. For a great many practical allocation problems, where large liability risks are involved, it seems to be an insurmountable obstacle to a market solution.

Next consider the examples referring to insufficiently tested pharmaceutical products, the cable-car accident, and the poor safety standards of small airlines. Here, there are two significant theoretical differences from the first case. One is that the risk is separable between those who are endangered. The other is that, in addition to an *ex post* compensation in the case of damage, there is the possibility of an *ex ante* compensation: the purchaser, i.e., the endangered person, decides whether or not to accept the risk.

These two differences exclude the public-goods aspect described above. Well-informed consumers will be aware of the risk imposed upon them by the BLOOS rule and will therefore buy the respective commodities or services only if the prices are sufficiently low to compensate for this risk. There are no external effects and the market solution can be expected to prevent a choice such as the one represented by point *S* in Figure 22.

However, the assumption of well-informed consumers is particularly misleading in the uncertainty case. If the probability of damage is very low, there is almost no chance of inferring the required information from observing empirical frequencies. Moreover, although producers are normally well informed regarding the possible risks of their products, for obvious reasons they make every attempt to keep their

information secret. Thus, in the case of product liability risks, the ignorant consumer seems to be the normal case.

So the Coase Theorem appears to be a rather weak argument against the BLOOS rule. There seem to be important cases where the BLOOS rule does indeed produce an artificial incentive to choose probability distributions that involve particularly large liability risks.

### 2.3.3. The Advantage of Compulsory Insurance

Since the market mechanism often seems unable to internalize the risks imposed on other parties through the operation of the BLOOS rule, other solutions must be sought.

In those cases where the lack of information on the part of the endangered party is the reason for misallocation, a good solution might be to force producers to reveal detailed product information to the consumers. For medicine, food, and many other articles various countries have successfully chosen this way.

In the other cases where the free-rider problem is the reason market forces are unable to react appropriately to the BLOOS rule, more direct government intervention may be required.

One possibility would be the introduction of a tax-subsidy mechanism[46]. But with such a solution neither the person causing the risk nor the person endangered would be freed from risk. If both are buying market insurance to reduce their risks, three agents would be concerned with the difficult task of estimating the loss distribution: the government, the insurer of the person sustaining the damage, and the insurer of the person causing it.

If we neglect the value question of who should pay whom, another solution that has frequently been chosen in practice seems much cheaper. This solution is the introduction of compulsory insurance[47]. Compulsory insurance can solve the allocation problem just as well and just as badly as a tax-subsidy mechanism, but has the advantage that, with a single action, both parties, the one causing and the one sustaining the damage, get rid of the risk.

The allocative implications of such a compulsory insurance are illustrated in Figure 23. Analogously to Figure 22, without insurance, point $S$ is chosen. The position of this point reveals both the shape of the end-of-period wealth distribution and the optimal loss prevention policy.

---

[46] If the Kaldor criterion is accepted, it is sufficient to follow PIGOU (1932, exp. pp. 192–196 and chapter XI) and tax the person causing the damage without paying out the revenue to the one sustaining it.

[47] The allocation of liability risks thus seems to be a good case for Arrow's proposal that the government ought to insure those risks that are not underwritten by private companies. Cf. the above reasoning concerning Figure 21.

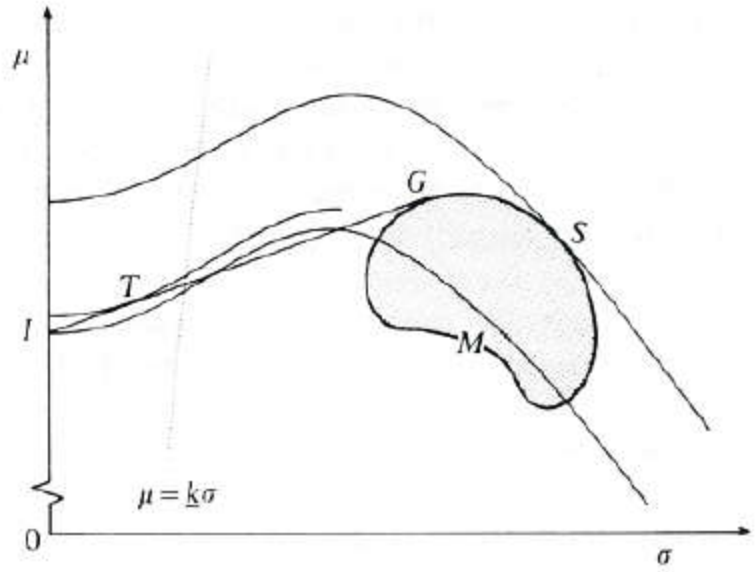


*Figure 23*

With compulsory insurance, the decision maker chooses another loss prevention policy which is described by point $G$. Under a full-coverage contract the corresponding level of end-of-period wealth is given by point $I$. Rather than requiring full coverage, the government might also allow a partial-coverage contract like that given by point $T$. In this case, however, care must be taken to ensure that the degree of coverage is sufficiently large so as to avoid a shifting of risk, i.e., the possibility of negative gross wealth. In the case of unbounded loss distributions this requirement of course leads back to the full-coverage case. For bounded loss distributions it means that only points on the insurance line to the left of the 'border line' $\mu = \underline{k}\sigma$ are admissible.

In principle, it is possible to solve the allocation problem implied by the BLOOS rule with the aid of a compulsory insurance system. It should not be forgotten, however, that the other allocative weaknesses of insurance studied above have not been overcome. Thus, in the case of large liability risks, a compromise must be made between two sources of misallocation, one that is brought about and one that is removed by insurance.

## 4. *Summary*

The analysis of insurance demand was carried out for two kinds of contracts: full-coverage contracts, where the individual has to choose between all or noting, and partial-coverage contracts, where any degree of coverage between zero and unity can be chosen. The first part of the analysis was concerned with the case of given risks that cannot be

manipulated by the insurance purchasers. It brought about the following results.

In the cases of property insurance and small-scale liability insurance there is scope for mutually advantageous contracts between the insurance purchaser and the company. The picture is different in the case of large liability risks. Since here insurance means that the insured person loses the advantage of shifting part of his risk to others, a market solution is not guaranteed. It is true, a market solution is possible in the case of strong risk aversion ($\varepsilon \geq 1$) since the purchaser has the lexicographic preference of preventing his net wealth from falling to zero. However, unfortunately only the hypothesis of weak risk aversion ($0 < \varepsilon < 1$) seems realistic. Only this hypothesis is compatible with the empirical observation that people develop a higher intensity of insurance demand as they grow older. In the case of weak risk aversion, people will not demand either full-coverage or partial-coverage contracts if the possible losses are sufficiently large relative to the decision maker's wealth.

The analysis of insurance demand for given risks can explain various phenomena observable in insurance markets. But from an economic point of view, it is particularly important to understand the allocative effects brought about by insurance. The following results were achieved in the second part of the analysis where the size of risk was considered to be subject to individual choice.

Under equivalence rating and unconditional compensation payments, property insurance leads to a favorable change in the loss prevention policy of the insurance purchaser: cheap insurance is substituted for excessive loss prevention costs. Unfortunately these ideal conditions are often violated in practice. First, in the case of community rating, the purchaser can manipulate the loss distribution without having to be afraid that the company will react with a change in the premium required. This possibility implies that, the higher the degree of coverage, the more he reduces his loss prevention effort beyond the optimum. In the exceptional case where the degree of coverage is above 100% there is even an incentive to destroy the object insured deliberately. A second allocative danger is brought about by the cost-compensation principle. This principle artificially reduces the price of repairs and recovery, thereby producing excessive demand leading to welfare losses.

While, in the case of property risks, insurance induces a reduction in loss prevention costs it may, in the case of large liability risks, have the opposite effect. Under equivalence rating this allocative effect is welcome. Without insurance, people neglect loss prevention because the Bloos rule makes it possible to transfer part of the risk to others. With insurance this external effect is removed and hence the 'right' amount of loss prevention effort is chosen. Since there is no guarantee that large

liability risks will be voluntarily insured, government intervention seems to be required. The introduction of compulsory insurance for large liability risks was seen to be an attractive possibility.