

ON THE MICROECONOMIC THEORY OF  
INVESTMENT UNDER UNCERTAINTY<sup>†</sup>

by

Robert C. Merton

WP #958-77      October 1977

<sup>†</sup> A revised version to appear  
in K. J. Arrow and M. D. Intriligator  
(eds.), Handbook of Mathematical  
Economics. Amsterdam: North-Holland  
Publishing Company (1978)

# On the Microeconomic Theory of Investment under Uncertainty\*

Robert C. Merton

## I. Introduction

Investment theory is the study of the individual behavior of households and economic organizations in the allocation of their resources to the available investment opportunities. For the purposes of investment theory, economic organizations are characterized as being members of one of two groups: "business firms" that hold as assets the physical means of production for the economy and finance their production decisions by issuing financial claims or securities; and "financial intermediaries" that hold financial claims as assets and finance these assets by issuing securities. Individuals or households are assumed to invest primarily in securities, and therefore invest only indirectly in physical assets. The markets in which these securities are traded are called the capital markets.

The natural starting point for the development of investment theory is to derive the investment behavior of individuals. It is traditional in economic theory to take the existence of households and their tastes as exogenous to the theory. However, this tradition does not extend to economic organizations and institutions. They are regarded as existing primarily because of the functions they serve instead of functioning primarily because they exist. Economic organizations are endogenous to the theory. To derive the functions of these economic organizations, therefore, the investment behavior of individuals must be derived first.

It is convenient to break the investment decision by individuals into two parts: (1) the "consumption-saving" choice where the individual decides how much of his wealth to allocate to current consumption and how much to invest for future consumption; and (2) the "portfolio selection" choice where he decides how to allocate his savings among the available investment opportunities. In general, the two decisions cannot be made independently. However, the format of the paper is to, first, solve the portfolio selection problem taking individual's consumption decisions and firms' production decisions as given, and to derive necessary conditions for financial equilibrium. Second, using these necessary conditions, the optimal production decision rules for firms are derived. Finally, the combined consumption and portfolio selection problem for individuals is solved.

If is, of course, not possible in a single paper to cover all the topics important to investment theory. However, two topics not covered here warrant special mention. First, while some necessary conditions for equilibrium are derived, I have not attempted to integrate these conditions into a general equilibrium theory for the economy. Second, no attempt has been made to make explicit how individuals and firms acquire the information needed to make their decisions, and in particular how they modify their behavior in environments where there are significant differences in the information available to various participants.<sup>1/</sup>

## II. One-Period Portfolio Selection

The basic investment choice problem for an individual is to determine the optimal allocation of his wealth among the available investment opportunities. The solution to the general problem of choosing the best investment mix is called portfolio selection theory. I begin the study of portfolio selection theory with its classic one-period formulation.

There are  $n$  different investment opportunities called securities and the random variable one-period return per dollar on security  $j$  is denoted by  $Z_j$  ( $j=1, \dots, n$ ) where a "dollar" is the "unit of account." Any linear combination of these securities which has a positive market value is called a portfolio. It is assumed that the investor chooses at the beginning of a period that feasible portfolio allocation which maximizes the expected value of a von Neumann-Morgenstern utility function<sup>2/</sup> for end-of-period wealth. I denote this utility function by  $U(W)$  where  $W$  is the end-of-period value of the investor's wealth measured in dollars. It is further assumed that  $U$  is an increasing strictly concave function on the range of feasible values for  $W$  and that  $U$  is twice-continuously differentiable.<sup>3/</sup> Because the criterion function for choice depends only on the distribution of end-of-period wealth, the only information about the securities that is relevant to the investor's decision is his subjective joint probability distribution for  $(Z_1, \dots, Z_n)$ .

In addition, it is assumed that:

- (A.1) ("Frictionless" markets) There are no transactions costs or taxes, and all securities are perfectly divisible.
- (A.2) ("Price Taker") The investor believes that his actions cannot affect the probability distribution of returns on the available securities.

Hence, if  $w_j$  is the fraction of the investor's initial wealth,  $W_0$ , allocated to security  $j$ , then  $\{w_1, \dots, w_n\}$  uniquely determines the probability distribution of his terminal wealth.

A "riskless security" is defined to be a security or feasible portfolio of securities whose return per dollar over the period is known with certainty.

(A.3) ("No-Arbitrage Opportunities") All riskless securities must have the same return per dollar. This common return will be denoted by  $R$ .

(A.4) ("No-Institutional" Restrictions) Short-sales of all securities, with full use of proceeds, is allowed without restriction. If there exists a riskless security, then the borrowing rate equals the lending rate.<sup>4/</sup>

Hence, the only restriction on the choice for the  $\{w_j\}$  is the budget constraint that  $\sum_1^n w_j = 1$ .

Given these assumptions, the portfolio selection problem can be formally stated as

$$(II.1) \quad \text{Max}_{\{w_1, \dots, w_n\}} E\{U(\sum_1^n w_j Z_j, W_0)\}$$

subject to  $\sum_1^n w_j = 1$  where "E" is the expectation operator for the subjective joint probability distribution. If  $(w_1^*, \dots, w_n^*)$  is a solution to (II.1), then it will satisfy the first-order conditions

$$(II.2) \quad E\{U'(Z^* W_0) Z_j\} = \lambda, \quad j = 1, 2, \dots, n$$

where the prime denotes derivative;  $Z^* \equiv \sum_1^n w_j^* Z_j$  is the random variable return per dollar on the optimal portfolio; and  $\lambda$  is the Lagrange multiplier

for the budget constraint. Together with the concavity assumptions on  $U$ , if the  $n \times n$  variance-covariance matrix of the returns  $(Z_1, \dots, Z_n)$  is non-singular, then there exists a unique interior solution. This condition on the returns distribution eliminates "redundant" securities (i.e., securities whose returns can be expressed as exact linear combinations of the returns on other available securities).<sup>5/</sup> It also rules out that any one of the securities is a riskless security.

If a riskless security is added to the menu of available securities (call it the  $(n+1)^{st}$  security), then it is the convention to express (II.1) as the following unconstrained maximization problem:

$$(II.3) \quad \text{Max}_{\{w_1, \dots, w_n\}} E\{U([\sum_{j=1}^n w_j (Z_j - R) + R]W_0)\}$$

where the portfolio allocations to the risky securities are unconstrained because the fraction allocated to the riskless security can always be chosen to satisfy the budget constraint. (I.e.,  $w_{n+1}^* = 1 - \sum_1^n w_j^*$ ). The first-order conditions can be written as

$$(II.4) \quad E\{U'(Z^* W_0)(Z_j - R)\} = 0 \quad j = 1, 2, \dots, n$$

where  $Z^*$  can be rewritten as  $\sum_1^n w_j^* (Z_j - R) + R$ . Again, if it is assumed that the variance-covariance matrix of the returns on the risky securities is nonsingular, then there exists a unique interior solution.

In both (II.1) or (II.3), no explicit consideration has been given for the treatment of bankruptcy (i.e.,  $Z^* < 0$ ). To rule out bankruptcy, the additional constraint that the probability of  $Z^* \geq 0$  be one could be imposed on the choices for  $(w_1^*, \dots, w_n^*)$ .<sup>6/</sup> If the reason for this constraint is to reflect institutional restrictions designed to avoid individual

bankruptcy, then it is too weak because the probability assessments on the  $\{Z_j\}$  are subjective. A more-realistic treatment would be to forbid borrowing and short-selling in conjunction with limited-liability securities where, by law,  $Z_j \geq 0$ . These rules can be formalized as restrictions on the allowable set of  $\{w_j\}$  such that  $w_j^* \geq 0$ ,  $j = 1, 2, \dots, n+1$ , and (II.1) or (II.3) can be solved using the methods of Kuhn and Tucker<sup>7/</sup> for inequality constraints. However, since the creation of limited-liability securities is itself a bankruptcy rule and only one of many that might be proposed, such explicit restrictions will not be imposed. Rather, it is simply assumed that there exists a bankruptcy law which allows for  $U(W)$  to be defined for  $W < 0$  and that this law is consistent with the continuity and concavity assumptions on  $U$ .

The optimal demand functions for risky securities,  $\{w_j^* w_0\}$ , and the resulting probability distribution for the optimal portfolio will, of course, depend on the risk preferences of the investor, his initial wealth, and the joint distribution for the securities' returns. It is well-known that the von Neumann-Morgenstern utility function can only be determined up to a positive affine transformation.<sup>8/</sup> Hence, the preference orderings of all choices available to the investor are completely specified by the Pratt-Arrow<sup>9/</sup> absolute risk-aversion function which can be written as

$$(II.5) \quad A(W) \equiv \frac{-U''(W)}{U'(W)},$$

and the change in absolute risk-aversion with respect to a change in wealth is, therefore, given by

$$(II.6) \quad \frac{dA}{dW} = A'(W) = A(W) \left[ A(W) + \frac{U'''(W)}{U''(W)} \right].$$

By the assumption that  $U(W)$  is increasing and strictly concave,  $A(W)$  is positive, and

such investors are called risk-averse. An alternative, but related, measure of risk-aversion is the relative risk-aversion function defined to be

$$(II.7) \quad \mathcal{R}(W) \equiv - \frac{U''(W)W}{U'(W)} = A(W)W,$$

and its change with respect to a change in wealth is given by

$$(II.8) \quad \mathcal{R}'(W) = A'(W)W + A(W) .$$

The certainty equivalent end-of-period wealth,  $W_c$ , associated with a given portfolio for end-of-period wealth whose random variable value is denoted by  $W$ , is defined to be that value such that

$$(II.9) \quad U(W_c) = E\{U(W)\} .$$

I.e.,  $W_c$  is the amount of money such that the investor is indifferent between having this amount of money for certain or the portfolio with random variable outcome  $W$ . The term "risk-averse" as applied to investors with concave utility functions is descriptive in the sense that the certainty equivalent end-of-period wealth is always less than the expected value of the associated portfolio,  $E\{W\}$ , for all such investors. The proof follows directly by Jensen's Inequality. Namely, if  $U$  is strictly concave, then

$$U(W_c) = E\{U(W)\} < U(E\{W\})$$

whenever  $W$  has positive dispersion, and because  $U$  is a nondecreasing function of  $W$ ,  $W_c < E\{W\}$  .

The certainty-equivalent can be used to compare the risk-aversions of two investors. An investor is said to be more risk-averse than a second investor if for every portfolio, the certainty-equivalent end-of-period wealth for the first investor is less than or equal to the certainty equivalent



end-of-period wealth associated with the same portfolio for the second investor with strict inequality holding for at least one portfolio.

While the certainty equivalent provides a natural definition for comparing risk-aversions across investors, Rothschild and Stiglitz<sup>10</sup> have attempted in a corresponding fashion to define the meaning of "increasing risk" for a security so that the "riskiness" of two securities or portfolios can be compared. In comparing two portfolios with the same expected values, the first portfolio with random variable outcome denoted by  $W_1$  is said to be less risky than the second portfolio with random variable outcome denoted by  $W_2$  if

$$(II.10) \quad E\{U(W_1)\} \geq E\{U(W_2)\}$$

for all concave  $U$  with strict inequality holding for some concave  $U$ . They bolster their argument for this definition by showing its equivalence to the following two other definitions:

(II.11) There exists a random variable  $Z$  such that  $W_2$  has the same distribution as  $W_1 + Z$  where the conditional expectation of  $Z$  given the outcome on  $W_1$  is zero. (I.e.,  $W_2$  is equal in distribution to  $W_1$  plus some "noise.")

(II.12) If the points of  $F$  and  $G$ , the distribution functions of  $W_1$  and  $W_2$ , are confined to the closed interval  $[a, b]$ , and  $T(y) \equiv \int_a^y [F(x) - G(x)]dx$ , then  $T(y) \geq 0$  and  $T(b) = 0$ . (I.e.,  $W_2$  has more "weight in its tails" than  $W_1$ .)

A feasible portfolio with return per dollar  $Z$  will be called an efficient portfolio if there exists an increasing, strictly concave function  $V$  such that  $E\{V'(Z)(Z_j - R)\} = 0$ ,  $j = 1, 2, \dots, n$ . Using the Rothschild-

Stiglitz definition of "less risky," a feasible portfolio will be an efficient portfolio only if there does not exist another feasible portfolio which is less risky than it is. All portfolios that are not efficient are called inefficient portfolios.

Proposition II.1. If the set of available securities contain no redundant securities, then no two efficient portfolios have the same risk in the Rothschild-Stiglitz sense.

Proof: Let  $Z_e^1$  and  $Z_e^2$  denote the random variable returns per dollar on two distinct efficient portfolios. Let "E" denote the expectation operator over their joint distribution. The proof goes by contradiction. If the two portfolios have the same risk, then  $E(Z_e^1) = E(Z_e^2)$  and  $E\{U(Z_e^1)\} = E\{U(Z_e^2)\}$  for every concave U. Consider the following artificial security with return per dollar Z given by  $Z = YZ_e^1 + (1 - Y)Z_e^2$  where  $Y = 1$  with probability  $\delta$  ( $0 < \delta < 1$ ) and  $Y = 0$  with probability  $(1 - \delta)$  and Y is independent of  $Z_e^1$  and  $Z_e^2$ . It follows that  $E_Z(Z) = E(Z_e^1) = E(Z_e^2)$  and  $E_Z\{U(Z)\} = \delta E\{U(Z_e^1)\} + (1 - \delta)E\{U(Z_e^2)\} = E\{U(Z_e^1)\} = E\{U(Z_e^2)\}$ . I.e., Z has the same risk as either  $Z_e^1$  or  $Z_e^2$ . Define  $Z_\delta \equiv E_Y(Z) = \delta Z_e^1 + (1 - \delta)Z_e^2$ . By the independence of Y and Jensen's Inequality,  $E\{U(Z_\delta)\} \geq E_Z\{U(Z)\}$  with equality holding if and only if  $Z = Z_\delta$  because U can be taken to be strictly concave. If the inequality holds, then  $Z_\delta$  is less risky than Z because  $E(Z_\delta) = E_Z(Z)$ . But  $Z_\delta$  is the random variable return on a feasible portfolio gotten by combining the two portfolios  $Z_e^1$  and  $Z_e^2$  with portfolio weights  $[\delta, (1 - \delta)]$ . Hence,  $Z_\delta$  is less risky than both  $Z_e^1$  and  $Z_e^2$ . But, this contradicts the hypothesis that these portfolios are efficient. Hence,  $Z = Z_\delta$ , and this is possible if and only if  $Z_e^1 = Z_e^2$ . If there are no redundant securities, then  $Z_e^1 = Z_e^2$  if and only if the

two portfolios contain identical holdings, but this contradicts the hypothesis that the two portfolios are distinct.

Hence, from the definition of an efficient portfolio and Proposition II.1, it follows that no two portfolios in the efficient set can be ordered with respect to one another. From (II.10), it follows immediately that every efficient portfolio is a possible optimal portfolio. I.e., for each efficient portfolio, there exists an increasing, concave  $U$  and an initial wealth  $W_0$  such that the efficient portfolio is a solution to (II.1) or (II.3). Further, from (II.10), all risk-averse investors will be indifferent between selecting their optimal portfolios from the set of all feasible portfolios or from the set of efficient portfolios. Hence, without loss of generality, I will assume that all optimal portfolios are efficient portfolios.

With these general definitions established, I now turn to the analysis of the optimal demand functions for risky assets and their implications for the distributional characteristics of the underlying securities. A note on notation: the symbol " $Z_e$ " will be used to denote the random variable return per dollar on an efficient portfolio, and a bar over a random variable (e.g.,  $\bar{Z}$ ) will denote the expected value of that random variable.

Theorem II.1. If  $Z$  denotes the random variable return per dollar on any feasible portfolio and if  $(Z_e - \bar{Z}_e)$  is riskier than  $(Z - \bar{Z})$  in the Rothschild and Stiglitz sense, then  $\bar{Z}_e > \bar{Z}$ .

Proof: By hypothesis,  $E\{U([Z - \bar{Z}]W_0)\} > E\{U([Z_e - \bar{Z}_e]W_0)\}$ . If  $\bar{Z} \geq \bar{Z}_e$ , then trivially,  $E\{U(ZW_0)\} > E\{U(Z_eW_0)\}$ . But,  $Z$  is a feasible portfolio and  $Z_e$  is an efficient portfolio. Hence, by contradiction,  $\bar{Z}_e > \bar{Z}$ .

Corollary II.1.a. If there exists a riskless security with return  $R$ , then  $\bar{Z}_e \geq R$  with equality holding only if  $Z_e$  is a riskless security.

Proof: The riskless security is a feasible portfolio with expected return

$R$ . If  $Z_e$  is riskless, then by assumption (A.3),  $\bar{Z}_e = R$ . If  $Z_e$  is not riskless, then  $(Z_e - \bar{Z}_e)$  is riskier than  $(R - R)$ . Therefore, by Theorem II.1,  $\bar{Z}_e > R$ .

Theorem II.2. The optimal portfolio for a nonsatiated, risk-averse investor will be the riskless security (i.e.,  $w_{n+1}^* = 1$ ,  $w_j^* = 0$ ,  $j = 1, 2, \dots, n$ ) if and only if  $\bar{Z}_j = R$  for  $j = 1, 2, \dots, n$ .

Proof: From (II.4),  $\{w_1^*, \dots, w_n^*\}$  will satisfy  $E\{U'(Z^* W_0)(Z_j - R)\} = 0$ ,  $j = 1, 2, \dots, n$ . If  $\bar{Z}_j = R$ ,  $j = 1, 2, \dots, n$ , then  $Z^* = R$  will satisfy these first-order conditions. By the strict concavity of  $U$  and the nonsingularity of the variance-covariance matrix of returns, this solution is unique. This proves the "if" part. If  $Z^* = R$  is an optimal solution, then we can rewrite (II.4) as  $U'(RW_0)E(Z_j - R) = 0$ . By the nonsatiation assumption,  $U'(RW_0) > 0$ . Therefore, for  $Z^* = R$  to be an optimal solution,  $\bar{Z}_j = R$ ,  $j = 1, 2, \dots, n$ . This proves the "only if" part.

Hence, from Corollary II.1.a and Theorem II.2, if a risk-averse investor chooses a risky portfolio, then the expected return on that portfolio exceeds the riskless rate, and a risk-averse investor will choose a risky portfolio if, at least, one available security has an expected return different from the riskless rate.

Define the notation  $E(Y|X_1, \dots, X_q)$  to mean the conditional expectation of the random variable  $Y$ , conditional on knowing the realizations for the random variables  $(X_1, \dots, X_q)$ .

Theorem II.3. If there exists a feasible portfolio with return  $Z_p$  such that for security  $s$ ,  $Z_s = Z_p + \epsilon_s$  where  $E(\epsilon_s) = E(\epsilon_s | Z_p, Z_j, j=1, \dots, n, j \neq s) = 0$ , then the fraction of every efficient portfolio allocated to security  $s$  is the same and equal to zero.

Proof: The proof follows by contradiction. Let  $Z_e$  be the return on an efficient portfolio with fraction  $\delta_s \neq 0$  allocated to security  $s$ . Let  $Z$  be the return on a portfolio with the same fractional holdings as  $Z_e$  except instead of security  $s$ , it holds the fraction  $\delta_s$  in feasible portfolio  $Z_p$ . Hence,  $Z_e = Z + \delta_s(Z_s - Z_p)$  or  $Z_e = Z + \delta_s \epsilon_s$ . By hypothesis,  $\bar{Z}_e = \bar{Z}$  and by construction,  $E(\epsilon_s | Z) = 0$ . Therefore, for  $\delta_s \neq 0$ ,  $Z_e$  is riskier than  $Z$  in the Rothschild-Stiglitz sense. But, this contradicts the hypothesis that  $Z_e$  is an efficient portfolio. Hence,  $\delta_s = 0$  for every efficient portfolio.

Corollary II.3.a. Let  $\psi$  denote the set of  $n$  securities with returns  $(Z_1, \dots, Z_{s-1}, Z_s, Z_{s+1}, \dots, Z_n)$  and  $\psi'$  denote the same set of securities except  $Z_s$  is replaced with  $Z_{s'}$ . If  $Z_{s'} = Z_s + \epsilon_s$  and  $E(\epsilon_s) = E(\epsilon_s | Z_1, \dots, Z_{s-1}, Z_s, Z_{s+1}, \dots, Z_n) = 0$ , then all risk averse investors would prefer to choose their optimal portfolios from  $\psi$  rather than  $\psi'$ .

The proof is essentially the same as the proof of Theorem II.3 with " $Z_s$ " replacing " $Z_p$ ". Unless the holdings of  $Z_s$  in every efficient portfolio is zero,  $\psi$  will be strictly preferred to  $\psi'$ .

Theorem II.3 and its corollary demonstrate that all risk averse investors would prefer any "unnecessary" uncertainty or "noise" to be eliminated. In particular, by this theorem the existence of lotteries is shown to be inconsistent with strict risk aversion on the part of all investors.<sup>11/</sup>

While the inconsistency of strict risk-aversion with observed behavior such as betting on the numbers can be "explained" by treating lotteries as consumption goods, it is difficult to use this argument to explain other implicit lotteries such as callable, sinking fund bonds where the bonds to be redeemed are selected at random.

As illustrated by the partitioning of the feasible portfolio set into its efficient and inefficient parts and the derived theorems, the Rothschild-Stiglitz definition of increasing risk is quite useful for studying the properties of optimal portfolios. However, it is important to emphasize that these theorems apply only to efficient portfolios and not to individual securities or inefficient portfolios. For example, if  $(Z_j - \bar{Z}_j)$  is riskier than  $(Z - \bar{Z})$  in the Rothschild-Stiglitz sense and if security  $j$  is held in positive amounts in an efficient or optimal portfolio (i.e.,  $w_j^* > 0$ ), then it does not follow that  $\bar{Z}_j$  must equal or exceed  $\bar{Z}$ . In particular, if  $w_j^* > 0$ , it does not follow that  $\bar{Z}_j$  must equal or exceed  $R$ . Hence, to know that one security is riskier than a second security using the Rothschild-Stiglitz definition of increasing risk provides no normative restrictions on holdings of either security in an efficient portfolio. And because this definition of riskier imposes no restrictions on the optimal demands, it cannot be used to derive properties of individual securities' return distributions from observing their relative holdings in an efficient portfolio. To derive these properties, a second definition of risk is required. However, discussion of this measure is delayed until Section III.

In closing this section, comparative statics results are derived for the optimal risky security demand in the special case of a single risky security and a riskless security (i.e.,  $n=1$ ). Additional results for the general case of many risky securities can be found in Fisher [22].

Define  $\kappa$  by

$$(II.13) \quad \kappa \equiv -E\{U''(W^*)(Z - R)^2 W_0\} > 0$$

where  $Z$  is the random variable return per dollar on the single risky security and  $W^* \equiv [w^*(Z - R) + R]W_0$  is the random variable end-of-period wealth for the optimal allocation.

Applying the Implicit Function Theorem to (II.4), the change in the proportion of the optimal portfolio allocated to the risky security with respect to a change in initial wealth can be written as

$$(II.14) \quad \frac{\partial w^*}{\partial W_0} = \frac{-E\{w^* R(W^*) U'(W^*) (W^* - RW_0)\}}{(w^* W_0)^2 \kappa} \quad \text{for } w^* \neq 0$$

$$= 0 \quad \text{for } w^* = 0,$$

where  $R$  is the relative risk-aversion function defined in (II.7). If  $U$  is such that  $R$  is a constant, then  $\partial w^* / \partial W_0 = 0$  because  $E\{U'(W^*) (W^* - RW_0)\} = 0$  from (II.4). By integrating (II.7) twice, the general class of concave utility functions that exhibit constant relative risk aversion can be written as

$$(II.15) \quad U(W) = C_1 \left[ \frac{W^\gamma - 1}{\gamma} \right] + C_2$$

where  $C_1$  and  $C_2$  are any constants such that  $C_1 > 0$  and  $\gamma < 1$ . Because von Neumann-Morgenstern utility functions are unique only up to a positive affine transformation, all distinct members of this class are determined by the single parameter  $\gamma$ . The relative risk-aversion function,  $R$ , equals  $(1 - \gamma) > 0$ . Hence, for investors with utility functions in this class, a larger value of  $\gamma$  implies a smaller value of relative risk aversion. Moreover, for all portfolios that can be ranked,<sup>12/</sup> it is straightforward to show that

$$\partial w^* / \partial \gamma > 0.$$

For each possible outcome for  $W^*$ , we have by the Mean Value Theorem that there exists a number  $\theta = \theta(W^*)$ ,  $0 \leq \theta \leq 1$ , such that

$$(II.16) \quad R(W^*) = R(RW_0) + R'(\eta)(W^* - RW_0)$$

where  $\eta \equiv RW_0 + \theta(W^* - RW_0)$  and  $R'$  is defined in (II.8). Substituting from (II.16) into (II.14) with  $w^* \neq 0$ , (II.14) can be rewritten as

$$(II.17) \quad \frac{\partial w^*}{\partial W_0} = \frac{-E\{w^* R'(\eta) U'(W^*) (W^* - RW_0)^2\}}{(w^* W_0)^2 \kappa}.$$

From (II.17), a sufficient condition for the optimal proportion in the risky security to decline with an increase in initial wealth is that  $R'w^* > 0$ .

Similarly, a sufficient condition for the optimal proportion to increase is that  $R'w^* < 0$ . Hence, for investors with strictly increasing (decreasing) relative risk aversion, an increase in initial wealth will induce a decrease (increase) in the absolute proportion of the optimal portfolio allocated to the risky security,  $|w^*|$ .

From II.14, the change in the dollar allocation to the risky security in the optimal portfolio with respect to a change in initial wealth can be written as

$$(II.18) \quad \frac{\partial (w^* W_0)}{\partial W_0} = \frac{-RE\{w^* A(W^*) U'(W^*) (W^* - RW_0)\}}{(w^*)^2 \kappa}, \quad w^* \neq 0$$

$$= 0, \quad w^* = 0$$

where  $A$  is the absolute risk-aversion function defined in (II.5). If  $U$  is such that  $A$  is a constant, then  $\partial (w^* W_0) / \partial W_0 = 0$ . By integrating (II.5) twice, the general class of concave utility functions that exhibit constant absolute



risk-aversion can be written as

$$(II.19) \quad U(W) = C_2 - C_1 e^{-\mu W}$$

where  $C_1$  and  $C_2$  are any constants such that  $C_1 > 0$  and  $\mu > 0$ . All distinct members of this class are determined by the single parameter  $\mu$  which is equal to the absolute risk-aversion function, and  $\partial(w^* W_0)/\partial\mu < 0$ .

Again, using the Mean Value Theorem,  $A(W^*)$  can be written as

$$(II.20) \quad A(W^*) = A(RW_0) + A'(\eta)(W^* - RW_0).$$

Substituting from (II.20) into (II.18) with  $w^* \neq 0$ , (II.18) can be rewritten as

$$(II.21) \quad \frac{\partial(w^* W_0)}{\partial W_0} = \frac{-RE\{w^* A'(\eta)U'(W^*)(W^* - RW_0)^2\}}{(w^*)^2 \kappa}.$$

From (II.21), a sufficient condition for the optimal dollar investment in the risky security to decline is that  $A'w^* > 0$ , and a sufficient condition for it to increase is that  $A'w^* < 0$ . Hence, for investors with strictly increasing (decreasing) absolute risk aversion, an increase in initial wealth will induce a decrease (increase) in the absolute dollar position in the risky security,  $|w^* W_0|$ .

Although the direction of change in both the proportional and absolute dollar holdings of the risky security depends upon the sign of  $w^*$ , the sign of  $w^*$  is determined solely by the sign of  $(\bar{Z} - R)$ . From Theorem II.2,  $w^* = 0$  if and only if  $\bar{Z} = R$ . If  $w^* \neq 0$ , then  $\bar{Z}^* > R$  because an optimal portfolio is an efficient portfolio. But  $\bar{Z}^* = R + w^*(\bar{Z} - R)$ . Hence,  $w^*(\bar{Z} - R) > 0$  if  $w^* \neq 0$ . Therefore, for any nonsatiated, risk-averse investor, the sign of  $w^*$  will equal the sign of  $(\bar{Z} - R)$ . Warning: this condition need not obtain with respect to an individual security when there is more than one risky

security. I.e., the sign of  $w_j^*$  need not equal the sign of  $(\bar{Z}_j - R)$  for  $n > 1$ .

From (II.14), the change in  $w^*$  with respect to a change in the risky security's expected return,  $\bar{Z}$ , can be written as

$$(II.22) \quad \frac{\partial w^*}{\partial \bar{Z}} = \frac{E\{U'(W^*)[1 - A(W^*)(W^* - RW_0)]\}}{\kappa}.$$

Inspection of (II.22) shows that the sign of  $\partial w^* / \partial \bar{Z}$  is ambiguous. However, three sufficient conditions for  $\partial w^* / \partial \bar{Z}$  to be positive are:

$$(II.23) \quad (i) \quad w^* = 0;$$

or

$$(ii) \quad A'(W) \leq 0;$$

or

$$(iii) \quad R(W) \leq \frac{W}{W - RW_0} \quad \text{for } RW_0 < W \leq W^+$$

where  $W^+$  equals the maximum possible outcome for  $W^*$ . Conditions (i) and (iii) follow directly from (II.22) and (ii) follows from the substitution for  $A(W^*)$  from (II.20) into (II.22).

From (II.14), the change in  $w^*$  with respect to a change in  $R$  can be written as

$$(II.24) \quad \frac{\partial w^*}{\partial R} = \frac{-E\{U'(W^*)[(w^*)^2 + A(W^*)w^*(1 - w^*)(W^* - RW_0)]\}}{(w^*)^2 \kappa}.$$

From (II.24),  $\partial w^* / \partial R$  can be either positive or negative. Two sufficient conditions for  $\partial w^* / \partial R$  to be negative are:

$$(II.25) \quad (i) \quad A'(W) \geq 0 \quad \text{and} \quad w^*(1 - w^*) \geq 0$$

or

$$(ii) \quad A'(W) \leq 0 \quad \text{and} \quad w^*(1 - w^*) \leq 0.$$

These conditions follow from the substitution for  $A(W^*)$  from (II.20) into (II.24).

Rothschild and Stiglitz<sup>13/</sup> have examined the effect on the optimal demand for the risky security of an increase in that security's riskiness. As with the other comparative statics results derived here, they found that the change in the optimal demand can be either positive or negative. A sufficient condition for an increase in risk to decrease the demand for the risky security is that  $f(Z) \equiv (Z - R)U'([w^*(Z - R) + R]W_0)$  be concave in  $Z$ . It is straightforward to show by differentiating  $f(Z)$  twice with respect to  $Z$  that if relative risk aversion is less than or equal to  $R$  and nondecreasing and if absolute risk aversion is nonincreasing, then  $f(Z)$  is strictly concave for  $w^* > 0$  and strictly convex for  $w^* < 0$ .

In summary, even for the simplest case of one risky security and a riskless security, the sign of the change in the optimal demand for the risky security with respect to changes in its probability distribution is consistently ambiguous unless restrictions are placed on the class of utility functions or on the class of probability distributions for  $Z$ .

$\text{cov}[V'(Z_e), Z_e] < 0.$ <sup>14/</sup> It is understood that in the following discussion "efficient portfolio" will mean "efficient portfolio with positive dispersion." Let  $Z_p$  denote the random variable return per dollar on any feasible portfolio p.

Definition. The measure of risk of portfolio p relative to efficient portfolio K with random variable return  $Z_e^K$ ,  $b_p^K$ , is defined by

$$b_p^K \equiv \text{cov}[Y(Z_e^K), Z_p]$$

and portfolio p is said to be riskier than portfolio p' relative to efficient portfolio K if  $b_p^K > b_{p'}^K$ .

Theorem III.1. If  $Z_p$  is the return on a feasible portfolio p and  $Z_e^K$  is the return on efficient portfolio K, then  $\bar{Z}_p - R = b_p^K (\bar{Z}_e^K - R)$ .

Proof: From the definition of  $V(Z_e^K)$ ,  $E\{V'(Z_e^K)(Z_j - R)\} = 0$ ,  $j = 1, 2, \dots, n$ .

Let  $\delta_j$  be the fraction of portfolio p allocated to security j. Then,

$$Z_p = \sum_1^n \delta_j (Z_j - R) + R, \text{ and } \sum \delta_j E\{V'(Z_e^K)(Z_j - R)\} = E\{V'(Z_e^K)(Z_p - R)\} = 0.$$

By a similar argument,  $E\{V'(Z_e^K)(Z_e^K - R)\} = 0$ . Hence,  $\text{cov}[V'(Z_e^K), Z_e^K] = (R - \bar{Z}_e^K)E\{V'(Z_e^K)\}$  and  $\text{cov}[V'(Z_e^K), Z_p] = (R - \bar{Z}_p)E\{V'(Z_e^K)\}$ . By

Corollary II.1.a,  $\bar{Z}_e^K > R$ . Therefore,  $\text{cov}[Y(Z_e^K), Z_p] = (R - \bar{Z}_p)/(R - \bar{Z}_e^K)$ .

Hence, the expected excess return on portfolio p,  $\bar{Z}_p - R$ , is in direct proportion to its risk, and because  $\bar{Z}_e^K > R$ , the larger is its risk, the larger is its expected return. Thus, Theorem III.1 provides the first argument why  $b_p^K$  is a natural measure of risk for individual securities.

A second argument goes as follows: consider an investor with utility function U and initial wealth  $W_0$  who solves the portfolio selection problem

$$\text{Max}_w E\{U([wZ_j + (1 - w)Z]W_0)\}$$

### III. Risk Measures for Securities and Portfolios in the One-Period Model

In the previous section, the Rothschild-Stiglitz measure for the risk of a security was defined, and the comparative statics of an increase in a security's risk was analyzed in the special case of a single risky security and a riskless security. However, in the more-than-one risky security portfolio problem, the Rothschild-Stiglitz measure is not a natural definition of risk for a security. In this section, a second definition of increasing risk is introduced, and it is argued that this second measure is a more appropriate definition for the risk of a security. Although this second measure will not in general provide the same orderings as the Rothschild-Stiglitz measure, it is further argued that the two measures are not in conflict, and indeed, are complementary.

If  $Z_e$  is the random variable return per dollar on an efficient portfolio with positive dispersion, then let  $V(Z_e)$  denote an increasing, strictly concave function such that

$$E\{V'(Z_e)(Z_j - R)\} = 0, \quad j = 1, 2, \dots, n.$$

I.e.,  $V$  is a concave utility function such that an investor with initial wealth  $W_0 = 1$  and these preferences would select this efficient portfolio as his optimal portfolio. While such a function  $V$  will always exist, it will not be unique. If  $\text{cov}(x_1, x_2)$  is the functional notation for the covariance between the random variables  $x_1$  and  $x_2$ , then define the random variable,  $Y(Z_e)$ , by

$$(III.1) \quad Y(Z_e) \equiv \frac{V'(Z_e) - E\{V'(Z_e)\}}{\text{Cov}[V'(Z_e), Z_e]}$$

$Y(Z_e)$  is well-defined as long as  $Z_e$  has positive dispersion because

ordering while the  $b_j^K$ -measure provides a complete ordering. Moreover, they can give different rankings. For example, suppose the return on security  $j$  is independent of the return on efficient portfolio  $K$ , then  $b_j^K = 0$  and  $\bar{Z}_j = R$ . Trivially,  $b_R^K = 0$  for the riskless security. Therefore, by the  $b_j^K$ -measure, security  $j$  and the riskless security have equal risk. However, if security  $j$  is not riskless, then by the Rothschild-Stiglitz measure, security  $j$  is more risky than the riskless security. Despite this, the two measures are not in conflict, and indeed, are complementary. The Rothschild-Stiglitz definition measures the "total risk" of a security in the sense that it compares the expected utility from holding a security alone with the expected utility from holding another security alone. Hence, it is the appropriate definition for identifying optimal portfolios and determining the efficient portfolio set. However, it is not useful for defining the risk of securities generally because it does not take into account that investors can mix securities together to form portfolios. The  $b_j^K$ -measure does take this into account because it measures the only part of an individual security's risk which is relevant to an investor: namely, the part that contributes to the total risk of his optimal portfolio. In contrast to the Rothschild-Stiglitz measure of total risk, the  $b_j^K$  measures the "systematic risk" of a security (relative to efficient portfolio  $K$ ). Of course, to determine the  $b_j^K$ , the efficient portfolio set must be determined. Because the Rothschild-Stiglitz measure does just that, the two measures are complementary.

Other properties of the  $b_p^K$  measure of systematic risk are:

- (P.1) If a feasible portfolio  $p$  has portfolio weights  $(\delta_1, \delta_2, \dots, \delta_n)$ , then  $b_p^K = \sum_1^n \delta_j b_j^K$ . The systematic risk of a portfolio is the weighted sum of the systematic risks of its component securities.

where  $Z$  is the return on a portfolio of securities and  $Z_j$  is the return on the security  $j$ . The optimal mix,  $w^*$ , will satisfy the first-order condition

$$(III.2) \quad E\{U'([w^* Z_j + (1 - w^*)Z]W_0)(Z_j - Z)\} = 0.$$

If the original portfolio of securities chosen was this investor's optimal portfolio (i.e.,  $Z = Z^*$ ), then the solution to (III.2) is  $w^* = 0$ . However, an optimal portfolio is an efficient portfolio. Therefore, by Theorem III.1,  $\bar{Z}_j - R = b_j^*(\bar{Z}^* - R)$ . Hence, the "risk-return tradeoff" provided in Theorem III.1 is a condition for personal portfolio equilibrium. Indeed, because security  $j$  may be contained in the optimal portfolio,  $w^* W_0$  is similar to an excess demand function.  $b_j^*$  measures the contribution of security  $j$  to the Rothschild-Stiglitz risk of the optimal portfolio in the sense that the investor is just indifferent to a marginal change in the holdings of security  $j$  provided that  $\bar{Z}_j - R = b_j^*(\bar{Z}^* - R)$ . Moreover, by the Implicit Function Theorem, we have from (III.2) that

$$(III.3) \quad \frac{\partial w^*}{\partial \bar{Z}_j} = \frac{w^* W_0 E\{U''(Z - Z_j)\} - E\{U'\}}{E\{U''(Z - Z_j)^2\}}$$

$$> 0 \text{ at } w^* = 0.$$

Therefore, if  $\bar{Z}_j$  lies above the "risk-return" line in the  $(\bar{Z} - b^*)$  plane, then the investor would prefer to increase his holdings in security  $j$  and if  $\bar{Z}_j$  lies below the line, then he would prefer to reduce his holdings. If the risk of a security increases, then the risk-averse investor must be "compensated" by a corresponding increase in that security's expected return if his current holdings are to remain unchanged.

The Rothschild-Stiglitz measure of risk is clearly different from the  $b_j^K$ -measure here. The Rothschild-Stiglitz measure provides only for a partial

(P.1) follows directly from the linearity of the covariance operator,  $\text{Cov}[x_1, x_2]$ , with respect to the random variable  $x_2$ .

Let  $p$  and  $p'$  denote any two feasible portfolios and let  $K$  and  $L$  denote any two efficient portfolios.

(P.2) If  $b_p^K = 0$  for some efficient portfolio  $K$ , then  $b_p^L = 0$  for any efficient portfolio  $L$ .

(P.2) follows from Theorem III.1 and Corollary II.1.a. If  $b_p^K = 0$ , then  $\bar{Z}_p = R$  because  $\bar{Z}_e^K > R$ . But if  $\bar{Z}_p = R$  and  $\bar{Z}_e^L > R$ , then  $b_p^L = 0$ .

(P.3) 
$$b_p^K \begin{matrix} \geq \\ < \end{matrix} b_{p'}^K, \quad \text{if and only if} \quad b_p^L \begin{matrix} \geq \\ < \end{matrix} b_{p'}^L.$$

(P.3) follows from (P.2) if  $b_p^K = b_{p'}^K = 0$ . Suppose  $b_p^K \neq 0$ , then (P.3) follows from Theorem (III.1) because

$$\frac{b_{p'}^K}{b_p^K} = \frac{\bar{Z}_{p'} - R}{\bar{Z}_p - R} = \frac{b_{p'}^L}{b_p^L}.$$

(P.3) provides a third argument why  $b_p^K$  is a natural measure of risk for individual securities. Namely, the ordering of securities by their systematic risk relative to a given efficient portfolio will be identical to their ordering relative to any other efficient portfolio. Hence, given the set of available securities, there is an unambiguous meaning to the statement "security  $j$  is riskier than security  $i$ ."

(P.4) If portfolio  $p$  is an efficient portfolio (call it the  $K^{\text{th}}$  one with  $\bar{Z}_p = \bar{Z}_e^K$ ), then for any efficient portfolio  $L$ ,  $b_p^L > 0$  and in particular,  $b_p^K = 1$ . Hence, all efficient portfolios have positive systematic risk relative to any efficient portfolio.



(P.4) follows from Theorem (III.1) and Corollary II.1.a.

(P.5) If the systematic risk of portfolio  $p$  is defined by its expected return,  $\bar{Z}_p$ , and if portfolio  $p$  is said to be riskier than portfolio  $p'$  if and only if  $\bar{Z}_p > \bar{Z}_{p'}$ , then this measure of systematic risk is equivalent to the  $b_p^K$  measure.

(P.5) follows directly from Theorem (III.1), Corollary II.1.a, and (P.3).

Although the expected return of a security provides an equivalent ranking to its  $b_p^K$  measure, the  $b_p^K$  measure is not vacuous. There exist nontrivial information sets which allow  $b_p^K$  to be determined without knowledge of  $\bar{Z}_p$ . For example, consider a model in which all investors agree on the joint distribution of the returns on securities. Suppose one knows the probability distribution of the optimal portfolio,  $Z^*W_0$ , and the utility function  $U$  for some investor. From (III.1), one, therefore, knows the distribution of  $Y(Z^*)$ . For security  $j$ , define the random variable  $\epsilon_j \equiv Z_j - \bar{Z}_j$ . Suppose further that one has enough information about the joint distribution of  $Y(Z^*)$  and  $\epsilon_j$  to compute  $\text{Cov}[Y(Z^*), \epsilon_j]$ , but does not know  $\bar{Z}_j$ .<sup>15/</sup> By the definition of covariance,  $\text{Cov}[Y(Z^*), \epsilon_j] = \text{Cov}[Y(Z^*), Z_j] = b_j^*$ . However, Theorem III.1 is a necessary condition for equilibrium in the securities market. Hence, one can deduce the equilibrium expected return on security  $j$  from  $\bar{Z}_j = R + b_j^*(\bar{Z}^* - R)$ . An analysis of the necessary information sets required to deduce the equilibrium structure of security returns is an important part of portfolio theory, and further discussion of this topic is provided in Sections IV and V.

#### IV. Spanning, Separation and Mutual Fund Theorems

Definition A set of  $M$  feasible portfolios with random variable returns  $(X_1, \dots, X_M)$  are said to span the space of portfolios contained in set  $\Psi$  if and only if for any portfolio in  $\Psi$  with return denoted by  $Z_p$ , there exists numbers  $(\delta_1, \dots, \delta_M)$ ,  $\sum_{j=1}^M \delta_j = 1$ , such that  $Z_p = \sum_{j=1}^M \delta_j X_j$ . If  $N$  is the number of securities available to generate the portfolios in  $\Psi$  and if  $M^*$  denotes the smallest number of feasible portfolios that span the space of portfolios contained in  $\Psi$ , then  $M^* \leq N$ .

As was illustrated in section II, very little can be derived about the structure of optimal portfolio demand functions unless further restrictions are imposed on the class of investors' utility functions or the class of probability distributions for securities' returns. A particularly fruitful set of such restrictions is the one that provides for a nontrivial (i.e.,  $M^* < N$ ) spanning of the feasible portfolio set. Indeed, the spanning property leads to a collection of "mutual fund" or "separation" theorems that are the core of modern financial theory.

A mutual fund is a financial intermediary that holds as its assets a portfolio of securities and issues as liabilities shares against these assets. Unlike the portfolio of an individual investor, the portfolio of securities held by a mutual fund need not be an efficient portfolio. The connection between mutual funds and the spanning property can be seen in the following theorem:

Theorem IV.1 If there exists  $M$  mutual funds whose portfolios span the portfolio set  $\Psi$ , then all investors will be indifferent between selecting their optimal portfolios from  $\Psi$  or from portfolio combinations of just the  $M$  mutual funds.

The proof of the theorem follows directly from the definition of spanning. If  $Z^*$  denotes the return on an optimal portfolio selected from  $\Psi$  and if  $X_j$  denotes the return on the  $j^{\text{th}}$  mutual fund's portfolio, then there exist portfolio weights  $(\delta_1^*, \dots, \delta_M^*)$  such that  $Z^* = \sum_{j=1}^M \delta_j^* X_j$ . Hence, any investor would be indifferent between the portfolio with return  $Z^*$  and the  $(\delta_1^*, \dots, \delta_M^*)$  combination of the mutual fund shares.

Although the theorem states "indifference," if there are information-gathering or other transactions costs and if there are economies of scale, then investors would prefer the mutual funds whenever  $M < N$ . By a similar argument, one would expect that investors would prefer to have the smallest number of funds necessary to span  $\Psi$ . Therefore, the smallest number of such funds,  $M^*$ , is a particularly important spanning set. Hence, the spanning property can be used to derive an endogenous theory for the existence of financial intermediaries with the functional characteristics of a mutual fund. Moreover, from these functional characteristics, a theory for their optimal management can be derived.

For the mutual fund theorems to have serious empirical content, the minimum number of funds required for spanning,  $M^*$ , must be significantly smaller than the number of available securities,  $N$ . When such spanning obtains, the investor's portfolio selection problem can be separated into two steps: first, individual securities are mixed together to form the  $M^*$  mutual funds; second, the investor allocates his wealth among the  $M^*$  funds' shares. If the investor knows that the funds span the space of optimal portfolios, then he need only know the joint distribution of  $(X_1, \dots, X_{M^*})$  to determine his optimal portfolio. It is for this reason that the mutual fund theorems are also called "separation" theorems. However, if the  $M^*$

funds can be constructed only if the fund managers know the preferences, endowments, and probability beliefs of each investor, then the formal separation property will have little operational significance.

In addition to providing an endogenous theory for mutual funds, the existence of a nontrivial spanning set can be used to deduce equilibrium properties of individual securities' returns and to derive optimal decision rules for business firms making physical investments. Moreover, in virtually every model of portfolio selection in which empirical implications beyond those presented in sections II and III are derived, some nontrivial form of the spanning property obtains.

While the determination of conditions under which nontrivial spanning will obtain is in a broad sense a subset of the traditional economic theory of aggregation, the first rigorous contributions in portfolio theory were made by Arrow [2], Markowitz [49], and Tobin [89]. In each of these papers and most subsequent papers, the spanning property is derived as an implication of the specific model examined, and therefore, such derivations provide only sufficient conditions. In two notable exceptions, Cass and Stiglitz [7] and Ross [72] "reverse" the process by deriving necessary conditions for nontrivial spanning to obtain. In this section, necessary and sufficient conditions for spanning are developed along the lines of Cass and Stiglitz and Ross, leaving until Section V discussion of the specific models of Arrow, Markowitz, and Tobin.

Let  $\Psi^f$  denote the set of all feasible portfolios that can be constructed from a riskless security with return  $R$  and  $n$  risky securities with a given joint probability distribution for their random variable returns  $(Z_1, \dots, Z_n)$ . Let  $\Omega$  denote the  $n \times n$  variance-covariance matrix of the returns on the  $n$  risky assets.

Theorem IV.2 Necessary conditions for the  $M$  feasible portfolios with returns  $(X_1, \dots, X_M)$  to span the portfolio set  $\Psi^f$  are (i) that the rank of  $\Omega \leq M$  and (ii) that there exists numbers  $(\delta_1, \dots, \delta_M)$ ,  $\sum_1^M \delta_j = 1$  such that the random variable  $\sum_1^M \delta_j X_j$  has zero variance.

Proof:

(i) the set of portfolio  $\Psi^f$  defines a  $(n+1)$  dimensional vector space. By definition, if  $(X_1, \dots, X_M)$  spans  $\Psi^f$ , then each risky security's return can be represented as a linear combination of  $(X_1, \dots, X_M)$ . Clearly, this is only possible if the rank of  $\Omega \leq M$ . (ii) The riskless security is contained in  $\Psi^f$ . Therefore, if  $(X_1, \dots, X_M)$  spans  $\Psi^f$ , then there must exist a portfolio combination of  $(X_1, \dots, X_M)$  which is riskless.

Proposition IV.1 If  $Z_p = \sum_1^n a_j Z_j + b$  is the return on some security or portfolio and if there are no "arbitrage opportunities" (Assumption (A.3)), then (1)  $b = [1 - \sum_1^n a_j]R$  and (2)  $Z_p = R + \sum_1^n a_j (Z_j - R)$ .

Proof:

Let  $Z^+$  be the return on a portfolio with fraction  $\delta_j^+$  allocated to security  $j$ ,  $j=1,2,\dots,n$ ;  $\delta_p$  allocated to the security with return  $Z_p$ ;  $(1-\delta_p-\sum_1^n \delta_j^+)$  allocated to the riskless security with return  $R$ . If  $\delta_j^+$  is chosen such that  $\delta_j^+ = -\delta_p a_j$ , then  $Z^+ = R + \delta_p (b - R[1 - \sum_1^n a_j])$ .  $Z^+$  is a riskless security, and therefore, by (A.3),  $Z^+ = R$ . But,  $\delta_p$  can be chosen arbitrarily. Therefore,  $b = [1 - \sum_1^n a_j]R$ . Substituting for  $b$ , it follows directly that  $Z_p = R + \sum_1^n a_j (Z_j - R)$ .

As long as there are no arbitrage opportunities, from Theorem IV.2 and Proposition IV.1, without loss of generality, it can be assumed that one

of the portfolios in any candidate spanning set is the riskless security.

If by convention,  $X_M = R$ , then in all subsequent analyses, the notation  $(X_1, \dots, X_m, R)$  will be used to denote a  $M$ -portfolio spanning set where  $m \equiv M-1$  is the number of risky portfolios (together with the riskless security) that span  $\Psi^f$ .

Theorem IV.3 A necessary and sufficient condition for  $(X_1, \dots, X_m, R)$  to span  $\Psi^f$  is that there exist numbers  $(a_{ij})$  such that  $Z_j = R + \sum_{i=1}^m a_{ij}(X_i - R)$ ,  $j=1, 2, \dots, n$ .

Proof:

If  $(X_1, \dots, X_m, R)$  span  $\Psi^f$ , then there exists portfolio weights  $(\delta_{1j}, \dots, \delta_{mj})$ ,  $\sum_{i=1}^m \delta_{ij} = 1$ , such that  $Z_j = \sum_{i=1}^m \delta_{ij} X_i$ . Noting that  $X_M = R$  and substituting  $\delta_{Mj} = 1 - \sum_{i=1}^m \delta_{ij}$ , we have that  $Z_j = R + \sum_{i=1}^m \delta_{ij}(X_i - R)$ . This proves necessity. If there exists numbers  $(a_{ij})$  such that  $Z_j = R + \sum_{i=1}^m a_{ij}(X_i - R)$ , then pick the portfolio weights  $\delta_{ij} = a_{ij}$  for  $i=1, \dots, m$ , and  $\delta_{Mj} = 1 - \sum_{i=1}^m \delta_{ij}$ , from which it follows that  $Z_j = \sum_{i=1}^M \delta_{ij} X_i$ . But every portfolio in  $\Psi^f$  can be written as a portfolio combination of  $(Z_1, \dots, Z_n)$  and  $R$ . Hence,  $(X_1, \dots, X_m, R)$  spans  $\Psi^f$  and this proves sufficiency.

Let  $\Omega_X$  be the  $m \times m$  variance-covariance matrix of the returns on the  $m$  portfolios with returns  $(X_1, \dots, X_m)$

Corollary IV.3a A necessary and sufficient condition for  $(X_1, \dots, X_m, R)$  to be the smallest number of feasible portfolios that span  $\Psi^f$  (i.e.,  $M^* = m+1$ ) is that the rank of  $\Omega$  equals the rank of  $\Omega_X = m$ .

Proof:

If  $(X_1, \dots, X_m, R)$  span  $\Psi^f$  and  $m$  is the smallest number of risky portfolios that does, then  $(X_1, \dots, X_m)$  must be linearly independent, and

therefore,  $\text{rank } \Omega_X = m$ . Hence,  $(X_1, \dots, X_m)$  form a basis for the vector space of security returns  $(Z_1, \dots, Z_n)$ . Therefore, the rank of  $\Omega$  must equal  $\Omega_X$ . This proves necessity. If the rank of  $\Omega_X = m$ , then  $(X_1, \dots, X_m)$  are linearly independent. Moreover,  $(X_1, \dots, X_m) \in \Psi^f$ . Hence, if the rank of  $\Omega = m$ , then there exists numbers  $(a_{ij})$  such that  $Z_j - \bar{Z}_j = \sum_{i=1}^m a_{ij} (X_i - \bar{X}_i)$  for  $j=1, 2, \dots, n$ . Therefore,  $Z_j = b_j + \sum_{i=1}^m a_{ij} X_i$  where  $b_j \equiv \bar{Z}_j - \sum_{i=1}^m a_{ij} \bar{X}_i$ . By the same argument used to prove Proposition IV.1,  $b_j = [1 - \sum_{i=1}^m a_{ij}]R$ . Therefore,  $Z_j = R + \sum_{i=1}^m a_{ij} (X_i - R)$ . By Theorem IV.3,  $(X_1, \dots, X_m, R)$  span  $\Psi^f$ .

It follows from Corollary IV.3.a, that a necessary and sufficient condition for nontrivial spanning of  $\Psi^f$  is that some of the risky securities are redundant securities. Note, however that this condition is sufficient only if securities are priced such that there are no arbitrage opportunities.

In all these derived theorems, the only restriction on investors' preferences was that they prefer more to less. In particular, it was not assumed that investors are necessarily risk averse. Although  $\Psi^f$  was defined in terms of a known joint probability distribution for  $(Z_1, \dots, Z_n)$  which implies homogeneous beliefs among investors, inspection of the proof of Theorem IV.3 shows that this condition can be weakened. If investors agree on a set of portfolios  $(X_1, \dots, X_m, R)$  such that  $Z_j = R + \sum_{i=1}^m a_{ij} (X_i - R)$ ,  $j=1, 2, \dots, n$ , and if they agree on the numbers  $(a_{ij})$ , then by Theorem IV.3,  $(X_1, \dots, X_m, R)$  span  $\Psi^f$  even if investors do not agree on the joint distribution of  $(X_1, \dots, X_m)$ . These appear to be the weakest restrictions on preferences and probability beliefs that can produce nontrivial spanning and the corresponding mutual fund theorem. Hence, to derive additional theorems, it is assumed that all investors

are risk-averse and that investors have homogeneous probability beliefs.

Define  $\Psi^e$  to be the set of all efficient portfolios contained in  $\Psi^f$ .

Proposition IV.2 If  $Z_e$  is the return on a portfolio contained in  $\Psi^e$ , then any portfolio that combines positive amounts of  $Z_e$  with the riskless security is also contained in  $\Psi^e$ .

Proof:

Let  $Z = \delta(Z_e - R) + R$  be the return on a portfolio with positive fraction  $\delta$  allocated to  $Z_e$  and fraction  $(1-\delta)$  allocated to the riskless security. Because  $Z_e$  is an efficient portfolio, there exists a strictly concave, increasing function  $V$  such that  $E\{V'(Z_e)(Z_j - R)\} = 0$ ,  $j=1,2,\dots,n$ . Define  $U(W) \equiv V(aW+b)$  where  $a \equiv 1/\delta > 0$  and  $b \equiv (\delta-1)R/\delta$ . Because  $a > 0$ ,  $U$  is a strictly concave and increasing function. Moreover,  $U'(Z) = aV'(Z_e)$ . Hence,  $E\{U'(Z)(Z_j - R)\} = 0$ ,  $j=1,2,\dots,n$ . Therefore, there exists a utility function such that  $Z$  is an optimal portfolio, and  $Z$  is, therefore, an efficient portfolio.

Hence, from Proposition IV.2, if  $(X_1, \dots, X_M)$  are the returns on  $M$  portfolios that are candidates to span the space of efficient portfolios  $\Psi^e$ , then without loss of generality, it can be assumed that one of the portfolios is the riskless security.

Theorem IV.4 If  $(X_1, \dots, X_m, R)$  span  $\Psi^e$  and if  $Z_j = \bar{Z}_j + \epsilon_j$  where  $E(\epsilon_j) = E(\epsilon_j | X_1, \dots, X_m) = 0$ , then  $\bar{Z}_j = R$ .

Proof:

By hypothesis,  $(X_1, \dots, X_m, R)$  span  $\Psi^e$  and therefore, for each efficient portfolio,  $Z_e^k$ , there exists portfolio weights such that  $Z_e^k = R + \sum_{i=1}^m \delta_i^k (X_i - R)$ . Hence,  $E(\epsilon_j | Z_e^k) = 0$  for every efficient portfolio.



Therefore, the systematic risk of security  $j$  with respect to efficient portfolio  $k$ ,  $b_j^k = 0$ . By Theorem III.1,  $\bar{Z}_j = R$ .

Hence, by Theorem IV.4, if in equilibrium, the total source of variation in a security's return is noise relative to a set of portfolios that span  $\Psi^e$ , then risk averse investors will be willing to hold the total quantity of that security outstanding even though they are not "compensated" by a positive expected excess return. Moreover, Theorem IV.4 suggests that the equilibrium expected return on a security will depend upon the joint distribution of its return with the set of spanning portfolios.

Theorem IV.5 If  $(X_1, \dots, X_m, R)$  span  $\Psi^e$  and if there exists numbers  $(a_{ij})$  such that  $Z_j = \bar{Z}_j + \sum_{i=1}^m a_{ij}(X_i - \bar{X}_i) + \epsilon_j$  where  $E(\epsilon_j) = E(\epsilon_j | X_1, \dots, X_m) = 0$ , then  $\bar{Z}_j = R + \sum_{i=1}^m a_{ij}(\bar{X}_i - R)$ .

Proof:

Let  $Z^+$  be the return on a portfolio with fraction  $\delta_i^+$  allocated to portfolio  $X_i$ ,  $i=1, \dots, m$ ;  $\delta$  allocated to security  $j$ , and  $1 - \delta - \sum_{i=1}^m \delta_i^+$  allocated to the riskless security. If  $\delta_i^+$  is chosen such that  $\delta_i^+ = -\delta a_{ij}$ , then  $Z^+ = R + \delta[\bar{Z}_j - R - \sum_{i=1}^m a_{ij}(\bar{X}_i - R)] + \delta \epsilon_j$ . By Theorem IV.4,  $\bar{Z}^+ = R$ . But  $\delta$  can be chosen arbitrarily. Therefore,  

$$\bar{Z}_j = R + \sum_{i=1}^m a_{ij}(\bar{X}_i - R).$$

Hence, if the return on a security can be written in a linear form relative to the spanning portfolios  $(X_1, \dots, X_m, R)$ , then its expected excess return is completely determined by the expected excess returns on the spanning portfolios and the weights  $a_{ij}$ .

The following theorem first proved by Ross [72] shows that security returns can always be written in this linear form relative to a set of

spanning portfolios.

Theorem IV.6 A necessary and sufficient condition for  $(X_1, \dots, X_m, R)$  to span the set of efficient portfolios  $\Psi^e$  is that there exists numbers  $(a_{ij})$  such that for  $j=1, \dots, n$ ,  $Z_j = R + \sum_{i=1}^m a_{ij}(X_i - R) + \epsilon_j$  where  $E(\epsilon_j) = E(\epsilon_j | X_1, \dots, X_m) = 0$ .

Proof:

The proof of sufficiency is straightforward. Let  $Z_p$  be the return on a feasible portfolio with fraction  $\delta_j$  invested in security  $j$ ,  $j=1, 2, \dots, n$  and  $1 - \sum_{j=1}^n \delta_j$  invested in the riskless security. Then  $Z_p = R + \sum_{j=1}^n \delta_j (Z_j - R)$ . By hypothesis,  $Z_p$  can be rewritten as  $Z_p = R + \sum_{i=1}^m \delta'_i (X_i - R) + \epsilon_p$  where  $\delta'_i \equiv \sum_{j=1}^n \delta_j a_{ij}$  and  $\epsilon_p \equiv \sum_{j=1}^n \delta_j \epsilon_j$ . Consider the feasible portfolio with return  $Z_p'$  constructed by allocating fraction  $\delta'_i$  to portfolio  $X_i$  and fraction  $1 - \sum_{i=1}^m \delta'_i$  to the riskless security. By construction,  $Z_p = Z_p' + \epsilon_p$  where  $E(\epsilon_p) = E(\epsilon_p | Z_p') = 0$ . Hence, for  $\epsilon_p \neq 0$ ,  $Z_p$  is riskier than  $Z_p'$  in the Rothschild-Stiglitz sense, and hence,  $Z_p$  cannot be an efficient portfolio. Therefore, all efficient portfolios can be generated by a portfolio combination of  $(X_1, \dots, X_m, R)$ . The proof of necessity is not presented here because it is long and not constructive. The interested reader can find the proof in Ross [72].

Since  $\Psi^e$  is contained in  $\Psi^f$ , any properties proved for portfolios that span  $\Psi^e$  must be properties of portfolios that span  $\Psi^f$ . From Theorems IV.3 and IV.6, the essential difference is that to span the efficient portfolio set, it is not necessary that linear combinations of the spanning portfolios exactly replicate the return on each available security. Hence, it is not necessary that there exist redundant securities for nontrivial spanning of

$\Psi^e$  to obtain. Of course, both theorems are empty of any empirical content if the size of the smallest spanning set,  $M^*$ , is equal to  $(n+1)$ .

As discussed in the introduction to this section, all the important models of portfolio selection exhibit the nontrivial spanning property for the efficient portfolio set. For all such models that do not restrict the class of admissible utility functions beyond that of risk aversion, the distribution of individual security returns must be such that  $Z_j = R + \sum_{i=1}^m a_{ij}(X_i - R) + \varepsilon_j$  where  $E(\varepsilon_j | X_1, \dots, X_m) = 0$  for  $j=1, \dots, n$ . Moreover, given some knowledge of the joint distribution of a set of portfolios that span  $\Psi^e$  with  $(Z_j - \bar{Z}_j)$ , there exists a method for determining the  $a_{ij}$  and  $\bar{Z}_j$ .

Proposition IV.3 If  $(X_1, \dots, X_m, R)$  span  $\Psi^e$  with  $(X_1, \dots, X_m)$  linearly independent with finite variances and if the return on security  $j$ ,  $Z_j$ , has a finite variance, then the  $(a_{ij})$ ,  $i=1, 2, \dots, m$  in Theorem IV.6 are given by the formula,  $i=1, 2, \dots, m$

$$a_{ij} = \sum_{k=1}^m v_{ik} \text{Cov}[X_k, Z_j]$$

where  $v_{ik}$  is the  $i$ - $k$ <sup>th</sup> element of  $\Omega_X^{-1}$ .

The proof of Proposition IV.3 follows directly from the condition that  $E(\varepsilon_j | X_k) = 0$  which implies that  $\text{Cov}[\varepsilon_j, X_k] = 0$ ,  $k=1, \dots, m$ . The condition that  $(X_1, \dots, X_m)$  be linearly independent is trivial in the sense that knowing the joint distribution of a spanning set, one can always choose a linearly independent subset. The only properties of the joint distributions required to compute the  $a_{ij}$  are the variances and covariances of the  $X_1, \dots, X_m$  and the covariances between  $Z_j$  and  $X_1, \dots, X_m$ . In particular, knowledge of  $\bar{Z}_j$  is not required because  $\text{Cov}[X_k, Z_j] = \text{Cov}[X_k, Z_j - \bar{Z}_j]$ . Hence, for  $m < n$  (and especially so, for  $m < n$ ), there exists a nontrivial information

set which allows the  $a_{ij}$  to be determined without knowledge of  $\bar{Z}_j$ . If  $\bar{X}_1, \dots, \bar{X}_m$  are known, then  $\bar{Z}_j$  can be computed by the formula in Theorem IV.5. By comparison with the example at the end of section III, the information set required there to determine  $\bar{Z}_j$  was a utility function and the joint distribution of its associated optimal portfolio with  $(Z_j - \bar{Z}_j)$ . Here, one must know a complete set of portfolios that span  $\Psi^e$ . However, only the second-moment properties of the joint distribution are required, and no utility function information other than risk aversion is required.

A special case of no little interest is when a single risky portfolio and the riskless security span the space of efficient portfolios. Indeed, the classic model of Markowitz and Tobin which is discussed in section V exhibits this strong form of separation. Moreover, most macroeconomic models have highly-aggregated financial sectors where investors' portfolio choices are limited to simple combinations of two securities: "bonds" and "stocks." The rigorous microeconomic foundation for such aggregation is precisely that  $\Psi^e$  is spanned by a single risky portfolio and the riskless security.

If  $X$  denotes the random variable return on a risky portfolio such that  $(X, R)$  spans  $\Psi^e$ , then the return on any efficient portfolio,  $Z_e$ , can be written as if it had been chosen by combining the risky portfolio with return  $X$  with the riskless security: Namely,  $Z_e = \delta(X - R) + R$  where  $\delta$  is the fraction allocated to the risky portfolio and  $(1 - \delta)$  is the fraction allocated to the riskless security. By Corollary II.1.a, the sign of  $\delta$  will be the same for every efficient portfolio, and therefore, all efficient portfolios will be perfectly positively correlated. If  $\bar{X} > R$ , then by Proposition IV.2,  $X$  will be an efficient portfolio and  $\delta > 0$  for every efficient portfolio.

Proposition IV.4 If  $(Z_1, \dots, Z_n)$  contain no redundant securities and if  $\delta_j$  denotes the fraction of portfolio X allocated to security j, and  $w_j^*$  denotes the fraction of any risk-averse investor's optimal portfolio allocated to security j,  $j=1, \dots, n$ , then for every such risk-averse investor,

$$\frac{w_j^*}{w_k^*} = \frac{\delta_j}{\delta_k}, \quad j, k=1, 2, \dots, n.$$

The proof follows immediately because every optimal portfolio is an efficient portfolio, and the holdings of risky securities in every efficient portfolio are proportional to the holdings in X. Hence, the relative holdings of risky securities will be the same for all risk averse investors.

Whenever Proposition IV.4 holds and if there exists numbers  $\delta_j^*$  such that  $\delta_j^*/\delta_k^* = \delta_j/\delta_k$ ,  $j, k=1, \dots, n$  and  $\sum_1^n \delta_j^* = 1$ , then the portfolio with proportions  $(\delta_1^*, \dots, \delta_n^*)$  is called the Optimal Combination of Risky Assets. If such a portfolio exists, then without loss of generality, it can always be assumed that  $X = \sum_1^n \delta_j^* Z_j$ .

Proposition IV.5 If  $(X, R)$  spans  $\Psi^e$ , then  $\Psi^e$  is a convex set.

Proof:

Let  $Z_e^1$  and  $Z_e^2$  denote the returns on two distinct efficient portfolios. Because  $(X, R)$  spans  $\Psi^e$ ,  $Z_e^1 = \delta_1(X-R) + R$  and  $Z_e^2 = \delta_2(X-R) + R$ . Because they are distinct,  $\delta_1 \neq \delta_2$ , and so, assume  $\delta_1 \neq 0$ . Let  $Z \equiv \lambda Z_e^1 + (1-\lambda)Z_e^2$  denote the return on a portfolio which allocates fraction  $\lambda$  to  $Z_e^1$  and  $(1-\lambda)$  to  $Z_e^2$  where  $0 \leq \lambda \leq 1$ . By substitution, the expression for Z can be rewritten as  $Z = \delta(Z_e^1 - R) + R$  where  $\delta \equiv [\lambda + \frac{\delta_2}{\delta_1}(1-\lambda)]$ . Because  $Z_e^1$  and  $Z_e^2$  are efficient portfolios, the sign of  $\delta_1$  is the same as the sign of  $\delta_2$ . Hence,  $\delta \geq 0$ . Therefore, by Proposition IV.2, Z

is an efficient portfolio. It follows by induction that for any integer  $k$  and numbers  $\lambda_i$  such that  $0 \leq \lambda_i \leq 1$ ,  $i=1, \dots, k$  and  $\sum_{i=1}^k \lambda_i = 1$ ,  $Z^k \equiv \sum_{i=1}^k \lambda_i Z_e^i$  is the return on an efficient portfolio. Hence,  $\Psi^e$  is a convex set.

Definition A market portfolio is defined to be a portfolio that holds all available securities in proportion to their market values. To avoid the problems of "double counting" caused by financial intermediaries and inter-household issues of securities, the equilibrium market value of a security for this purpose is defined to be the equilibrium value of the aggregate demand by individuals for the security. In models where all physical assets are held by business firms and business firms hold no financial assets, an equivalent definition is that the market value of a security equals the equilibrium value of the aggregate amount of that security issued by business firms. If  $V_j$  denotes the market value of security  $j$  and  $V_R$  denotes the value of the riskless security, then

$$\delta_j^M = \frac{V_j}{\sum_{j=1}^n V_j + V_R}, \quad j=1, 2, \dots, n$$

where  $\delta_j^M$  is the fraction of security  $j$  held in a market portfolio.

Theorem IV.7 If  $\Psi^e$  is a convex set, and if the securities' market is in equilibrium, then a market portfolio is an efficient portfolio.

**Proof:**

Let there be  $K$  risk-averse investors in the economy with the initial wealth of investor  $k$  denoted by  $W_0^k$ . Define  $Z^k \equiv R + \sum_{j=1}^n w_j^k (Z_j - R)$  to

be the return per dollar on investor  $k$ 's optimal portfolio where  $w_j^k$  is the fraction allocated to security  $j$ . In equilibrium,  $\sum_{j=1}^K w_j^k w_0^k = V_j$ ,  $j=1,2,\dots,n$  and  $\sum_{j=1}^K w_j^k \equiv W_0 = \sum_{j=1}^n V_j + V_R$ . Define  $\lambda_k \equiv W_0^k / W_0$ ,  $k=1,\dots,K$ . Clearly,  $0 \leq \lambda_k \leq 1$  and  $\sum_{k=1}^K \lambda_k = 1$ . By definition of a market portfolio,  $\sum_{j=1}^K w_j^k \lambda_k = \delta_j^M$ ,  $j=1,2,\dots,n$ . Multiplying by  $(Z_j - R)$  and summing over  $j$ , it follows that  $\sum_{k=1}^K \lambda_k \sum_{j=1}^n w_j^k (Z_j - R) = \sum_{k=1}^K \lambda_k (Z^k - R) = \sum_{j=1}^n \delta_j^M (Z_j - R) = Z_M - R$  where  $Z_M$  is defined to be the return per dollar on the market portfolio. Because  $\sum_{k=1}^K \lambda_k = 1$ ,  $Z_M = \sum_{k=1}^K \lambda_k Z^k$ . But every optimal portfolio is an efficient portfolio. Hence,  $Z_M$  is a convex combination of the returns on  $K$  efficient portfolios. Therefore, if  $\Psi^e$  is convex, then the market portfolio is contained in  $\Psi^e$ .

Because a market portfolio can be constructed without the knowledge of preferences, the distribution of wealth or the joint probability distribution for the outstanding securities, models in which the market portfolio can be shown to be efficient are more likely to produce testable hypotheses. In addition, the efficiency of the market portfolio provides a rigorous microeconomic justification for the use of a "representative man" in aggregated economic models. Moreover, it is currently fashionable in the real world to advise "passive" investment strategies that simply mix the market portfolio with the riskless security. Provided that the market portfolio is efficient, by Proposition IV.2, no investor following such strategies could ever be convicted of "inefficiency." Unfortunately, necessary and sufficient conditions for the market portfolio to be efficient have not as yet been derived.

However, even if the market portfolio were not efficient, it does have the following important property:

Proposition IV.6 In all portfolio models with homogeneous beliefs and risk-averse investors, the equilibrium expected return on the market portfolio exceeds the return on the riskless security.

The proof follows directly from the proof of Theorem IV.7 and Corollary II.1.a. Clearly,  $\bar{Z}_M - R = \sum_{k=1}^K \lambda_k (\bar{Z}^k - R)$ . By Corollary II.1.a,  $\bar{Z}^k \geq R$  for  $k=1, \dots, K$ , with strict inequality holding if  $Z^k$  is risky. But,  $\lambda_k \geq 0$ . Hence,  $\bar{Z}_M > R$  if any risky securities are held by any investor. Note that using no information other than market prices and quantities of securities outstanding, the market portfolio (and combinations of the market portfolio and the riskless security) is the only risky portfolio where the sign of its equilibrium expected excess return can always be predicted.

Returning to the special case where  $\Psi^e$  is spanned by a single risky portfolio and the riskless security, it follows immediately from Proposition IV.5 and Theorem IV.7 that the market portfolio is efficient. Because all efficient portfolios are perfectly positively correlated, it follows that the risky spanning portfolio can always be chosen to be the market portfolio (i.e.,  $X = Z_M$ ). Therefore, every efficient portfolio (and hence, every optimal portfolio) can be represented as a simple portfolio combination of the market portfolio and the riskless security with a positive fraction allocated to the market portfolio. If all investors want to hold risky securities in the same relative proportions, then the only way in which this is possible is if these relative proportions are identical to those in the market portfolio. Indeed, if there were one best investment strategy, and if this "best" strategy were widely known, then whatever the original statement of the strategy, it must lead to simply this imperative: hold the market portfolio.



Because for every security,  $\delta_j^M \geq 0$ , it follows from Proposition IV.4, that in equilibrium, every investor will hold non-negative quantities of risky securities, and therefore, it is never optimal to short sell risky securities. Hence, in models where  $m=1$ , the introduction of restrictions against short-sales will not affect the equilibrium.

Theorem IV.8 If  $(Z_M, R)$  span  $\Psi^e$ , then the equilibrium expected return on security  $j$ , can be written as

$$\bar{Z}_j = R + \beta_j (\bar{Z}_M - R)$$

where

$$\beta_j \equiv \frac{\text{Cov}[Z_j, Z_M]}{\text{Var}(Z_M)}$$

The proof follows directly from Theorem IV.6 and Proposition IV.3. This relationship, called the Security Market Line, was first derived by Sharpe [84] as a necessary condition for equilibrium in the mean-variance model of Markowitz and Tobin when investors have homogeneous beliefs. This relationship has been central to the vast majority of empirical studies of securities' returns published during the last decade. Indeed, the switch in notation from " $a_{ij}$ " to " $\beta_j$ " in this special case reflects the almost universal adoption of the term "the 'beta' of a security" to mean the covariance of that security's return with the market portfolio divided by the variance of the return on the market portfolio.

In the special case of Theorem IV.8,  $\beta_j$  measures the systematic risk of security  $j$  relative to the efficient portfolio  $Z_M$  (i.e.,  $\beta_j = b_j^M$  as defined in section III), and therefore, beta provides a complete ordering of the risk of individual securities. As is often the case in research, useful

concepts are derived in a special model first. The term "systematic risk" was first coined by Sharpe and was measured by beta. The definition in Section III is a natural generalization. Moreover, unlike the general risk measure of Section III,  $\beta_j$  can be computed from a simple covariance between  $Z_j$  and  $Z_M$ . Securities whose returns are positively correlated with the market are called "pro-cyclical," and will be priced to have positive equilibrium expected excess returns. Securities whose returns are negatively correlated are called "counter-cyclical," and will have negative equilibrium expected excess returns. In general, the sign of  $b_j^k$  can not be determined by the sign of the correlation coefficient between  $Z_j$  and  $Z_e^k$ . However, because  $\partial Y(Z_e^k) / \partial Z_e^k > 0$  for each realization of  $Z_e^k$ ,  $b_j^k > 0$  does imply a generalized positive "association" between the return on  $Z_j$  and  $Z_e^k$ . Similarly,  $b_j^k < 0$ , implies a negative "association."

Let  $\Psi_{\min}$  denote the set of portfolios contained in  $\Psi^f$  such that there exists no other portfolio in  $\Psi^f$  with the same expected return and a smaller variance. Let  $Z(\mu)$  denote the return on a portfolio contained in  $\Psi_{\min}$  such that  $\bar{Z}(\mu) = \mu$ , and let  $\delta_j^\mu$  denote the fraction of this portfolio allocated to security  $j$ ,  $j=1, \dots, n$ .

**Theorem IV.9** If  $(Z_1, \dots, Z_n)$  contain no redundant securities, then (a) for each value  $\mu$ ,  $\delta_j^\mu$ ,  $j=1, \dots, n$  are unique; (b) there exists a portfolio contained in  $\Psi_{\min}$  with return  $X$  such that  $(X, R)$  spans  $\Psi_{\min}$ ; (c)  $\bar{Z}_j - R = a_j(\bar{X} - R)$  where  $a_j \equiv \text{Cov}(Z_j, X) / \text{Var}(X)$ ,  $j=1, 2, \dots, n$ .

Proof:

Let  $\sigma_{ij}$  denote the  $i$ - $j$ <sup>th</sup> element of  $\Omega$  and because  $(Z_1, \dots, Z_n)$  contain no redundant securities,  $\Omega$  is nonsingular. Hence, let  $v_{ij}$  denote the

$i$ - $j$ <sup>th</sup> element of  $\Omega^{-1}$ . All portfolios in  $\Psi_{\min}$  with expected return  $\mu$  must have portfolio weights that are solutions to the problem: minimize  $\sum_{i=1}^n \sum_{j=1}^n \delta_i \delta_j \sigma_{ij}$  subject to the constraint  $\bar{Z}(\mu) = \mu$ . Trivially, if  $\mu = R$ , then  $Z(R) = R$  and  $\delta_j^R = 0$ ,  $j=1,2,\dots,n$ . Consider the case when  $\mu \neq R$ . The  $n$  first-order conditions are:

$$0 = \sum_{j=1}^n \delta_j^\mu \sigma_{ij} - \lambda_\mu (\bar{Z}_i - R), \quad i=1,2,\dots,n$$

where  $\lambda_\mu$  is the Lagrange multiplier for the constraint. Multiplying by  $\delta_i^\mu$  and summing, we have that  $\lambda_\mu = \text{Var}[Z(\mu)]/(\mu - R)$ . By definition of  $\Psi_{\min}$ ,  $\lambda_\mu$  must be the same for all  $Z(\mu)$ . Because  $\Omega$  is nonsingular, the set of linear equations has the unique solution

$$\delta_j^\mu = \lambda_\mu \sum_{i=1}^n v_{ij} (\bar{Z}_i - R), \quad i=1,2,\dots,n.$$

This proves (a). From this solution,  $\delta_j^\mu / \delta_k^\mu$ ,  $j, k=1,2,\dots,n$  are the same for every value of  $\mu$ . Hence, all portfolios in  $\Psi_{\min}$  with  $\mu \neq R$  are perfectly correlated. Hence, pick any portfolio in  $\Psi_{\min}$  with  $\mu \neq R$  and call its return  $X$ . Then every  $Z(\mu)$  can be written in the form  $Z(\mu) = \delta_\mu (X - R) + R$ . Hence,  $(X, R)$  span  $\Psi_{\min}$  which proves (b), and from Theorem IV.6 and Proposition IV.3, (c) follows directly.

From Theorem IV.9,  $a_k$  will be equivalent to  $b_k^K$  as a measure of a security's systematic risk provided that the  $Z(\mu)$  chosen for  $X$  is such that  $\mu > R$ . Like  $\beta_k$ , the only information required to compute  $a_k$  are the joint second-moments of  $Z_k$  and  $X$ . Which of the two equivalent measures will be more useful obviously depends upon the information set that is available.

However, as the following theorem demonstrates, the  $a_k$  measure is the natural choice in the case when there exists a spanning set for  $\Psi^e$  with  $m=1$ .

Theorem IV.10 If  $(X, R)$  span  $\Psi^e$  and if  $X$  has a finite variance, then  $\Psi^e$  is contained in  $\Psi_{\min}$ .

Proof:

Let  $Z_e$  be the return on any efficient portfolio. By Theorem IV.6,  $Z_e$  can be written as  $Z_e = R + a_e(X-R)$ . Let  $Z_p$  be the return on any portfolio in  $\Psi^f$  such that  $\bar{Z}_e = \bar{Z}_p$ . By Theorem IV.6,  $Z_p$  can be written as  $Z_p = R + a_p(X-R) + \epsilon_p$  where  $E(\epsilon_p) = E(\epsilon_p|X) = 0$ . Therefore,  $a_p = a_e$  if  $\bar{Z}_p = \bar{Z}_e$ .  $\text{Var}(Z_p) = a_p^2 \text{Var}(X) + \text{Var}(\epsilon_p) \geq a_p^2 \text{Var}(X) = \text{Var}(Z_e)$ . Hence,  $Z_e$  is contained in  $\Psi_{\min}$ . Moreover,  $\Psi^e$  will be the set of all portfolios in  $\Psi_{\min}$  such that  $\mu \geq R$ .

Thus, whenever there exists a spanning set for  $\Psi^e$  with  $m=1$ , the means, variances, and covariances of  $(Z_1, \dots, Z_n)$  are sufficient statistics to completely determine all efficient portfolios. Such a strong set of conclusions suggests that the class of joint probability distributions for  $(Z_1, \dots, Z_n)$  which admit a two-fund separation theorem will be highly specialized. However, as the following theorems demonstrate, the class is not empty.

Theorem IV.11 If  $(Z_1, \dots, Z_n)$  have a joint normal probability distribution, then there exists a portfolio with return  $X$  such that  $(X, R)$  span  $\Psi^e$ .

Proof:

Using the procedure applied in the proof of Theorem IV.9, construct a risky portfolio contained in  $\Psi_{\min}$ , and call its return  $X$ . Define the random variables,  $\epsilon_k \equiv Z_k - R - a_k(X-R)$   $k=1, \dots, n$ . By part (c) of

that theorem,  $E(\epsilon_k) = 0$ , and by construction,  $\text{Cov}[\epsilon_k, X] = 0$ . Because  $Z_1, \dots, Z_n$  are normally distributed,  $X$  will be normally distributed. Hence,  $\epsilon_k$  is normally distributed, and because  $\text{Cov}[\epsilon_k, X] = 0$ ,  $\epsilon_k$  and  $X$  are independent. Therefore,  $E(\epsilon_k) = E(\epsilon_k | X) = 0$ . From Theorem IV.6, it follows that  $(X, R)$  spans  $\Psi^e$ .

It is straightforward to prove that if  $(Z_1, \dots, Z_n)$  can have arbitrary means, variances, and covariances, then a necessary condition for there to exist a portfolio with return  $X$  such that  $(X, R)$  span  $\Psi^e$  is that  $(Z_1, \dots, Z_n)$  be joint normally distributed. However, it is important to emphasize the word "arbitrary." For example, the joint probability density function,  $p(Z_1, \dots, Z_n)$  is called a symmetric function if for each set of admissible outcomes for  $(Z_1, \dots, Z_n)$ ,  $p(Z_1, \dots, Z_n)$  remains unchanged when any two arguments of  $p$  are interchanged.

Theorem IV.12 If  $p(Z_1, \dots, Z_n)$  is a symmetric function with respect to all its arguments, then there exists a portfolio with return  $X$  such that  $(X, R)$  spans  $\Psi^e$ .

Proof:

By hypothesis,  $p(Z_1, \dots, Z_1, \dots, Z_n) = p(Z_1, \dots, Z_1, \dots, Z_n)$  for each set of given values  $(Z_1, \dots, Z_n)$ . Therefore, from the first-order conditions for portfolio selection, (II.4), every risk averse investor will choose  $w_1^* = w_i^*$ . But, this is true for  $i=1, \dots, n$ . Hence, all investors will hold all risky securities in the same relative proportions. Therefore, if  $X$  is the return on a portfolio with an equal dollar investment in each risky security, then  $(X, R)$  will span  $\Psi^e$ .

Samuelson [77] was the first to examine this class of density functions in a portfolio context. An obvious example of such a joint distribution is when

$Z_1, \dots, Z_n$  are independently and identically distributed which implies that  $p(Z_1, \dots, Z_n)$  is of the form  $p(Z_1) \dots p(Z_n)$ .

A second example comes from the Linear-Factor model developed by Ross [71]. Suppose the returns on securities are generated by

$$(IV.1) \quad Z_j = \bar{Z}_j + \sum_{i=1}^m a_{ij} Y_i + \epsilon_j \quad j=1, \dots, n$$

where  $E(\epsilon_j) = E(\epsilon_j | Y_1, \dots, Y_m) = 0$  and without loss of generality,  $E(Y_i) = 0$  and  $\text{cov}[Y_i, Y_j] = 0$ ,  $i \neq j$ . The random variables  $\{Y_i\}$  represent common factors that are likely to affect the returns on a significant number of securities. If it is possible to construct a set of  $m$  portfolios with returns  $(X_1, \dots, X_m)$  such that  $X_i$  and  $Y_i$  are perfectly correlated,  $i=1, 2, \dots, m$ , then the conditions of Theorem IV.6 will be satisfied and  $(X_1, \dots, X_m, R)$  will span  $\Psi^e$ .

Although, in general, it will not be possible to construct such a set, by imposing some mild additional restrictions on  $\{\epsilon_j\}$ , Ross [71] has derived an asymptotic spanning theorem as the number of available securities,  $n$ , becomes large. While the rigorous derivation is rather tedious, a rough description goes as follows: Let  $Z_p$  be the return on a portfolio with fraction  $\delta_j$  allocated to security  $j$ ,  $j=1, 2, \dots, n$ . From (IV.1),  $Z_p$  can be written as

$$(IV.2) \quad Z_p = \bar{Z}_p + \sum_{i=1}^m a_{ip} Y_i + \epsilon_p$$

where  $\bar{Z}_p = R + \sum_{j=1}^n \delta_j (\bar{Z}_j - R)$ ;  $a_{ip} \equiv \sum_{j=1}^n \delta_j a_{ij}$ ;  $\epsilon_p \equiv \sum_{j=1}^n \delta_j \epsilon_j$ . Consider the set of portfolios (called "well-diversified" portfolios) that have the property  $\delta_j \equiv u_j/n$ , where  $|u_j| \leq M_j < \infty$  and  $M_j$  is independent of  $n$ ,  $j=1, \dots, n$ . Virtually by the definition of a common factor, it is reasonable to assume that for every  $n \gg m$ , a significantly positive fraction of all securities,  $\lambda_i$ , have

$a_{ij} \neq 0$ , and this will be true for each common factor  $i$ ,  $i=1, \dots, m$ . Similarly, because the  $\{\epsilon_j\}$  denotes the variations in securities' returns not explained by common factors, it is also reasonable to assume for large  $n$  that for each  $j$ ,  $\epsilon_j$  is uncorrelated with virtually all other securities' returns. Hence, if the number of common factors,  $m$ , is fixed, then it should be possible to construct a set of well-diversified portfolios  $\{X_k\}$  such that for  $X_k$ ,  $a_{ik} = 0$ ,  $i=1, \dots, m$ ,  $i \neq k$  and  $a_{kk} \neq 0$  for all  $n \gg m$ . It follows from (IV.2), that  $X_k$  can be written as

$$(IV.3) \quad X_k = \bar{X}_k + a_{kk} Y_k + \frac{1}{n} \sum_{j=1}^n u_j^k \epsilon_j, \quad k=1, \dots, m.$$

But  $|u_j^k|$  is bounded, independently of  $n$ , and virtually all the  $\epsilon_j$  are uncorrelated. Therefore, by the Law of Large Numbers, as  $n \rightarrow \infty$ ,  $X_k \rightarrow \bar{X}_k + a_{kk} Y_k$ . So, as  $n$  becomes very large,  $X_k$  and  $Y_k$  become perfectly correlated, and by Theorem IV.6, asymptotically,  $(X_1, \dots, X_m, R)$  will span  $\Psi^e$ . In particular, if  $m=1$ , then asymptotically, two-fund separation will obtain independent of any other distributional characteristics of  $Y_1$  or the  $\{\epsilon_j\}$ .

It is interesting to note that empirical studies of stock market securities' returns have rarely found more than two or three statistically-significant common factors.<sup>16/</sup> Given that there are tens of thousands of different corporate liabilities traded in U.S. securities markets, the assumptions used by Ross are not without some empirical foundation. Indeed, whenever nontrivial spanning of  $\Psi^e$  obtains and the set of risky spanning portfolios can be identified, much of the structure of individual securities returns can be empirically estimated. For example, if  $(X_1, \dots, X_m, R)$  span  $\Psi^e$ , then by Theorem IV.6 and Proposition IV.3, Ordinary Least Squares Regression of the realized excess returns on security  $j$ ,  $Z_j - R$ , on the realized excess

returns of the spanning portfolios,  $(X_1 - R, \dots, X_m - R)$ , will always give unbiased estimates of the  $a_{ij}$ . Of course, for these estimators to be efficient, further restrictions on the  $\{\varepsilon_j\}$  are required to satisfy the Gauss-Markov Theorem.

Although the analyses derived here have been expressed in terms of restrictions on the joint distribution of security returns without explicitly mentioning security prices, it is obvious that these derived restrictions impose restrictions on prices through the identity that  $Z_j \equiv V_j / V_{j0}$  where  $V_j$  is the random variable, end-of-period aggregate value of security  $j$  and  $V_{j0}$  is its initial value. Hence, given the characteristics of any two of these variables, the characteristics of the third are uniquely determined. For the study of equilibrium pricing, the usual format is to derive the equilibrium  $V_{j0}$ , given the distribution of  $V_j$ .

Theorem IV.13 If  $(X_1, \dots, X_m, R)$  spans  $\Psi^e$ ,  $M^* = m + 1$ , and all securities have finite variances, then a necessary condition for equilibrium in the securities' market is that

$$(IV.3) \quad V_{j0} = \frac{\bar{V}_j - \sum_{k=1}^m \sum_{i=1}^m v_{ik} \text{Cov}[X_k, V_j] (\bar{X}_i - R)}{R}, \quad j=1, \dots, n$$

where  $v_{ik}$  is the  $i$ - $k$ <sup>th</sup> element of  $\Omega_x^{-1}$ .

Proof:

Because  $M^* = m+1$ ,  $\Omega_x$  is nonsingular. From the identity  $V_j \equiv Z_j V_{j0}$  and Theorem IV.6,  $V_j = V_{j0} [R + \sum_{i=1}^m a_{ij} (X_i - R) + \varepsilon_j]$  where  $E(\varepsilon_j | X_1, \dots, X_m) = E(\varepsilon_j) = 0$ . Taking expectations  $\bar{V}_j = V_{j0} [R + \sum_{i=1}^m a_{ij} (\bar{X}_i - R)]$ . Noting that  $\text{cov}[X_k, V_j] = V_{j0} \text{Cov}[X_k, Z_j]$ , we have from Proposition IV.3, that  $V_{j0} a_{ij} = \sum_{k=1}^m v_{ik} \text{Cov}[X_k, V_j]$ . By substituting for  $a_{ij}$  in the  $\bar{V}_j$  expression and rearranging terms, the theorem is proved.



Hence, from Theorem IV.13, a sufficient set of information to determine the equilibrium value of security  $j$  is the first and second moments for the joint distribution of  $(X_1, \dots, X_m, V_j)$ . Moreover, the valuation formula has the following important "linearity" properties:

Corollary IV.13.a If the hypothesized conditions of Theorem IV.13 hold and if the end-of-period value of some security is given by  $V = \sum_1^n \lambda_j V_j$  where  $\lambda_j$  are numbers, then in equilibrium,

$$V_o = \sum_1^n \lambda_j V_{jo}.$$

The proof of the corollary follows by substitution for  $V$  in formula (IV.3). This property of formula (IV.3) is called "value-additivity."

Corollary IV.13.b If the hypothesized conditions of Theorem IV.13 hold and if the end-of-period value of some security is given by  $V = qV_j + u$  where  $E(u) = E(u|X_1, \dots, X_m) = \bar{u}$  and  $E(q) = E(q|X_1, \dots, X_m, V_j) = \bar{q}$ , then in equilibrium,

$$V_o = \bar{q}V_{jo} + \bar{u}/R.$$

The proof follows by substitution for  $V$  in formula (IV.3) and by applying the hypothesized conditional expectation conditions to show that  $\text{Cov}[X_k, V] = \bar{q} \text{Cov}[X_k, V_j]$ . Hence, to value two securities whose end-of-period values differ only by multiplicative or additive "noise," one can simply substitute the expected values of the noise terms. As will be shown later, both corollaries are central to the theory of investment by business firms.

If nontrivial spanning of  $\Psi^e$  is to obtain, the joint probability distribution for securities' returns cannot be arbitrary. How restrictive these conditions are cannot be answered in the abstract. First, the

introduction of general equilibrium pricing conditions on securities will impose some restrictions on the joint distribution of returns. Second, the discussed benefits to individuals from having a set of spanning mutual funds may induce the creation of financial intermediaries or additional financial securities which together with pre-existing securities will satisfy the conditions of Theorem IV.6. An important example of the latter is the Arrow model discussed in section V.

An alternative approach to the development of nontrivial spanning theorems is to derive a class of utility functions for investors such that for arbitrary joint probability distributions for the available securities, investors within the class can generate their optimal portfolios from the spanning portfolios. Let  $\Psi^u$  denote the set of optimal portfolios selected from  $\Psi^f$  by investors with strictly concave Von Neumann-Morgenstern utility functions  $U_i$ . Cass and Stiglitz [7] have proved the following theorem:

Theorem IV.14 There exists a portfolio with return  $X$  such that  $(X, R)$  spans  $\Psi^u$  if and only if  $A_i(W) = 1/(a_i + bW) > 0$  where  $A_i$  is the absolute risk aversion function for investor  $i$  in  $\Psi^u$ . <sup>17/</sup>

The family of utility functions whose absolute risk aversion functions can be written as  $1/(a+bW) > 0$  is called the "HARA" (Hyperbolic Absolute Risk Aversion) family. <sup>18/</sup> By appropriate choices for  $a$  and  $b$ , various members of the family will exhibit increasing, decreasing, or constant absolute and relative risk aversion. Hence, if each investor's utility function could be approximated by some member of the HARA family, then it might appear that this alternative approach would be fruitful. However, it should be emphasized that the "b" in the statement of Theorem IV.14

does not have a subscript  $i$ , and therefore, for separation to obtain, all investors in  $\Psi^u$  must have virtually the same utility function.<sup>19/</sup> Moreover, they must agree on the joint probability distribution for  $(Z_1, \dots, Z_n)$ . Hence, the only significant way in which investors can differ is in their endowments of initial wealth.

Cass and Stiglitz also examine the possibilities for general nontrivial spanning ( $1 \leq m < n$ ) by restricting the class of utility functions and conclude, "... it is the requirement that there be any mutual funds, and not the limitation on the number of mutual funds which is the restrictive feature of the property of separability."<sup>20/</sup> Hence, the Cass and Stiglitz analysis is essentially a negative report on this approach to developing spanning theorems.

In closing this section, two further points should be made. First, although virtually all the spanning theorems require the generally implausible assumption that all investors agree upon the joint probability distribution for securities, it is not so unreasonable when applied to the theory of financial intermediation and mutual fund management. In a world where the economic concepts of "division of labor" and "comparative advantage" have content, then it is quite reasonable to expect that an efficient allocation of resources would lead to some individuals (the "fund managers") gathering data and actively estimating the joint probability distributions and the rest either buying this information directly or delegating their investment decisions by "agreeing to agree" with the fund managers' estimates. If the distribution of returns is such that nontrivial spanning of  $\Psi^e$  does not obtain, then there are no gains to financial intermediation over the direct sale of the distribution estimates. However, if nontrivial spanning

does obtain and the number of risky spanning portfolios,  $m$ , is small, then a significant reduction in redundant information processing and transactions can be produced by the introduction of mutual funds. If a significant coalition of individuals can agree upon a common source for the estimates and if they know that, based on this source, a group of mutual funds offered spans  $\Psi^e$ , then they need only be provided with the joint distribution for these mutual funds to form their optimal portfolios. On the supply side, if the characteristics of a set of spanning portfolios can be identified, then the mutual fund managers will know how to structure the portfolios of the funds they offer.

The second point concerns the riskless security. It has been assumed throughout that there exists a riskless security. Although some of the specifications will change slightly, virtually all the derived theorems can be shown to be valid in the absence of a riskless security.<sup>21/</sup> However, the existence of a riskless security vastly simplifies many of the proofs.

## V. Two Special Models of One-Period Portfolio Selection

The two most-cited models in the literature of portfolio selection are the Time-State Preference model of Arrow [2] and Debreu [9] and the Mean-Variance model of Markowitz [49] and Tobin [89]. Because these models have been central to the development of the microeconomic theory of investment, there are already many review and survey articles devoted just to each of these models.<sup>22/</sup> Hence, only a cursory description of each model is presented here with specific emphasis on how each model fits within the framework of the analyses presented in the other sections. Moreover, while, under appropriate conditions, both models can be interpreted as multiperiod, intertemporal portfolio selection models, such an interpretation will be delayed until section VII.

The structure of the Arrow-Debreu model is described as follows: Consider an economy where all possible configurations for the economy at the end of the period can be described in terms of  $M$  possible states of nature. The states are mutually exclusive and exhaustive. It is assumed that there are  $N$  risk averse individuals with initial wealth  $W_0^k$  and von Neumann Morgenstern utility function  $U^k(W)$  for investor  $k$ ,  $k = 1, \dots, N$ . Each individual acts on the basis of subjective probabilities for the states of nature denoted by  $\Pi_k(\theta)$ ,  $\theta=1, \dots, M$ ;  $k=1, \dots, N$ . While these subjective probabilities can differ across investors, it is assumed for each investor that  $0 < \Pi_k(\theta) < 1$ ,  $\theta=1, \dots, M$ . As was assumed in section II, there are  $n$  risky securities with returns per dollar  $Z_j$  and initial market value,  $V_{j0}$ ,  $j=1, \dots, n$ , and the "perfect market" assumptions of that section, (A.1) - (A.4), are assumed here as well. Moreover, if state  $\theta$  obtains, then the return on security  $j$  will be  $Z_j(\theta)$ ,

and all investors agree on the functions  $Z_j(\theta)$ ,  $j=1, \dots, n$ ;  $\theta=1, \dots, M$ . Because the set of states is exhaustive,  $[Z_j(1), \dots, Z_j(M)]$  describe all the possible outcomes for the returns on security  $j$ . In addition, there are available  $M$  "pure" securities with the properties that,  $i=1, \dots, M$ , one unit (share) of pure security  $i$  will be worth \$1 at the end of the period if state  $i$  obtains and will be worthless if state  $i$  does not obtain. If  $P_i$  denotes the price per share of pure security  $i$  and if  $X_i$  denotes its return per dollar, then for  $i=1, \dots, M$ ,  $X_i$  as a function of the states of nature, can be written as  $X_i(\theta) = 1/P_i$  if  $\theta=i$  and  $X_i(\theta) = 0$  if  $\theta \neq i$ . All investors agree on the functions  $\{X_i(\theta)\}$ ,  $i, \theta=1, \dots, M$ .

Let  $Z = Z(N_1, \dots, N_M)$  denote the return per dollar on a portfolio of pure securities that holds  $N_j$  shares of pure security  $j$ ,  $j=1, \dots, M$ . If  $V_0(N_1, \dots, N_M) \equiv \sum_{j=1}^M N_j P_j$  denotes the initial value of this portfolio, then the return per dollar on the portfolio, as a function of the states of nature, can be written as  $Z(\theta) = N_\theta/V_0$ ,  $\theta=1, \dots, M$ .

Proposition V.1 There exists a riskless security, and its return per dollar  $R$  equals  $1/(\sum_{j=1}^M P_j)$ .

Consider the pure-security portfolio that holds one share of each pure security ( $N_j=1$ ,  $j=1, \dots, M$ ). The return per dollar,  $Z$ , is the same in every state of nature and equals  $1/V_0(1, \dots, 1)$ . Hence, there exists a riskless security and by (A.3), its return  $R$  is given by  $1/(\sum_{j=1}^M P_j)$ .

Proposition V.2 For each security  $j$  with return  $Z_j$ , there exists a portfolio of pure securities whose return per dollar exactly replicates  $Z_j$ .

Proof:

Let  $Z^j \equiv Z(Z_j(1), \dots, Z_j(M))$  denote the return on a portfolio of pure securities with  $N_\theta = Z_j(\theta)$ ,  $\theta=1, \dots, M$ . It follows that  $V_o(Z_j(1), \dots, Z_j(M)) = \sum_{i=1}^M P_i Z_j(i)$  and  $Z^j(\theta) = Z_j(\theta)/V_o$ ,  $\theta=1, \dots, M$ . Consider a three-security portfolio with return  $Z_p$  where fraction  $V_o$  is invested in  $Z^j$ ; fraction  $-1$  is invested in  $Z_j$ ; and fraction  $1 - V_o - (-1) = (2-V_o)$  is invested in the riskless security. The return per dollar on this portfolio as a function of the states of nature can be written as

$$Z_p(\theta) = (2-V_o)R + V_o Z^j(\theta) - Z_j(\theta) = (2-V_o)R$$

which is the same for all states. Hence,  $Z_p$  is a riskless security and by (A.3),  $Z_p(\theta) = R$ . Therefore,  $V_o = 1$ , and  $Z^j(\theta) = Z_j(\theta)$ ,  $\theta=1, \dots, M$ .

Proposition V.3 The set of pure securities with returns  $(X_1, \dots, X_M)$  spans the set of all feasible portfolios that can be constructed from the  $M$  pure securities and the  $n$  other securities.

The proof follows immediately from Proposition V.1 and V.2. Hence, whenever a complete set of pure securities exists or can be constructed from the available securities, then every feasible portfolio can be replicated by a portfolio of pure securities. Models in which such a set of pure securities exists are called "complete markets" models in the sense that any additional securities or markets would be redundant. Necessary and sufficient conditions for such a set to be constructed from the available  $n$  risky securities alone are that  $n \geq M$ ; Assumption (A.3) holds; and the rank of the variance-covariance matrix of returns,  $\Omega$ , equals  $M$ .

The connection between the pure securities of the Arrow-Debreu model and the mutual fund theorems of section IV are obvious. To put this model in comparable form, we can choose the alternative spanning set  $(X_1, \dots, X_m, R)$  where  $m = M-1$ . From Theorem IV.3, the returns on the risky securities can be written as

$$(V.1) \quad Z_j = R + \sum_{i=1}^M a_{ij} (X_i - R) \quad , \quad j=1, \dots, m$$

where the numbers  $(a_{ij})$  are given by Proposition IV.3

Note that nowhere in the derivation were the subjective probability assessments of the individual investors required. Hence, individual investors need not agree on the joint distribution for  $(X_1, \dots, X_m)$ . However, by Theorem IV.3, investors cannot have arbitrary beliefs in the sense that they must agree on the  $(a_{ij})$  in (V.1).

Proposition V.4 If  $V_j(\theta)$  denotes the end-of-period value of security  $j$  if state  $\theta$  obtains, then a necessary condition for equilibrium in the securities' market is that

$$V_{j0} = \sum_{k=1}^M p_k V_j(k) \quad , \quad j=1, \dots, n$$

The proof follows immediately from the proof of Proposition V.2. It was shown there that  $V_0 = \sum_{k=1}^M p_k Z_j(k) = 1$ . Multiplying both sides by  $V_{j0}$  and noting the identity  $V_j(k) \equiv V_{j0} Z_j(k)$ , it follows that  $V_{j0} = \sum_{k=1}^M p_k V_j(k)$ .

However, by Theorem IV.3 and Proposition V.3, it follows that the  $\{V_{j0}\}$  can also be written as

$$(V.2) \quad V_{j0} = \frac{\bar{V}_j - \sum_{i=1}^m \sum_{k=1}^m v_{ik} \text{Cov}[X_k, V_j](\bar{X}_i - R)}{R} \quad , \quad j=1, \dots, n$$



where  $v_{ik}$  is the  $i$ - $k^{\text{th}}$  element of  $\Omega_X^{-1}$ . Hence, from (V.2) and Proposition V.4, it follows that the  $a_{ij}$  in (V.1) can be written as

$$(V.3) \quad a_{ij} = [Z_j(i) - R] / [1/P_i - R], \quad i=1, \dots, m; j=1, \dots, n.$$

From (V.3), given the prices of the securities,  $\{P_i\}$  and  $\{V_{jo}\}$ , the  $\{a_{ij}\}$  will be agreed upon by all investors if and only if they agree upon the  $\{V_j(i)\}$  functions.

While it is commonly believed that the Arrow-Debreu model is completely general with respect to assumptions about investors' beliefs, the assumption that all investors agree on the  $\{V_j(i)\}$  functions can impose nontrivial restrictions on these beliefs. In particular, when there is production, it will in general be inappropriate to define the states, tautologically, by the end-of-period values of the securities, and therefore, investors will at least have to agree on the technologies specified for each firm.<sup>23/</sup> However, as discussed in section IV, it is unlikely that a model without some degree of homogeneity in beliefs (other than agreement on currently observed variables) can produce testable restrictions. Among models that do produce such testable restrictions, the assumptions about investors' beliefs in the Arrow-Debreu model are probably the most general.

Finally, for the purposes of portfolio theory, the Arrow-Debreu model is a special case of the spanning models of section IV which serves to illustrate the generality of the linear structure of those models.

The most elementary type of portfolio selection model in which all securities are not perfect substitutes is one where every portfolio can be characterized by two numbers: its "risk" and its "return." The mean-variance portfolio selection model of Markowitz [49] and Tobin [89] is such

a model. In this model, each investor chooses his optimal portfolio so as to maximize a utility function of the form,  $H[E(W), \text{Var}(W)]$ , subject to his budget constraint where  $W$  is his random variable end-of-period wealth. The investor is said to be "risk-averse in a mean-variance sense" if  $H_1 > 0$ ;  $H_2 < 0$ ; and  $H_{11}H_{22} - H_{12}^2 > 0$  where subscripts denote partial derivatives.

In an analogous fashion to the general definition of an efficient portfolio in section II, a feasible portfolio will be called a mean-variance efficient portfolio if there exists a risk averse mean-variance utility function such that this feasible portfolio would be preferred to all other feasible portfolios. Let  $\Psi_{mv}^e$  denote the set of mean-variance efficient portfolios. As defined in section IV,  $\Psi_{min}$  is the set of feasible portfolios such that there exists no other portfolio with the same expected return and a smaller variance. For a given initial wealth  $W_0$ , every risk-averse investor would prefer the portfolio with the smallest variance among those portfolios with the same expected return. Hence,  $\Psi_{mv}^e$  is contained in  $\Psi_{min}$ .

Proposition V.5 If  $(Z_1, \dots, Z_n)$  are the returns on the available risky securities, then there exists a portfolio contained in  $\Psi_{mv}^e$  with return  $X$  such that  $(X, R)$  spans  $\Psi_{mv}^e$  and  $\bar{Z}_j - R = a_j (\bar{X} - R)$  where  $a_j \equiv \text{Cov}(Z_j, X) / \text{Var}(X)$ ,  $j=1, 2, \dots, m$ .

The proof follows immediately from theorem IV.9<sup>24/</sup>. Hence, all the properties derived in the special case of two-fund spanning ( $m=1$ ) in section IV apply to the mean-variance model. Indeed, because all such investors would prefer a higher expected return for the same variance of return,  $\Psi_{mv}^e$  is the set of all portfolios contained in  $\Psi_{min}$  such that their expected returns are equal to or exceed  $R$ . Hence, the mean-variance model is also a special case of the spanning models developed in section IV.

If investors have homogeneous beliefs, then the equilibrium version of the mean-variance model is called the Capital Asset Pricing Model.<sup>25/</sup> It follows from Proposition IV.5, and Theorem IV.7, that in equilibrium, the market portfolio can be chosen as the risky spanning portfolio. From Theorem IV.8, the equilibrium structure of expected returns must satisfy the Security Market Line.

Because of the mean-variance model's attractive simplicity and its strong empirical implications, a number of authors<sup>26/</sup> have studied the conditions under which such a criterion function is consistent with the expected utility maxim. Like the studies of general spanning properties cited in section IV, these studies examined the question in two parts: (i) what is the class of probability distributions such that the expected value of a arbitrary concave utility function can be written solely as a function of mean and variance; and (ii) what is the class of strictly concave von Neumann-Morgenstern utility functions whose expected value can be written solely as a function of mean and variance for arbitrary distributions. Since the class of distributions in (i) was shown in section IV to be equivalent to the class of finite variance distributions that admit two-fund spanning of the efficient set, the analysis will not be repeated here. To answer (ii), it is straightforward to show that a necessary condition is that  $U$  be of the form,  $W - bW^2$ , with  $b > 0$ . This member of the HARA family is called the quadratic, and will only satisfy the von Neumann axioms if  $W \leq 1/2b$ , for all possible outcomes for  $W$ . Even if  $U$  is defined to be  $\text{Max}[W - bW^2, 1/4b]$  so that  $U$  satisfies the axioms for all  $W$ , for general distributions, its expected value can be written as a function of just  $E(W)$  and  $\text{Var}(W)$ , only if the maximum possible outcome for  $W$  is less than  $1/2b$ .

Although both the Arrow-Debreu and Markowitz-Tobin models were shown to be special cases of the spanning models in section IV, they deserve special attention because they are unquestionably the genesis of these general models.

## VI. Investment Theory for the Firm

In the preceding section of the paper, the portfolio selection problem for individuals was solved and a set of necessary conditions for financial equilibrium were derived. Neither the current consumption-saving choice by individuals nor the allocation of resources for physical production were explicitly considered. Hence, the preceding analyses are best viewed as a partial equilibrium study of the financial markets taking current consumption and production plans as fixed. In this section, the theory of optimal investment in physical assets is presented, and the connection between production theory and the financial markets is made explicit. However, the optimal choice of current consumption by individuals is still taken as given leaving until Section VII its explicit examination.

There are two essential differences between the portfolio selection problem and the optimal allocation of physical investment problem. First, because it was assumed that individuals behave "competitively" with respect to the securities market (i.e., assumption (A.2) in Section II), only linear allocations of resources are allowed among the available investments in the portfolio selection problem. In general, physical production technologies can be nonlinear functions of their inputs. Also, because the available investments in the portfolio problem are securities, it is possible for an individual to invest negative amounts in specific investments. For production technologies, the amount of physical investment must be nonnegative.

Second, in developed market economies, most of the physical production is carried on by business firms where the production decisions are made by managers who are generally not the (sole) owners of the firm. Hence, it is

important to know the conditions under which an efficient allocation of resources among the available production technologies will obtain when there exists an institutional separation between the owners of the resources and the managers of these resources. In essence, does there exist a set of investment decision rules such that if firm managers follow these rules, the firm will operate "as if" the owners of the firm had made the production decisions directly. Hence, unlike the utility function of an individual which is taken to be exogeneous in the portfolio selection problem, the criterion function for production decisions by the firm is derived, and is therefore endogeneous.

It is well known in the theory of production under certainty that if firms are competitive and make their production decisions so as to maximize the market value of the firm, then a competitive equilibrium is a pareto optimum. Arrow [ 2 ] and Debreu [ 9 ] extended these results to uncertainty within the framework of their "complete markets" model. However, the extent to which these results carry over to "incomplete markets" has not as yet been determined. Diamond [10] and more recently Leland [46], Eckern and Wilson [12], Radner [69], and Hart [31] have derived conditions under which stockholders will unanimously agree on a production decision. However, the "derived" criterion function for the firm will not, in general, be "to maximize market value." While the nonoptimality of the value maximization rule is not surprising when firms do not behave competitively, Jensen and Long [40], Fama [16], and Stiglitz [88] claim that it can be nonoptimal even when firms are competitive. Merton and Subrahmanyam [62] argue that the posited firm behavior in the Jensen-Fama-Stiglitz models is not competitive and, therefore, the findings using their models are not inconsistent with those proven for complete markets.

While these issues have not yet been resolved and will not be here, it is hoped that the analysis of this section will shed some light on the controversy.

A firm is defined by the single production technology it owns and let there be  $n$  production technologies where the random variable end-of-period value of firm  $j$ ,  $V_j(I_j; \theta_j)$ , can be written as a function of the initial investment in its technology,  $I_j$ , and an exogeneously specified random variable  $\theta_j$ . Suppose there exists a set of portfolios with random variable returns per dollar  $(X_1, \dots, X_m, R)$  that span the efficient portfolio set. Let  $V_{j0}(I_j)$  be the equilibrium initial value of firm  $j$  after  $I_j$  has been invested in its technology. Therefore,  $V_{j0}(I_j)$  will satisfy the formula (IV.3) in Theorem IV.13.

Consider a risk averse individual with utility function  $U(W)$ . Prior to investment and trading, his initial endowment contains  $\lambda_j$  fractional ownership of firm  $j$ ,  $j = 1, \dots, n$ , in addition to other exogeneous assets with market value  $\bar{W}_0$ . His initial wealth,  $W_0$ , can be written as

$$(VI.1) \quad W_0(I_1, \dots, I_n) = \sum_{j=1}^n \lambda_j [V_{j0}(I_j) - I_j] + \bar{W}_0$$

Consider the expanded portfolio selection problem where the investor chooses his optimal portfolio allocation and the amount of initial investment allocated to those firms in which he has some positive initial ownership. As a natural extension of assumption (A.2) in the standard portfolio problem of Section II, it is assumed that the investor believes that his actions (including his choices for  $I_j$ ) cannot affect the probability distribution of returns on the set of spanning portfolios  $(X_1, \dots, X_m, R)$ .

By hypothesis,  $(X_1, \dots, X_m, R)$  span  $\psi^e$ . Hence, without loss of generality, we can formulate the problem as:

$$(VI.2) \quad \text{Max}_{\{w_i, I_j\}} E\{U([\sum_1^m w_i (X_i - R) + R]W_0)\}$$

where  $w_i$  is the fraction of his initial wealth allocated to spanning portfolio  $X_i$ ,  $i = 1, \dots, m$ ;  $(1 - \sum_1^m w_i)$  is the fraction allocated to the riskless security;  $W_0$  is given by (VI.1); and "E" denotes the expectation operator over his subjective probability distribution for  $(X_1, \dots, X_m)$ .

Define  $W^* \equiv [\sum_1^m w_i^* (X_i - R) + R][\sum_1^n \lambda_j [V_{jo}(I_j^*) - I_j^*] + \bar{W}_0]$  where  $(w_1^*, \dots, w_m^*)$  are his optimal portfolio weights and  $(I_1^*, \dots, I_n^*)$  are his optimal choices for initial investment in the  $n$  firms. For an interior solution, the optimal choices will satisfy the first order conditions

$$(VI.3) \quad E\{U'(W^*)(X_i - R)\} = 0, \quad i = 1, \dots, m$$

and

$$(VI.4) \quad E\{U'(W^*)\lambda_j Z^*[V'_{jo}(I_j^*) - 1]\} = 0, \quad j = 1, \dots, n,$$

where  $Z^*$  is the return per dollar on his optimal portfolio and  $V'_{jo}(I_j) \equiv \partial V_{jo} / \partial I_j$ . By the first-order conditions (VI.3),  $E\{U'(W^*)Z^*\} = R E\{U'(W^*)\} > 0$  by non-satiation. Hence, for all firms where the individual has positive initial ownership ( $\lambda_j > 0$ ), the first-order conditions (VI.4) can be rewritten as

$$(VI.5) \quad V'_{jo}(I_j^*) - 1 = 0.$$



From (VI.1), optimality conditions (VI.5) simply imply that for a given distribution of the spanning portfolios, the investor would prefer the allocation of physical investment across firms to be the one that maximizes his initial wealth. But the initial wealth function in (VI.1) does not depend upon the investor's utility function. Hence, every investor with a positive ownership of firm  $j$  will agree that the amount of physical investment in firm  $j$  should be chosen so as to maximize its market value.

And if the investors agree on the  $V_{jo}(I_j)$  function, then they will agree on the optimal choice for  $I_j$ . Note: This latter condition can be satisfied even if investors do not have homogeneous beliefs about the probability distribution for  $(X_1, \dots, X_m)$ . For example, for the Arrow-Debreu model in Section V, it was shown that all investors will agree on the valuation formula (IV.3), even though they have heterogeneous beliefs about the probability distribution for the states.

In general, with heterogeneous beliefs, investors will disagree on the amount of investment that the firm should take to maximize its market value. However, as long as each shareholder perceives changes in the firm's investment as having no effect on the return distributions of the efficient set, they will agree that any change in investment which increases the market value of the firm will be preferred to ones that do not.

Theorem VI.1 If changes in the investment made by firm  $j$  do not change the distribution of the efficient portfolio set, then all shareholders of firm  $j$  will agree that its investment should be chosen to maximize its market value. In addition, if all

such shareholders agree on the valuation function  $V_{j0}(I_j)$ , then the value-maximizing firm will operate "as if" the owners of the firm had made the investment decision directly.

Of course, the critical issue is under what conditions will the hypothesis of Theorem VI.1 obtain? One clear example is if, for every choice  $I_j$ , the resulting equilibrium random variable return on firm  $j$  is such that it is a "redundant" security (as defined in Sections II and IV), then changes in the investment made by firm  $j$  will not affect the equilibrium distribution of the efficient portfolio set. This condition will occur whenever, from the point of view of investors, there exist other securities that are perfect substitutes for security  $j$ . While "perfect substitutability" is sufficient, it is not necessary as the following analysis demonstrates.

Consider the same risk averse individual in the previous analysis, but for simplicity it is assumed that he has an initial endowment of positive fractional ownership of firm  $j$  only. Assume that  $(X_1, \dots, X_m, Z_j, R)$  span the efficient portfolio set. The investor believes that his actions cannot affect the distribution of returns for  $(X_1, \dots, X_m, R)$  and he can only affect the distribution of  $Z_j$  through his choice of  $I_j$ . (VI.2) can be reformulated as

$$(VI.6) \quad \text{Max}_{\{w_i, I_j\}} E\{U([\sum_{i=1}^m w_i (X_i - R) + w_{m+1} (Z_j - R) + R]W_0)\}$$

where  $w_{m+1}$  is the fraction of his initial wealth allocated to firm  $j$ . The first-order conditions for an interior solution can be written as

$$(VI.7) \quad \begin{aligned} E\{U'(W^*)(X_i - R)\} &= 0, & i = 1 \dots, m \\ E\{U'(W^*)(Z_j - R)\} &= 0 \end{aligned}$$

and

$$(VI.8) \quad E\{U'(W^*)[(V'_{j0}(I_j^*) - 1)\lambda_j Z^* + W_{m+1}^* W_0^* Z'_j(I_j^*)]\} = 0$$

where  $W_0^* \equiv \lambda_j [V_{j0}(I_j^*) - I_j^*] + \bar{W}_0$  and  $Z'_j(I_j)$  is shorthand for  $\partial[V_j(I_j; \theta_j)/V_{j0}(I_j)]/\partial I_j$  for each given value of the random variable  $\theta_j$ .

Using (VI.7), (VI.8) can be rewritten as

$$(VI.9) \quad \lambda_j R E\{U'(W^*)\}(V'_{j0}(I_j^*) - 1) + w_{m+1}^* W_0^* E\{U'(W^*)Z'_j(I_j^*)\} = 0.$$

Hence, without further conditions on the distribution of the marginal return on security  $j$ ,  $Z'_j(I_j^*)$ , the value-maximizing choice for  $I_j$  will not be optimal. Of course, if the post-investment holdings of security  $j$  by the investor are zero, ( $w_{m+1}^* = 0$ ), then the value-maximizing choice is optimal for him, but this will not sustain a post-investment equilibrium unless all investors would choose  $w_{m+1}^* = 0$ .

However, if the random variable marginal return can be written in the form

$Z'_j(I_j^*) = \gamma_j + \sum_{i=1}^m \mu_{ij}(X_i - R) + \eta_j(Z_j - R) + \epsilon_j$  where  $E(\epsilon_j) = E(\epsilon_j | X_1, \dots, X_m, Z_j) = 0$ , then using (VI.7), (VI.9) can be rewritten as

$$(VI.10) \quad \lambda_j R (V'_{j0}(I_j^*) - 1) + w_{m+1}^* W_0^* \gamma_j = 0.$$

If  $\gamma_j = 0$ , then value maximization is again optimal. Even if  $\gamma_j \neq 0$ , if the condition is imposed that  $I_j^*$  be chosen such that post-investment investors in firm  $j$  would not want to change it, then  $\lambda_j V_{j0}(I_j^*) = w_{m+1}^* W_0^*$ , and (VI.10) can be rewritten as

$$(VI.11) \quad R(V'_{jo}(I_j^*) - 1) + \gamma_j V_{jo}(I_j^*) = 0.$$

Hence, if the firm chooses its investment so as to satisfy (VI.11), then all (pre- and post-investment) stockholders will agree on the investment chosen by the firms. (VI.11) is an example of stockholder unanimity and the conditions imposed on  $Z'_j$  are the same (except for our including  $\epsilon_j$ ) as in the Proposition proved by Eckern and Wilson [12, p. 175].

If the available technologies exhibit stochastic constant returns (as in the model of Diamond [10]) and if the technologies are freely available, then a necessary condition for equilibrium is that the value of firm  $j$  be no larger than the value of its factor inputs. In our structure,  $V_{jo}(I_j^*) \leq I_j^*$ . If the investment process is "reversible" or if ex ante investment decisions are agreed upon by ex post stockholders, then  $V_{jo}(I_j^*) = I_j^*$ . But stochastic constant returns implies that  $V_j(I_j; \theta_j) = I_j \theta_j$  and therefore  $Z_j(I_j^*) = \theta_j$ . Hence,  $Z'_j(I_j^*) = 0$ , and from (VI.9), value maximization is optimal.

In summary, if the choice of investment for each firm taken individually is perceived as having no impact on the return distribution of the efficient portfolio set, then value maximization will produce the same investment allocation as would have been chosen by the stockholders of the firms if they had made the decision directly.

There appears to be a connection between this condition and the conditions for a competitive securities market. Since a change in a firm's investment choice will change its market value and can change its own return distribution, the traditional definition of a competitive market as one in which no single participant's action can affect prices in that market makes no sense

for a securities market. However, it seems to me that a reasonable condition for a securities market to be competitive is that no single participant's action can affect the distribution of returns on any set of portfolios that span the efficient portfolio set. In the case of a certainty world, this condition implies that no single participant's actions can change the interest rate. In the Arrow-Debreu model, this condition implies that no single participant's actions can affect the prices of the pure securities  $(P_1, \dots, P_M)$ . In the Capital Asset Pricing model, it implies that no single participant's actions can affect the return distribution for either the market portfolio or the interest rate. Like the traditional definition, this condition implies that no single participant's actions can affect anything "that matters" to everyone.

While this condition is appealing, there are a number of issues that must be resolved before it could be accepted as a definition. First, unlike prices, return distributions are not observable. In models with homogeneous expectations, this is not a problem because, given a set of prices, the return distributions are unambiguously defined. However, in incomplete markets models with heterogeneous beliefs, there are obvious complications. Second, the definition, existence, and optimality properties of a "competitive" equilibrium would require derivation. Third, while it is already a standard assumption of portfolio theory that the actions of a single individual or financial institution will not affect the distribution of returns on securities, the case for individual firms investment decisions is more subtle. In discussing the value maximization criterion in incomplete markets, Radner notes that with the exception of the Diamond model, "I have not seen a formulation of this hypothesis that enables the producer to act as a price-taker; i.e., that does not imply that the producer is able to calculate the effect of his actions

on the equilibrium prices.<sup>27/</sup> In general, his comment would also apply to the assumption that producers take the distribution of the efficient set as given. While this issue merits careful study, my suspicion is that if significant non-trivial spanning of the efficient set obtains, then a producer might not be able to calculate the effect of his actions on the equilibrium return distribution for the efficient set. If these issues can be satisfactorily decided, then much of the current controversy surrounding the theory of the firm with incomplete securities markets will be settled.

In discussing the investment decision by firms, it was assumed that each firm had a single production technology and the investment decision was to choose the intensity at which the technology is operated. However, firms can also make investments by buying other firms' technologies. To examine the effects of this type of investment, suppose that the first  $k$  technologies are owned by a single firm and so, instead of  $n$ , there are only  $n - k + 1$  firms. If there are no economies of scale in such a consolidation (i.e., no "synergy"), then the end-of-period value for the consolidated firm can be written as

$$(VI.12) \quad V(I_1, \dots, I_k; \theta_1, \dots, \theta_k) \equiv \sum_1^k V_j(I_j; \theta_j).$$

Theorem VI.2 Suppose  $(X_1, \dots, X_m, R)$  span the efficient portfolio set and (VI.12) holds. If the return distribution for  $(X_1, \dots, X_m, R)$  remains the same as it was prior to consolidation, then, post consolidation, the initial value of the consolidated firm is given by

$$V_0 = \sum_1^k V_{j0}(I_j).$$

The proof follows immediately from the "value additivity" properly derived in Corollary IV.13.a. Indeed, if the investment decision of any single firm (including the consolidated firm) cannot affect the distribution of  $(X_1, \dots, X_m, R)$ ; (VI.12) holds; and firms make their investment decision according to the value maximization rule, then from Theorem VI.2, the investment allocation chosen by the consolidated firm will be identical to the allocation that would have been chosen by the  $k$  individual firms. In this case, consolidation of firms has no effect on individual welfare because both the initial wealth of each individual and the return distribution on his optimal portfolio will remain unchanged. Indeed, if (VI.12) holds and the (pre-consolidation) allocation of investment is optimal, then individuals cannot be any better off after consolidation. However, it is possible that, post-consolidation, investors could be worse off. For example, suppose there do not exist "perfect substitute" securities for at least some of the individual firms and pre-consolidation, not all investors chose to hold these  $k$  firms in proportion to their market values. Since, post-consolidation, investors can only hold these  $k$  firms in the same relative proportions, some pre-consolidation optimal portfolios will not be feasible after consolidation, and therefore some individuals could be worse off.

In an analogous fashion to the individual's portfolio selection behavior, it has been argued that firms acquire other firms to reduce risk. However, unless an acquisition provides a production opportunity otherwise unavailable or unless the acquired firm was not operating its technology efficiently, diversification by the firm cannot improve the welfare of individuals, and in some cases it can reduce it. This result serves as a warning against the indiscriminate use of models that treat firms "like" individuals and ascribe

to the firms exogeneously given utility functions rather than deriving endogeneous criterion functions for ranking their choices.

In the discussion of the investment decision by firms, the method of financing these investments was not made explicit. Implicitly, it was assumed that firms used internally available funds or issued additional financial securities where all financial claims against the firm were of a single type called equity. Of course, it is well known that, in addition to equity, firms also issue other types of financial claims (e.g., debt, preferred stock, and convertible bonds). The choice of the menu of financial liabilities is called the firm's financing decision. Although, in general, the optimal investment and financing decisions by a firm are determined simultaneously, it is useful to study the financing decision by taking the investment decision as given.

Consider firm  $j$  with random variable end-of-period value  $V^j$  and  $q$  different financial claims. The  $k^{\text{th}}$  such financial claim is defined by the function  $f_k[V^j]$  which describes how the holders of this security will share in the end-of-period value of the firm. The production technology and choice of investment intensity,  $V_j(I_j; \theta_j)$  and  $I_j$ , are taken as given. If it is assumed that the end-of-period value of the firm is independent of its choice of financial liabilities,<sup>28/</sup> then  $V^j = V_j(I_j; \theta_j)$ , and  $\sum_1^q f_k \equiv V_j(I_j; \theta_j)$  for every outcome  $\theta_j$ .

Suppose when firm  $j$  is all equity financed, there exists an equilibrium such that the initial value of firm  $j$  is given by  $V_{j0}(I_j)$ .



Theorem VI.3 If firm  $j$  is financed by  $q$  different claims defined by the functions  $f_k(V_j^j)$ ,  $k = 1, \dots, q$  and if there exists an equilibrium such that the return distributions of the efficient portfolio set remains unchanged from the equilibrium in which firm  $j$  was all equity financed, then

$$\sum_1^q f_{ko} = V_{jo}(I_j)$$

where  $f_{ko}$  is the equilibrium initial value of financial claim  $k$ .

Proof: In the equilibrium in which firm  $j$  is all equity financed, the end-of period random variable value of firm  $j$  is  $V_j(I_j; \theta_j)$  and the initial value,  $V_{jo}(I_j)$ , is given by formula (IV.3) where  $(X_1, \dots, X_m, R)$  spans the efficient set. Consider now that firm  $j$  is financed by the  $q$  different claims. The random variable end-of-period value of firm  $j$ ,  $\sum_1^q f_k$ , is still given by  $V_j(I_j; \theta_j)$ . By hypothesis, there exists an equilibrium such that the distribution of the efficient portfolio set remains unchanged, and therefore the distribution of  $(X_1, \dots, X_m, R)$  remains unchanged. By inspection of formula (IV.3), the initial value of firm  $j$  will remain unchanged, and therefore,  $\sum_1^q f_{ko} = V_{jo}(I_j)$ .

Hence, for a given investment policy, the way in which the firm finances this investment will not affect the market value of the firm unless the choice of financial instruments changes the return distributions of the efficient portfolio set. Theorem VI.3 is representative of a class of theorems that describe the impact of financing policy on the market value of a firm when the investment decision is held fixed, and this class is generally referred to as "Modigliani-Miller Theorems" after the pioneering work in this direction by Modigliani and Miller.<sup>29/</sup>

Clearly, a sufficient condition for Theorem VI.3 to obtain is that each of the financial claims issued by the firm are "redundant securities" (as defined in Sections II and IV). This condition will be satisfied by the subclass of financial securities that provide for linear sharing rules. I.e.,  $f_k(V) = a_k V + b_k$  where  $\sum_1^q a_k = 1$  and  $\sum_1^q b_k = 0$ . They are redundant securities because the investor can replicate exactly the payoff structure of each claim by a portfolio combination of the (all-equity) firm and the riskless security. Hence, in this case, Theorem VI.3 will obtain as a special case of Proposition IV.1. Indeed, an example of this subclass is in the original Modigliani-Miller paper where they examined the effect on firm values of borrowing by firms under the assumption that borrowing (either by firms or individuals) is riskless and there is no bankruptcy.

It should be pointed out that the linear structure for firm borrowings only applies if there is no chance of default on the debt. Consider the case of a single homogeneous debt issue where the firm promises to pay B dollars at the end of the period and in the event the firm does not pay (i.e., defaults), then ownership of the firm is transferred to the debtholders. If the equity of the firm has limited liability, then the payoff function to the debt,  $f_1$ , can be written as

$$(VI.13) \quad f_1(V^j) = \text{Min}[V^j, B]$$

and the payoff function to the equity,  $f_2$ , can be written as

$$(VI.14) \quad f_2(V^j) = \text{Max}[0, V^j - B].$$

Hence,  $f_1$  and  $f_2$  will have a linear sharing-rule structure only if the probability that  $v^j < B$  is zero.

While the linearity of the sharing rules is sufficient, it is not a necessary condition for Theorem VI.3 to obtain as Stiglitz [87] has shown for the Arrow-Debreu and Capital Asset Pricing Models, and as will be demonstrated in Section VII.

## VII. Intertemporal Consumption and Portfolio Selection Theory

As with the preceeding analyses here, most papers on investment theory under uncertainty have assumed that individuals act so as to maximize the expected utility of end-of-period wealth and that intra-period revisions are not allowed. Therefore, all events which take place after next period are irrelevant to their decisions. Of course, individuals, and therefore firms, do care about events beyond "next period," and they can review their allocations periodically. Hence, the one-period, static analyses will only be valid under those conditions such that an intertemporally-maximizing individual acts, each period, as if he were a one-period, expected utility-of-wealth maximizer. In this section, the lifetime consumption-portfolio selection problem is solved, and conditions are derived under which the one-period static portfolio problem will be an appropriate "surrogate" for the dynamic, multi-period portfolio problem.

As in the early contributions by Hakansson [29], Leland [45], Samuelson [78] and Merton [53], the problem of choosing optimal portfolio and consumption rules for an individual who lives T years is formulated as follows. The individual chooses his consumption and portfolio allocation for each period so as to maximize<sup>30/</sup>

$$(VII.1) \quad E_0 \{ \sum_0^{T-1} U[C(t), t] + B[W(T), T] \}$$

where  $C(t)$  is consumption chosen at age  $t$ ;  $W(t)$  is wealth at age  $t$ ; " $E_t$ " is the conditional expectation operator conditional on knowing all relevant

information available as of time  $t$ ; the utility function (during life)  $U$  is assumed to be strictly concave in  $C$  and the "bequest" function  $B$  is also assumed to be concave in  $W$ .

It is assumed that there are  $n$  risky securities with random variable returns between time  $t$  and  $t+1$  denoted by  $Z_1(t+1), \dots, Z_n(t+1)$ , and there is a riskless security whose return between  $t$  and  $t+1$ ,  $R(t)$ , will be known with certainty as of time  $t$ .<sup>31/</sup> When the individual "arrives" at date  $t$ , he will know the value of his portfolio,  $W(t)$ . He chooses how much to consume,  $C(t)$ , and then reallocates the balance of his wealth,  $W(t) - C(t)$ , among the available securities. Hence, the accumulation equation between  $t$  and  $t+1$  can be written as<sup>32/</sup>

$$(VII.2) \quad W(t+1) = [\sum_1^n w_j(t) [Z_j(t+1) - R(t)] + R(t)] [W(t) - C(t)]$$

where  $w_j(t)$  is the fraction of his portfolio allocated to security  $j$  at date  $t$ ,  $j=1, \dots, n$ . Because the fraction allocated to the riskless security can always be chosen to equal  $1 - \sum_1^n w_j(t)$ , the choices for  $w_1(t), \dots, w_n(t)$  are unconstrained.

It is assumed that there exist  $m$  state variables,  $\{S_k(t)\}$ , such that the stochastic processes for  $\{Z_1(t+1), \dots, Z_n(t+1), S_1(t+1), \dots, S_m(t+1)\}$  are Markov with respect to  $S_1(t), \dots, S_m(t)$ , and  $S(t)$  will denote the  $m$ -vector of state variable values at time  $t$ .<sup>33/</sup>

The method of stochastic dynamic programming is used to derive the optimal consumption and portfolio rules. Define the function,  $J[W(t), S(t), t]$ , by

$$(VII.3) \quad J[W(t), S(t), t] \equiv \text{Max } E_t \{ \sum_t^{T-1} U[C(\tau), \tau] + B[W(T), T] \}.$$

$J$ , therefore, is the (utility) value of the balance of the individual's optimal consumption investment program from date  $t$  forward, and, in this context, is called the "derived" utility of wealth function. By the Principle of Optimality, (VII.3) can be rewritten as

$$(VII.4) \quad J[W(t), S(t), t] = \text{Max}\{U[C(t), t] + E_t\{J[W(t+1), S(t+1), t+1]\}\}$$

where "Max" is over the current decision variables  $[C(t), w_1(t), \dots, w_n(t)]$ . Substituting for  $W(t+1)$  in (VII.4) from (VII.2) and differentiating with respect to each of the decision variables, we can write the  $n+1$  first order conditions for a regular interior maximum as <sup>34/</sup>

$$(VII.5) \quad 0 = U_C[C^*(t), t] - E_t\{J_W[W(t+1), S(t+1), t+1](\sum_{j=1}^n w_j^*(Z_j - R) + R)\}$$

and

$$(VII.6) \quad 0 = E_t\{J_W[W(t+1), S(t+1), t+1](Z_j - R)\}, \quad j=1, 2, \dots, n$$

where  $U_C \equiv \partial U / \partial C$ ;  $J_W \equiv \partial J / \partial W$ ; and  $(C^*, w_j^*)$  are the optimum values for the decision variables. Henceforth, except where needed for clarity, the time indices will be dropped. Using (VII.6), (VII.5) can be rewritten as

$$(VII.7) \quad 0 = U_C[C^*, t] - R E_t\{J_W\}.$$

To solve for the complete optimal program, one first solves (VII.6) and (VII.7) for  $C^*$  and  $w^*$  as functions of  $W(t)$  and  $S(t)$  when  $t = T-1$ . This can be done because  $J[W(T), S(T), T] = B[W(T), T]$ , a given function.

Substituting the solutions for  $C^*(T-1)$  and  $w^*(T-1)$  in the right-hand side of (VII.4), (VII.4) becomes an equation and therefore one has  $J[W(T-1), S(T-1), T-1]$ . Using (VII.6), (VII.7), and (VII.4), one can proceed to solve for the optimal rules in earlier periods in the usual "backwards" recursive fashion of dynamic programming. Having done so, one will have a complete schedule of optimal consumption and portfolio rules for each date expressed as functions of the (then) known state variables  $W(t)$ ,  $S(t)$ , and  $t$ . Moreover, as Samuelson [ 78 ] has shown, the optimal consumption rules will satisfy the "envelope condition" expressed as

$$(VII.8) \quad J_W[W(t), S(t), t] = U_C[C^*(t), t].$$

I.e., at the optimum, the marginal utility of wealth (future consumption) will just equal the marginal utility of (current) consumption. Moreover, from (VII.8), it is straightforward to show that  $J_{WW} < 0$  because  $U_{CC} < 0$ . Hence,  $J$  is a strictly concave function of wealth.

A comparison of the first order conditions for the static portfolio selection problem, (II.4) in Section II, with the corresponding conditions (VII.6) for the dynamic problem will show that they are formally quite similar. Of course, they do differ in that, for the former case, the utility function of wealth is taken to be exogeneous while, in the latter, it is derived. However, the more fundamental difference in terms of derived portfolio selection behavior is that  $J$  is not only a function of  $W$  but also a function of  $S$ . The analogous condition in the static case would be that the end-of-period utility function of wealth is also state dependent.

To see that this difference is not trivial, consider the Rothschild-Stiglitz of riskier that was used in the one-period analysis to partition the feasible portfolio set into its efficient and inefficient parts. Let  $W_1$  and  $W_2$  be the random variable, end-of-period values of two portfolios with identical expected values. If  $W_2$  is equal in distribution to  $W_1 + Z$  where  $E(Z|W_1) = 0$ , then from (II.10) and (II.11),  $W_2$  is riskier than  $W_1$  and every risk averse maximizer of the expected utility of end-of-period wealth would prefer  $W_1$  to  $W_2$ . However, consider an intertemporal maximizer with a strictly concave, derived utility function  $J$ . It will not, in general be true that  $E_t\{J[W_1, S(t+1), t+1]\} > E_t\{J[W_2, S(t+1), t+1]\}$ . Therefore, although the intertemporal maximizer selects his portfolio for only one period at a time, the optimal portfolio selected may be one that would never be chosen by any risk-averse, one-period maximizer. Hence, the portfolio selection behavior of an intertemporal maximizer will, in general, be operationally distinguishable from the behavior of a static maximizer.

To adapt the Rothschild-Stiglitz definition to the intertemporal case, a stronger condition is required. Namely, if  $W_2$  is equal in distribution to  $W_1 + Z$  where  $E[Z|W_1, S(t+1)] = 0$ , then every risk-averse intertemporal maximizer would prefer to hold  $W_1$  rather than  $W_2$  in the period  $t$  to  $t+1$ . The proof follows immediately from the concavity of  $J$  and Jensen's Inequality. Namely,  $E_t\{J[W_2, S(t+1), t+1]\} = E_t\{E(J[W_2, S(t+1), t+1]|W_1, S(t+1))\}$ . By Jensen's Inequality,  $E(J[W_2, S(t+1), t+1]|W_1, S(t+1)) < J(E(W_2|W_1, S(t+1)), S(t+1), t+1) = J(W_1, S(t+1), t+1)$ , and, therefore,  $E_t\{J[W_2, S(t+1), t+1]\} < E_t\{J[W_1, S(t+1), t+1]\}$ . Hence, "noise" as denoted by  $Z$  must not only be noise relative to  $W_1$  but noise relative to the state variables  $S_1(t+1), \dots, S_m(t+1)$ . All the



analyses of the preceeding sections can be formally adapted to the intertemporal framework by simply requiring that the "noise" terms there,  $\epsilon$ , have the additional property that  $E_t(\epsilon|S(t+1)) = E_t(\epsilon) = 0$ . Hence, in the absence of further restrictions on the distributions, the resulting efficient portfolio set for intertemporal maximizers will be larger than in the static case.

However, under certain conditions,<sup>35/</sup> the portfolio selection behavior of intertemporal maximizers will be "as if" they were one-period maximizers. For example, if  $E_t[Z_j(t+1)] \equiv \bar{Z}_j(t+1) = E_t[Z_j(t+1)|S(t+1)]$ ,  $j=1, 2, \dots, n$ , then the additional requirement that  $E_t(\epsilon|S(t+1)) = 0$  will automatically be satisfied for any feasible portfolio, and the original Rothschild-Stiglitz "static" definition will be valid. Indeed, in the cited papers by Hakansson, Leland, Samuelson, and Merton, it is assumed that the security returns  $\{Z_1(t), \dots, Z_n(t)\}$  are serially independent and identically distributed in time which clearly satisfies this condition. Define the investment opportunity set at time t to be the joint distribution for  $\{Z_1(t+1), \dots, Z_n(t+1)\}$  and the return on the riskless security,  $R(t)$ . The Hakansson et al papers assume that the investment opportunity set is constant through time. The condition  $\bar{Z}_j(t+1) = E_t[Z_j(t+1)|S(t+1)]$ ,  $j=1, \dots, n$ , will also be satisfied if changes in the investment opportunity set are either completely random or time dependent in a nonstochastic fashion. Moreover, with the possible exception of a few perverse cases, these are the only conditions on the investment opportunity set under which  $\bar{Z}_j(t+1) = E_t[Z_j(t+1)|S(t+1)]$ ,  $j=1, \dots, n$ . Hence, for arbitrary concave utility functions, the one-period analysis will be a valid surrogate for the intertemporal analysis only if changes in the investment opportunity set satisfy these conditions.

Of course, by inspection of (VII.6), if  $J$  were of the form  $V[W(t),t] + H[S(t),t]$  so that  $J_W = V_W$  is only a function of wealth and time, then for arbitrary investment opportunity sets, such an intertemporal investor will act "as if" he is a one-period maximizer. Unfortunately, the only concave utility function that will produce such a  $J$  function and satisfy the additivity specification in (VII.1) is  $U[C,t] = a(t) \log[C]$  and  $B[W,T] = b(T) \log[W]$  where either  $a = 0$  and  $b > 0$  or  $a > 0$  and  $b \geq 0$ . While some have argued that this utility function is of special normative significance,<sup>36/</sup> any model whose results depend singularly upon all individuals having the same utility function where, in addition, the utility function must have a specific form, can only be viewed as an example, and not the basis for a theory.

Hence, in general, the one-period, static analysis will not be rich enough to describe the investment behavior of individuals in an intertemporal framework. Indeed, without additional assumptions, the only derived restrictions on optimal demand functions and equilibrium security returns are the ones that rule out arbitrage. Hence, to deduce additional properties, further assumptions about the dynamics of the investment opportunity set will be needed. However, before these assumptions are introduced, I make a brief digression.<sup>37/</sup>

There are three time intervals or horizons involved in the consumption-portfolio problem. First, there is the trading horizon which is the minimum length of time between which successive transactions by economic agents can be made in the market. In a sequence-of-markets analysis, it is the length of time between successive market openings, and is therefore part of the specification of the structure of markets in the economy. While this

structure will depend upon the tradeoff between the costs of operating the market and its benefits, this time scale is not determined by the individual investor, and is the same for all investors in the economy. Second, there is the decision horizon which is the length of time between which the investor makes successive decisions, and it is the minimum time between which he would take any action. For example, an investor with a fixed decision interval of one month, who makes a consumption decision and portfolio allocation today will under no conditions make any new decisions or take any action prior to one month from now. This time scale is determined by the costs to the individual of processing information and making decisions, and is chosen by the individual. Third, there is the planning horizon which is the maximum length of time for which the investor gives any weight in his utility function. Typically, this time scale would correspond to the balance of his lifetime and is denoted by "T" in the formulation (VII.1).

The static approach to portfolio selection implicitly assumes that the individual's decision and planning horizons are the same: "one period." While the intertemporal approach distinguishes between the two, when individual demands are aggregated to determine market equilibrium relationships, it is implicitly assumed in both approaches that the decision interval is the same for all investors, and therefore corresponds to the trading interval. If "h" denotes the length of time in the trading interval, then every solution derived has, as an implicit argument, h. Clearly, if one were to change h, then the derived behavior of investors would change, as indeed would any deduced equilibrium relationships.<sup>38/</sup> I might mention, somewhat

parenthetically, that empirical researchers almost uniformly neglect to recognize that  $h$  is part of a model's specification. For example, in Theorem IV.6, the returns on securities were shown to have a linear relationship to the returns on a set of spanning portfolios. However, because the  $n$ -period return on a security is the product (and not the sum) of the one-period returns, this linear relationship can only obtain for single time interval,  $h$ . If we define a fourth time interval, the observation horizon, to be the length of time between successive observations of the data by the researcher, then the usual practice is to implicitly assume that the decision and trading intervals are equal to the observation interval. This is done whether the observation interval is daily, weekly, monthly, or annually!

If the "Frictionless" markets assumption (A.1) is extended to include no costs of information processing or of operating the markets, then it follows that all investors would prefer to have " $h$ " as small as physically possible. Indeed, the aforementioned general assumption that all investors have the same decision interval will, in general, only be valid if all such costs are zero. This said, it is natural to consider the limiting case when  $h$  tends to zero, and trading takes place continuously in time.

Returning from this digression on time scales, consider an economy where the trading interval,  $h$ , is sufficiently small such that the state description of the economy can change only "locally" during the interval  $(t, t+h)$ . Formally, the Markov stochastic processes for the state variables,  $S(t)$ , are assumed to satisfy the property that one-step transitions are permitted only to the nearest neighboring states. The analogous condition in the limiting case of continuous time is that the sample paths for  $S(t)$

are continuous functions of time. I.e., for every realization of  $S(t+h)$  except possibly on a set of measure zero,  $\lim_{h \rightarrow 0} [S_k(t+h) - S_k(t)] = 0$ ,  $k=1, \dots, m$ . If, however, in the continuous limit, the uncertainty of "end-of-period" returns is to be preserved, then, an additional requirement is that  $\lim_{h \rightarrow 0} [S_k(t+h) - S_k(t)]/h$  exists almost nowhere. I.e., even though the sample paths are continuous, the increments to the states are not, and therefore, in particular, "end-of-period" rates of return will not be "predictable" even in the continuous time limit. The class of stochastic processes that satisfy these conditions are called diffusion processes.<sup>39/</sup>

Although such processes are almost nowhere differentiable in the usual sense, under some mild regularity conditions, there is a generalized theory of stochastic differential equations which allows their instantaneous dynamics to be expressed as the solution to the system of equations.<sup>40/</sup>

$$(VII.9) \quad dS_i(t) = G_i(S, t)dt + H_i(s, t)dq_i(t) \quad , \quad i=1, \dots, m,$$

where  $G_i(S, t)$  is the instantaneous expected change in  $S_i(t)$  per unit time at time  $t$ ;  $H_i^2$  is the instantaneous variance of the change in  $S_i(t)$  where it is understood that these statistics are conditional on  $S(t) = S$ . The  $dq_i(t)$  are Weiner processes with the instantaneous correlation coefficient per unit of time between  $dq_i(t)$  and  $dq_j(t)$  given by the function  $\eta_{ij}(S, t)$ ,  $i, j=1, \dots, m$ .<sup>41/</sup> Moreover, the functions  $\{G_i, H_i, \eta_{ij}, i, j=1, \dots, m\}$  completely describe the transition probabilities for  $S(t)$  between any two dates.<sup>42/</sup>

Under the assumption that the returns on securities can be described by diffusion processes, Merton [54] has solved the continuous-time analog to the discrete-time formulation in (VII.1). Namely,

$$(VII.10) \quad \text{Max } E_0 \left\{ \int_0^T U[C(t), t] dt + B[W(T), T] \right\}$$

Adapting the notation in that paper,<sup>43/</sup> the rate of return dynamics on the security  $j$  can be written as

$$(VII.11) \quad \frac{dP_j}{P_j} = \alpha_j(S, t) dt + \sigma_j(S, t) dZ_j, \quad j=1, \dots, n$$

where  $\alpha_j$  is the instantaneous conditional expected rate of return per unit time;  $\sigma_j^2$  is its instantaneous conditional variance per unit time; and  $dZ_j$  are Weiner processes with the instantaneous correlation coefficient per unit time between  $dZ_j(t)$  and  $dZ_k(t)$  given by the function  $\rho_{jk}(S, t)$ ,  $j, k=1, \dots, n$ . In addition to the  $n$  risky securities, there is a riskless security whose instantaneous rate of return per unit time is the interest rate  $r(t)$ .<sup>44/</sup> To complete the model's dynamics description, define the functions  $\mu_{ij}(S, t)$  to be the instantaneous correlation coefficients per unit time between  $dq_i(t)$  and  $dZ_j(t)$ ,  $i=1, \dots, m$ ;  $j=1, \dots, n$ .<sup>45/</sup>

If  $J$  is defined by

$$(VII.12) \quad J[W(t), S(t), t] \equiv \text{Max } E_t \left\{ \int_t^T U[C(\tau), \tau] d\tau + B[W(T), T] \right\},$$

then the continuous-time analog to (VII.4) can be written as<sup>46/</sup>

$$\begin{aligned}
 \text{(VII.13)} \quad 0 = & \text{Max}\{U[C,t] + J_t + J_W[(\sum_1^n w_j(\alpha_j - r) + r)W - C] + \sum_1^m J_i G_i + \\
 & \frac{1}{2} J_{WW} \sum_1^n \sum_1^n w_i w_j \sigma_{ij} W^2 + \frac{1}{2} \sum_1^m \sum_1^m J_{ij} H_i H_j \eta_{ij} + \\
 & \sum_1^m \sum_1^n J_{iW} w_j \sigma_j H_i \mu_{ij} W\}
 \end{aligned}$$

where the subscripts  $t, W, i$  on  $J$  denote partial derivatives with respect to the arguments  $t, W$ , and  $S_i$ , ( $i=1, \dots, m$ ) of  $J$ , respectively and  $\sigma_{ij} \equiv \sigma_i \sigma_j \rho_{ij}$  is the instantaneous covariance of the returns of security  $i$  with security  $j$ ,  $i, j=1, \dots, n$ . As was the case in (VII.4), the "Max" in (VII.13) is over the current decision variables  $[C(t), w_1(t), \dots, w_n(t)]$ . If  $C^*$  and  $w^*$  are the optimum rules, then the  $(n+1)$  first-order conditions for (VII.13) can be written as

$$\text{(VII.14)} \quad 0 = U_C[C^*, t] - J_W[W, S, t]$$

and

$$\text{(VII.15)} \quad 0 = J_W(\alpha_j - r) + J_{WW} \sum_1^n w_i^* \sigma_{ij} W + \sum_1^m J_{iW} \sigma_j H_i \mu_{ij}, \quad j=1, \dots, n.$$

(VII.14) is identical to the "envelope condition," (VII.8), in the discrete time analysis. However, unlike (VII.6) in the discrete time analysis, (VII.15) is a system of equations which is linear in the optimal demands for risky securities. Hence, if none of the risky securities is redundant, then (VII.15) can be solved explicitly for the optimal demand functions using standard matrix inversion. I.e.,

$$(VII.16) \quad w_j^*(t)W(t) = K \sum_1^n v_{kj}(\alpha_k - r) + \sum_1^m B_i \zeta_{ij}, \quad j=1, \dots, n$$

where  $v_{kj}$  is the  $k$ - $j$ <sup>th</sup> element of the inverse of the instantaneous variance-covariance matrix of returns  $[\sigma_{ij}]$ ;  $\zeta_{ij} \equiv \sum_1^n v_{kj} \sigma_k H_{1i} \mu_{1k}$ ;  $K \equiv -J_W/J_{WW}$ ; and  $B_i \equiv -J_{iW}/J_{WW}$ ,  $i=1, \dots, m$ .

As an immediate consequence of (VII.16), we have the following mutual fund theorem:

Theorem VII.1 If the returns dynamics are described by (VII.9) and (VII.11), then there exists  $(m+2)$  mutual funds constructed from linear combinations of the available securities such that, independent of preferences, wealth distribution, or planning horizon, individuals will be indifferent between choosing from a linear combination of these  $(m+2)$  funds or a linear combination of the original  $n$  risky securities and the riskless security.

Proof: Let mutual fund #1 be the riskless security; let mutual fund #2 hold fraction,  $\delta_j \equiv \sum_1^n v_{kj}(\alpha_k - r)$ , in security  $j$ ,  $j=1, \dots, n$ , and the balance  $(1 - \sum_1^n \delta_j)$  in the riskless security; let mutual fund  $\#(2+i)$  hold fraction,  $\delta_j^i \equiv \zeta_{ij}$ , in security  $j$ ,  $j=1, \dots, n$  and the balance  $(1 - \sum_1^n \delta_j^i)$  in the riskless security for  $i=1, \dots, m$ . Consider a portfolio of these mutual funds which allocates  $d_2(t) = K$  dollars to fund #2;  $d_{2+i}(t) = B_i$  dollars to fund  $\#(2+i)$ ,  $i=1, \dots, m$ ; and  $d_1(t) = W(t) - \sum_2^{2+m} d_i(t)$  dollars to fund #1. By inspection of (VII.16), this portfolio of funds exactly replicates the optimal portfolio holdings chosen from among the original  $n$



risky securities and the riskless security. However, the fractional holdings of these securities by the  $(m+2)$  funds do not depend upon the preferences, wealth, or planning horizon of the individuals investing in the funds. Hence, every investor can replicate his optimal portfolio by investing in the  $(m+2)$  funds.

Of course, as with the mutual fund theorems of Section IV, Theorem VII.1 is vacuous if  $m \geq n + 1$ . However, for  $m \ll n$ , the  $(m+2)$  portfolios provide for a nontrivial spanning of the efficient portfolio set, and it is straightforward to show that the instantaneous returns on individual securities will satisfy the same linear specification relative to these spanning portfolios as was derived in Theorem IV.6 for the one-period analysis.

It was shown in the discrete-time analysis that if  $E_t[Z_j(t+1)|S(t)] = \bar{Z}_j(t+1)$ ,  $j=1, \dots, n$ , then the intertemporal maximizer will act "as if" he were a static maximizer of the expected utility of end-of-period wealth. The corresponding condition in the continuous-time case translates into  $\mu_{ij} = 0$ ,  $i=1, \dots, m$  and  $j=1, \dots, n$ . Under these conditions, the optimal demand functions in (VII.16) can be rewritten as

$$(VII.17) \quad w_j^*(t)W(t) = K \sum_{k=1}^n v_{jk}(\alpha_k - r), \quad j=1, \dots, n.$$

From inspection of (VII.17), the relative holdings of risky securities,  $w_j^*(t)/w_1^*(t)$ , will be the same for all investors, and, hence, the efficient

portfolio set will be spanned by just two funds: a single risky fund and the riskless security. Moreover, by the procedure used to prove Theorem IV.9 and Theorem IV.10 in the static analysis, the efficient portfolio set here can be shown to be generated by the set of portfolios with minimum (instantaneous) variance for a given expected rate of return. Hence, under these conditions, the continuous-time intertemporal maximizer will act "as if" he were a static, Markowitz-Tobin mean-variance maximizer. Although the demand functions are formally identical to those derived from the mean-variance model, the continuous-time analysis is valid for any concave utility function and does not assume that security returns are normally distributed. Indeed, if, for example, the investment opportunity set  $\{\alpha_j, r, \sigma_{ij}, i, j=1, \dots, n\}$  is constant through time, then from (VII.11), the return on each risky security will be lognormally distributed which implies that all securities will have limited liability.<sup>47/</sup>

In the general case described in Theorem VII.1, the qualitative behavioral differences between an intertemporal maximizer and a static maximizer can be clarified further by analyzing the characteristics of the derived spanning portfolios.

As already shown, fund #1 and fund #2 are the "usual" portfolios that would be mixed to provide an optimal portfolio for a static maximizer. Hence, the intertemporal behavioral differences are characterized by funds #(2+i),  $i=1, \dots, m$ . At the level of demand functions, the "differential demand" for risky security  $j$ ,  $\Delta D_j^*$ , is defined to be the difference between the demand for that security by an intertemporal maximizer at time  $t$  and the demand for that security by a static maximizer of the expected utility of "end-of-period" wealth where the absolute risk aversion and current

wealth of the two maximizers are the same. Noting that  $K \equiv -J_W/J_{WW}$  is the reciprocal of the absolute risk aversion of the derived utility of wealth function, from (VII.16), we have that

$$(VII.18) \quad \Delta D_j^* = \sum_1^m B_1 \zeta_{1j} \quad , j=1, \dots, n.$$

Lemma VII.1 Define  $dY_1 \equiv dS_1 - \left( \sum_1^n \delta_j^+ \left( \frac{dP_j}{P_j} - rdt \right) + rdt \right)$ . The set of portfolio weights  $\{\delta_j^+\}$  that minimize the (instantaneous) variance of  $dY_1$  are given by  $\delta_j^+ = \zeta_{1j}$ ,  $j=1, \dots, n$  and  $i=1, \dots, m$ .

Proof: The instantaneous variance on  $dY_1$  is equal to  $[H_1^2 - 2 \sum_1^n \delta_j^+ H_1 \sigma_{1j} \mu_{1j} + \sum_1^n \sum_1^n \delta_j^+ \delta_k^+ \sigma_{jk}]$ . Hence, the minimizing set of  $\{\delta_j^+\}$  will satisfy  $0 = -H_1 \sigma_{1j} \mu_{1j} + \sum_1^n \delta_k^+ \sigma_{jk}$ ,  $j=1, \dots, n$ . By matrix inversion,  $\delta_j^+ = \zeta_{1j}$ .

The instantaneous rate of return on fund #(2+i) is exactly  $[rdt + \sum_1^n \zeta_{1j} \left( \frac{dP_j}{P_j} - rdt \right)]$ . Hence fund #(2+i) can be described as that feasible portfolio whose rate of return most closely replicates the instantaneous change in state variable  $S_i(t)$ , and this is true for  $i=1, \dots, m$ .

Consider the special case where there exist securities that are instantaneously perfectly correlated with changes in each of the state variables. Without loss of generality, assume that the first  $m$  securities are the securities such that  $dP_i/P_i$  is perfectly positively correlated with  $dS_i$ ,  $i=1, \dots, m$ . In this case,<sup>48/</sup> the demand functions (VII.16) can be rewritten in the form

$$\begin{aligned}
 \text{(VII.19)} \quad w_i^*(t)W(t) &= K \sum_1^n v_{ik} (\alpha_k - r) + B_i H_i / \sigma_i, \quad i=1, \dots, m \\
 &= K \sum_1^n v_{ik} (\alpha_k - r), \quad i=m+1, \dots, n.
 \end{aligned}$$

Hence, the relative holdings of securities  $m+1$  through  $n$  will be the same for all investors, and the differential demand functions can be rewritten as

$$\begin{aligned}
 \text{(VII.20)} \quad \Delta D_i^* &= B_i H_i / \sigma_i \quad i=1, \dots, m \\
 &= 0 \quad i=m+1, \dots, n.
 \end{aligned}$$

The composition of fund  $\#(2+i)$  reduces to a simple combination of security  $i$  and the riskless security.

The behavior implied by the demand functions in (VII.16) can be more easily interpreted if they are rewritten in terms of the direct utility and optimal consumption functions. The optimal consumption function has the form  $C^*(t) = C^*(W, S, t)$ , and from (VII.14), it follows immediately that

$$\text{(VII.21)} \quad K = -U_C[C^*, t] / (U_{CC}[C^*, t] \partial C^* / \partial W)$$

$$B_i = -(\partial C^* / \partial S_i) / (\partial C^* / \partial W), \quad i=1, \dots, m.$$

Because  $\partial C^* / \partial W > 0$ , it follows that the sign of  $B_i$  equals the sign of  $(-\partial C^* / \partial S_i)$ . An unanticipated change in a state variable is said to be unfavorable if, ceteris paribus, such a change would reduce current optimal

consumption. E.g., an unanticipated increase in  $S_i$  would be unfavorable if  $\partial C^*/\partial S_i < 0$ . Inspection of (VII.20), for example, shows that for such an individual, the differential demand for security  $i$  (which is perfectly positively correlated with changes in  $S_i$ ) will be positive. If there is an unanticipated increase in  $S_i$ , then, ceteris paribus, there will be an unanticipated increase in his wealth. Because  $\partial C^*/\partial W > 0$ , this increase in wealth will tend to offset the negative impact on  $C^*$  caused by the increase in  $S_i$ , and therefore, the unanticipated variation in  $C^*$  will be reduced. In effect, by holding more of this security, he expects to be "compensated" by larger wealth in the event that  $S_i$  changes in the unfavorable direction. Of course, if  $\partial C^*/\partial S_i > 0$ , then he will take a differentially short position. However, in all cases, investors will allocate their wealth to the funds  $\#(2+i)$ ,  $i=1, \dots, m$ , so as to "hedge" against unfavorable changes in the state variables  $S(t)$ .<sup>49/</sup>

In the usual static model, such hedging behavior will not be observed because the utility function depends only upon end-of-period wealth, and therefore it is implicitly assumed that  $\partial C^*/\partial S_i = 0$ ,  $i=1, \dots, m$ . Thus, in the intertemporal model, securities have, in addition to their manifest function of providing an "efficient" risk-return tradeoff for end-of-period wealth, a latent function of allowing consumers to "hedge" against other uncertainties.<sup>50/</sup>

It was shown in Section IV that a necessary and sufficient condition for a set of portfolios to span the efficient portfolio set is that the returns on every security can be written as a linear function of the returns on the spanning portfolios plus noise. Moreover, for the spanning property

to have operational significance, the size of the minimum spanning set,  $M^*$ , must be very much smaller than the number of available securities,  $n+1$ . This linearity requirement would appear to virtually rule out nontrivial spanning unless the markets are already complete in the Arrow-Debreu sense. To see this consider the following: suppose there exists a nontrivial spanning set  $(X_1, \dots, X_m, R)$  and the return on all-equity financed firm  $j$  satisfies the linearity condition. Suppose firm  $j$  changes its liabilities structure by issuing debt and retiring some equity where the end-of-period values of the two claims are given by (VI.13) and (VI.14) in Section VI. Although the payoffs to these claims are perfectly functionally related to the end-of-period value of the firm, this functional relationship is non-linear if there is a positive probability of default. Hence, the returns on either of these claims will not, in general, be expressible as a linear function of the returns on the spanning portfolios plus noise. Therefore, upon such a change in firm  $j$ 's liability structure, the size of the spanning set would, in general, have to increase. In effect, because the payoff to these liabilities cannot be replicated by a portfolio combination of the previously-existing securities, there are no "perfect substitutes" for the created liabilities among these securities. While "perfect substitutability" is not a necessary condition for the spanning set to remain unchanged, without it, it is unlikely that this set will remain unchanged. Because firms can and do issue many different liabilities, it is, therefore, unlikely that the size of the minimum spanning set will differ significantly from the number of securities outstanding in incomplete market models with arbitrary concave utility functions.

While the issuing of nonlinear liabilities has this effect in the static and discrete-time dynamic models, it does not for the continuous-time model described here. To illustrate, consider the following example which uses a type of analysis first presented by Black and Scholes [5] in the context of option pricing. Let  $V$  be the value of some security whose return dynamics are described by the diffusion process.

$$(VII.23) \quad \frac{dV}{V} = \alpha dt + \sigma dZ$$

where  $\alpha$  and  $\sigma$  are, at most, functions of  $V$  and  $t$  so that the process is Markov. Let  $W$  be the price of some security whose value as of a specified date in the future,  $T^*$ , is given by the function  $g[V(T^*)]$  where  $g[0] = 0$  and  $g[V(T^*)]/V(T^*)$  is bounded.

Let  $f(V,t)$  be the solution to the partial differential equation<sup>51/</sup>

$$(VII.24) \quad \frac{1}{2} \sigma^2 V^2 f_{VV} + rVf_V - rf + f_t = 0$$

subject to the boundary conditions

$$(VII.25) \quad \begin{aligned} (a) \quad & f[0,t] = 0 \\ (b) \quad & f[V,T^*] = g[V] \\ (c) \quad & f/V \text{ bounded,} \end{aligned}$$

where in this example the interest rate  $r$  is taken to be constant over time and subscripts denote partial derivatives.

Consider the continuous-time portfolio strategy where the investor allocates the fraction  $w(t)$  to the security with value  $V(t)$  and  $[1 - w(t)]$  to the riskless security. If  $w(t)$  is a right-continuous function and  $P(t)$  denotes the value of the portfolio at time  $t$ , then the dynamics for  $P$  must satisfy

$$(VII.26) \quad dP = [w(\alpha - r) + r]Pdt + w\sigma P dZ.$$

Suppose the particular portfolio strategy chosen is  $w(t) = f_V[V, t]V(t)/P(t)$ . Substituting this strategy into (VII.26), we have that

$$(VII.27) \quad dP = [f_V V(\alpha - r) + rP]dt + f_V \sigma V dZ.$$

Since  $f$  is twice-continuously differentiable, Itô's Lemma<sup>52/</sup> can be used to express the stochastic process for  $f$  as

$$(VII.28) \quad df = \left[ \frac{1}{2} \sigma^2 V^2 f_{VV} + \alpha V f_V + f_t \right] dt + f_V \sigma V dZ.$$

But  $f$  satisfies (VII.24), and therefore (VII.28) can be rewritten as

$$(VII.29) \quad df = [f_V V(\alpha - r) + rf]dt + f_V \sigma V dZ.$$

Comparing (VII.29) with (VII.27), if the initial investment in the portfolio,  $P(0)$ , is chosen such that  $P(0) = f[V(0), 0]$ , then  $P(t) = f[V, t]$  for all  $V$  and  $t \leq T^*$ . But  $P(t)$  is a feasible portfolio strategy with the properties that  $P(t) = 0$  if  $V(t) = 0$  and  $P(T^*) = g[V(T^*)]$ . Hence, if the



value of the other security,  $W(t)$ , does not equal  $P(t)$  for each  $t$ , then there will exist an arbitrage opportunity. Therefore,  $W(t) = f[V(t), t]$ . Since  $g$  can be a nonlinear function, it has been shown that there exists a dynamic portfolio strategy that will exactly replicate the payoff to a nonlinear function of the price of the security. The application of this procedure to the debt-equity case of (VI.13) and (VI.14) is immediate. If  $V$  denotes the value of the firm, then the debt issue will satisfy (VII.24) with  $g[V] = \min[V, B]$  and the equity will satisfy the same equation but with  $g[V] = \max[0, V - B]$ . In addition, this type of analysis can be used to show that the Modigliani-Miller Theorem (Theorem VI.3) will obtain for nonlinear sharing rules.<sup>53/</sup> Finally, inspection of (VII.24) shows that knowledge of the expected return on  $V$ ,  $\alpha$ , is not required to solve for the values of the various claims, and indeed the only non-observable input required is the variance rate,  $\sigma^2$ . Moreover, (VII.24) often yields closed-form solutions and in those cases where it does not, there exist efficient numerical solution techniques. Hence, this type of analysis has led to a copious theoretical and empirical literature on the pricing of corporate liabilities and contingent claims generally.<sup>54/</sup>

In summary, the combined assumptions of continuous trading and local state variable changes substantially simplifies the analysis required in the discrete-time case. Under these conditions, the crucial non-trivial spanning properties will obtain, and the creation of firm-contingent liabilities will not, in general, affect the equilibrium. Although continuous-trading is only a theoretical proposition, it has been shown that the continuous-trading solutions will be an asymptotically valid approximation

to the discrete time solution as the trading interval  $h$  becomes small.<sup>55/</sup> Indeed, actual securities markets are open virtually every day, and hence the assumption that  $h$  is small is not without empirical foundation. In the intertemporal version of the Arrow-Debreu model with complete markets, it is well known that the market need only be open "once" because individuals will have no need for further trade. The continuous trading model is, of course, the opposite extreme. However, the continuous-time model appears to have many of the properties of the Arrow-Debreu model without nearly so many securities. Hence, it may be that a good substitute for having so many markets and securities is to have fewer markets and securities but the existing markets open more frequently. However, the study of this possibility will be left as a topic for future research.

## FOOTNOTES

- \* This paper is not a survey of the economics and finance literature on investment theory which is already copious and whose rate of expansion has in recent years accelerated. Rather the paper is designed to be a self-contained, but cryptic, introduction to some of the major problems in investment theory. In many cases, I have referenced excellent survey articles which themselves contain extensive bibliographies rather than attempt to reference all the important individual contributions to the subject. I gratefully acknowledge financial support from the National Science Foundation for this paper.
1. An important issue within this topic is the "information efficiency" of the stock market as derived in Fama [13, 14] and Samuelson [76, 81]. Grossman [25] and Grossman and Stiglitz [26] provide some additional insights. Hirshleifer [36] has shown that, in the absence of production, private sector expenditures on information gathering may be "socially wasteful." Moreover, significant further advances in general equilibrium theory will require explicit recognition of information production and its dissemination, and the development of such an integrated theory is just at its beginning.
  2. von Neumann and Morgenstern [91]. For an axiomatic description, see Herstein and Milnor [34]. Although the original axioms require that  $U$  be bounded, the continuity axiom can be extended to allow for unbounded functions. See Samuelson [82] for a discussion of this and the St. Petersburg Paradox.
  3. The strict concavity assumption implies that investors are everywhere risk-averse. While strictly convex or linear utility functions on the entire range imply behavior that is grossly at variance with observed behavior, the strict concavity assumption also rules out Friedman-Savage type utility functions whose behavioral implications are reasonable. The strict concavity also implies  $U'(W) > 0$  which rules out individual satiation.
  4. The "borrowing rate" is the rate on riskless-in-terms-of-default loans. While virtually every individual loan involves some chance of default, the empirical "spread" in the rate on margin loans to individuals suggests that this assumption is not a "bad approximation" for portfolio selection analysis. An explicit analysis of risky loan evaluation is provided in Section VII.
  5. For a trivial example, shares of General Motors with odd serial numbers are technically different from shares of GM with even serial numbers, and are, therefore, technically different securities. However, because their returns are identical, they are perfect substitutes from

the point of view of investors. In portfolio theory, securities are operationally defined by their return distributions, and therefore two securities with identical returns are indistinguishable.

6. If  $U$  is such that  $U'(0) = \infty$ , and by extension,  $U'(W) = \infty$ ,  $W < 0$ , then from (II.2) or (II.4), it is easy to show that the probability of  $Z^* \leq 0$  is a set of measure zero.
7. Kuhn and Tucker [42]. The analysis of the "no short-sales" case is complicated and leads to virtually no theorems. Although in actual markets there are some restrictions on short sales, these restrictions may not be too important because limited liability securities (e.g. put options) can and have been created that provide essentially the same type of return as a short sale.
8. The behavior associated with the utility function  $V(W) \equiv aU(W) + b$ ,  $a > 0$ , is indistinguishable from the behavior associated with  $U(W)$ .
9. Pratt [67]. Note:  $A(W)$  is invariant to any positive affine transformation of  $U(W)$ .
10. Rothschild and Stiglitz [73] and [74]. There is an extensive literature not discussed here that uses this type of risk measure to determine when one portfolio "stochastically dominates" another. Cf. Hadar and Russell [27, 28] and Bawa [3].
11. I believe that Christian von Weizsäcker proved a similar theorem in unpublished notes some years ago. However, I do not have a reference.
12. Because the iso-elastic family is not well defined for  $W < 0$ , portfolios with a positive probability of negative end-of-period wealth cannot be ranked.
13. Rothschild and Stiglitz [74, pp. 70-74].
14. For a proof, see Theorem 236 in Hardy, Littlewood, and Pólya [30].
15. A sufficient amount of information would be the joint distribution of  $Z^*$  and  $\epsilon_j$ . The necessary amount of information will depend upon the functional form of  $U'$ . However, in no case will a necessary condition be knowledge of  $Z_j$ .
16. Cf. King [41], Livingston [48], Farrar [17], Feeney and Hester [19], Farrell [18]. Unlike standard "factor analysis," the number of common factors here does not depend upon the fraction of total variation in an individual security's return that can be "explained." Rather what is important is the number of factors necessary to "explain" the covariation between pairs of individual securities.

17. For this family of utility functions, the probability distribution for securities cannot be completely arbitrary without violating the von Neumann-Morgenstern axioms. For example, it is required that for every realization of  $W$ ,  $W > -a/b$  for  $b > 0$  and  $W < -a/b$  for  $b < 0$ . The latter condition is especially restrictive.
18. Many authors have studied the properties of this family. See Merton [54, p389] for references.
19. As discussed in footnote 17, the range of values for  $a_i$  cannot be arbitrary for a given  $b$ . Moreover, the sign of  $b$  uniquely determines the sign of  $A'(W)$ .
20. Cass and Stiglitz [7, p. 144].
21. Cf. Ross [72] for spanning proofs in the absence of a riskless security. Black [4] and Merton [55] derive the two-fund theorem for the mean-variance model with no riskless security.
22. For the Arrow-Debreu model, see Hirshleifer [33, 34, 35], Myers [66], and Radner [68]. For the Mean-Variance model, see Jensen [38, 39] and Sharpe [85].
23. If the states are defined in terms of end-of-period values of the firm in addition to "environmental" factors, then the firms' production decisions will, in general, alter the state space description which violates the assumptions of the model. Moreover, I see no obvious reason why individuals are any more likely to agree upon the  $\{V_i(i)\}$  function than upon the probability distributions for the environmental factors. If sufficient information is available to partition the states into fine enough categories to produce agreement on the  $\{V_i(i)\}$  functions, then, given this information, it is difficult to imagine how rational individuals would have heterogeneous beliefs about the probability distributions for these states. As with the standard certainty model, agreement on the technologies is necessary for pareto optimality in this model. However, as Peter Diamond has pointed out to me, it is not sufficient. Sufficiency demands the stronger requirement that everyone be "right" in their assessment of the technologies.
24. In particular, the optimal portfolio demand functions are of the form derived in the proof of Theorem IV.9. For a complete analytic derivation, see Merton [55].
25. Sharpe [84], Lintner [47], and Mossin [65] are generally credited with independent derivations of the model. Black [4] extended the model to include the case of no riskless security.

26. Cf. Borch [ 6], Feldstein [20], Tobin [90], and Samuelson [77].
27. Radner [68, p. 460].
28. By this assumption, I have formally ruled out financial securities that alter the tax liabilities of the firm (e.g., interest deductions) or ones that can induce "outside" costs (e.g., bankruptcy costs). However, by redefining  $V_j(I_j; \theta_j)$  as the pre-tax and bankruptcy value of the firm and letting one of the  $f_k$  represent the government's (tax) claim and another the lawyers' (bankruptcy) claim, then the analysis in the text will be valid for these extended securities as well.
29. Miller and Modigliani [64]. See also Stiglitz [87] and Miller [63].
30. The additivity of the utility function and the single-consumption-good assumptions are made for analytical simplicity and because the principle topic of this paper is investment allocation and not the individual consumption choice. Fama [15] analyzes the problem for non-additive utilities. Although  $T$  is treated as known in the text, the analysis is essentially the same for an uncertain lifetime with  $T$  a random variable. Cf. Richard [70] and Merton [54].
31. This definition of a riskless security is purely technical and without normative significance. For example, investing solely in the riskless security will not allow for a certain consumption stream because  $R(t)$  will vary stochastically over time. On the other hand, a  $T$ -period, riskless-in-terms-of-default coupon bond which allows for a certain consumption stream is not a riskless security because its one-period return is uncertain. For further discussion, see Merton [57].
32. It is assumed that all income comes from investment in securities. The analysis would be the same with wage income provided that investors can sell shares against future income. However, because institutionally this cannot be done, the "nonmarketability" of wage income will cause systematic effects on the portfolio and consumption decisions.
33. Many non-Markov stochastic processes can be transformed to fit the Markov format by expanding the number of state variables. Cf. Cox and Miller [ 8, pp. 16-18]. To avoid including "surplus" state variables, it is assumed that  $\{S(t)\}$  represent the minimum number of variables necessary to make  $\{Z_j(t+1)\}$  Markov.

34. Sufficient conditions for existence and uniqueness of the solutions are: (1) strict concavity of  $U$  and  $B$ ; (2) no redundant securities; (3) no arbitrage opportunities. Cf. Dreyfus [11] for the dynamic programming technique.
35. See Fama [15] for a general discussion of these conditions.
36. See Latané [44], Markowitz [50], and Rubinstein [75] for arguments in favor of this view, and Samuelson [80], Goldman [24], and Merton and Samuelson [6] for arguments in opposition to this view.
37. This digression is adapted from Merton [58, pp. 662-663].

38. If investor behavior were invariant to  $h$ , then investors would choose the same portfolio if they were "frozen" into their investments for ten years as they would if they could revise their portfolios every day.

39. See Feller [21]; Itô and McKean [37], and Cox and Miller [8].

40. (VII.9) is a short-hand expression for the stochastic integral

$$S_i(t) = S_i(0) + \int_0^t G_i(S, \tau) d\tau + \int_0^t H_i(S, \tau) dq_i$$

where  $S_i(t)$  is the solution to (VII.9) with probability one. For a general discussion and proofs, see Arnold [1], Itô and McKean [37], McKean [51], and McShane [52].

41.  $\int_0^t dq_i = q_i(t) - q_i(0)$  will be normally distributed with a zero mean and variance equal to  $t$ .
42. See Feller [21, pp. 320-321]; Cox and Miller [8, p. 215]. The transition probabilities will satisfy the Kolmogorov or Fokker-Planck partial differential equations.
43. Merton [54, p. 377].  $dP_i/P_i$  in continuous time corresponds to  $Z_j(t+1) - 1$  in the discrete-time analysis.
44.  $r(t)$  corresponds to  $R(t) - 1$  in the discrete-time analysis, and is the "force of interest," continuous rate. While the rate earned between  $t$  and  $(t+dt)$ ,  $r(t)$ , is known with certainty as of time  $t$ ,  $r(t)$  can vary stochastically over time.
45. Unlike in the Arrow-Debreu model, for example, it is not assumed here that the returns are necessarily completely described by the changes in the state variables,  $dS_i$ ,  $i=1, \dots, m$ . I.e., the  $dZ_j$  need not be instantaneously perfectly correlated with some linear combination of  $dq_1, \dots, dq_m$ . Rather it is only assumed that  $(dP_1/P_1, \dots, dP_n/P_n, dS_1, \dots, dS_m)$  is Markov in  $S(t)$ .

46. See Merton [54, p. 381] and Kushner [43, Chapter IV, Theorem 7].
47. See Merton [54, p. 384-488]. It is also shown there that the return will be lognormal on the risky fund which, together with the riskless security, spans the efficient portfolio set.
48. This case is similar in spirit to the Arrow-Debreu complete markets model.
49. This behavior will obtain even when the return on fund  $\#(2+i)$  is not instantaneously perfectly correlated with  $dS_i$ .
50. For further discussion of this analysis, descriptions of specific sources of uncertainty, and extensions to discrete-time examples, see Merton [57, 58, 59]. In the case of multiple consumption goods with uncertain relative prices, similar behavior obtains. However  $C^*$  is a vector and  $J_W$  is the "shadow" price of the "composite" consumption bundle. Hence, the corresponding derived "hedging" behavior is to minimize the unanticipated variations in  $J_W$ .
51. (VII.24) is a linear partial differential equation of the parabolic type. If  $\sigma^2$  is a continuous function and  $g$  is piece-wise continuous, then there exists a unique solution that satisfies boundary conditions (VII.25). The usual method for solving this equation is Fourier transforms.
52. Itô's Lemma is for stochastic differentiation the analog to the Fundamental Theorem of the calculus for deterministic differentiation. For a statement of the Lemma and applications in economics, see Merton [54, 56]. For its rigorous proof, see McKean [51, p. 44].
53. For a proof and extensions to general contingent claim securities, see Merton [60].
54. The literature based on the Black-Scholes type analysis has grown so rapidly that no attempt has been made here to provide a complete set of references. An excellent survey article is Smith [86]. In addition to the pricing of corporate liabilities, it has been applied to the pricing of loan guarantees, insurance contracts, "dual" funds, term-structure bonds, and even tenure for university professors.
55. See Samuelson [79] and Merton and Samuelson [61]. In Merton [58, p. 663] there is a brief discussion of the special cases in which the limiting discrete-time solutions do not approach the continuous-time solutions.



## BIBLIOGRAPHY

1. Arnold, L., Stochastic Differential Equations: Theory and Applications. New York: John Wiley (1974).
2. Arrow, K.J., "The Role of Securities in the Optimal Allocation of Risk Bearing," Review of Economic Studies, 31 (1964), pp. 91-96.
3. Bawa, V. S., "Optimal Rules for Ordering Uncertain Prospects," Journal of Financial Economics, 2, (1975), pp. 95-121.
4. Black, F., "Capital Market Equilibrium with Restricted Borrowing," Journal of Business, 45 (1972), pp. 444-455.
5. \_\_\_\_\_ and M. Scholes., "The Pricing of Options and Corporate Liabilities," Journal of Political Economy 81, (1973), pp. 637-654.
6. Borch, K., "A Note on Uncertainty and Indifference Curves," Review of Economic Studies, 36 (1969), pp. 1-4.
7. Cass, D. and J. E. Stiglitz, "The Structure of Investor Preferences and Asset Returns, and Separability in Portfolio Allocation: A Contribution to the Pure Theory of Mutual Funds," Journal of Economic Theory, 2 (1970), pp. 122-160.
8. Cox, D.A. and H. D. Miller, The Theory of Stochastic Processes. New York: John Wiley (1968).
9. Debreu, G., Theory of Value. New York: John Wiley (1959).
10. Diamond, P., "The Role of a Stock Market in a General Equilibrium Model with Technological Uncertainty," American Economic Review, 57 (1967), pp. 759-776.
11. Dreyfus, S. E., Dynamic Programming and the Calculus of Variations. New York: Academic Press (1965).
12. Eckern, S. and R. Wilson, "On the Theory of the Firm in an Economy with Incomplete Markets," Bell Journal of Economics and Management Science, 5 (1974), pp. 171-180.
13. Fama, E., "The Behavior of Stock Market Prices," Journal of Business 38 (1965), pp. 34-105.
14. \_\_\_\_\_, "Efficient Capital Markets: A Review of Theory and Empirical Work," Journal of Finance, 25 (1970), pp. 383-417.

15. \_\_\_\_\_, "Multi-period Consumption-Investment Decisions," American Economic Review, 60 (1970), pp. 163-174.
16. \_\_\_\_\_, "Perfect Competition and Optimal Production Decisions under Uncertainty," Bell Journal of Economics and Management Science, 3 (1972) pp. 509-530
17. Farrar, D. E., The Investment Decision Under Uncertainty, Englewood Cliffs, New Jersey: Prentice Hall (1962).
18. Farrell, J. L., "Analyzing Covariation of Returns to Determine Homogeneous Stock Groupings," Journal of Business, 47 (1974), pp. 186-207.
19. Feeney, G. J., and D. Hester, "Stock Market Indices: A Principal Components Analysis," in D. Hester and J. Tobin, (eds.), Risk Aversion and Portfolio Choice. New York: John Wiley (1967).
20. Feldstein, M. S., "Mean-Variance Analysis in the Theory of Liquidity Preference and Portfolio Selection," Review of Economic Studies, 36 (1969), pp. 5-12.
21. Feller, W., An Introduction to Probability Theory and Its Applications, Volume 2. New York: John Wiley (1966).
22. Fischer, S., "Assets, Contingent Commodities, and the Slutsky Equations," Econometrica, 40 (1972), pp. 371-385.
23. Friend, I and J. Bicksler, (eds.). Studies in Risk and Return, Volumes I & II. Cambridge, Mass: Ballinger (1975).
24. Goldman, M. B., "A Negative Report on the 'Near-Optimality' of the Max-Expected-Log Policy as Applied to Bounded Utilities for Long-Lived Programs," Journal of Financial Economics, 1 (1974), pp. 97-103.
25. Grossman, S., "On the Efficiency of Competitive Stock Markets where Traders have Diverse Information," Journal of Finance, 31 (1976), pp. 573-585.
26. \_\_\_\_\_ and J. E. Stiglitz, "Information and Competitive Price Systems," American Economic Review, 66 (1976), pp. 246-253.
27. Hadar, J. and W. R. Russell, "Rules for Ordering Uncertain Prospects," American Economic Review, 59 (1969), pp. 25-34.
28. \_\_\_\_\_, "Stochastic Dominance and Diversification," Journal of Economic Theory, 3 (1971), pp. 288-305.

29. Hakansson, N., "Optimal Investment and Consumption Strategies under Risk for a Class of Utility Functions," Econometrica, 38 (1970), pp. 587-607.
30. Hardy, G. H., J. E. Littlewood, and G. Pólya., Inequalities. Cambridge: The University Press (1959).
31. Hart, O., "On the Optimality of Equilibrium When the Market Structure is Incomplete," Journal of Economic Theory, 7 (1975)
32. Herstein, I., and J. Milnor., "An Axiomatic Approach to Measurable Utility," Econometrica, 21 (1953).
33. Hirshleifer, J., "Investment Decision under Uncertainty: Choice-Theoretic Approaches," Quarterly Journal of Economics 79 (1965), pp. 509-536.
34. \_\_\_\_\_, "Investment Decisions Under Uncertainty: Applications of the State-Preference Approach," Quarterly Journal of Economics, 80 (1966), pp. 252-277.
35. \_\_\_\_\_, Investment, Interest and Capital. Englewood Cliffs, New Jersey: Prentice-Hall (1970).
36. \_\_\_\_\_, "Where Are We in the Theory of Information," American Economic Review, 63 (1973), pp. 31-39.
37. Ito, K and H. P. McKean, Jr., Diffusion Processes and their Sample Paths, New York: Academic Press (1964).
38. Jensen, M. C. (ed.), Studies in the Theory of Capital Markets. New York: Praeger Publishers (1972).
39. \_\_\_\_\_, "Capital Markets: Theory and Evidence," Bell Journal of Economics and Management Science, 3 (1972), pp. 357-398.
40. \_\_\_\_\_ and J. B. Long, Jr., "Corporate Investment Under Uncertainty and Pareto Optimality in the Capital Markets," Bell Journal of Economics and Management Science, 3 (1972), pp. 151-174.
41. King, B. R., "Market and Industry Factors in Stock Price Behavior," Journal of Business, 39 (1966 Supplement), pp. 139-190.
42. Kuhn, H. W. and A. W. Tucker., "Nonlinear Programming," in U. Neyman (ed.), Proceedings of the Second Berkeley Symposium of Mathematical Statistics and Probability. Berkeley: University of California Press (1951).

43. Kushner, H. J., Stochastic Stability and Control. New York: Academic Press (1967).
44. Latane, H., "Criteria for Choice Among Risky Ventures," Journal of Political Economy, 67 (1959), pp. 144-155.
45. Leland, H., "Dynamic Portfolio Theory," Ph.D. Thesis: Cambridge, Mass, Harvard University (1968).
46. \_\_\_\_\_, "Production Theory and the Stock Market," Bell Journal of Economics and Management Science, 5 (1974), pp. 125-144.
47. Lintner, J., "The Valuation of Risk Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets," Review of Economics and Statistics, 47 (1965), pp. 13-37.
48. Livingston, M., "Industry Movements of Common Stocks," Journal of Finance, 32 (1977), pp. 861-874.
49. Markowitz, H., Portfolio Selection: Efficient Diversification of Investment. New York: John Wiley (1959).
50. \_\_\_\_\_, "Investment for the Long Run: New Evidence for an Old Rule," Journal of Finance, 31 (1976), pp. 1273-1286.
51. McKean, Jr., H. P., Stochastic Integrals. New York: Academic Press (1969).
52. McShane, E. J., Stochastic Calculus and Stochastic Models. New York: Academic Press (1974).
53. Merton, R. C., "Lifetime Portfolio Selection under Uncertainty: The Continuous-Time Case," Review of Economics and Statistics, 51 (1969), pp. 247-257.
54. \_\_\_\_\_, "Optimum Consumption and Portfolio Rules in a Continuous Time Model," Journal of Economic Theory, 3 (1971), pp. 373-413.
55. \_\_\_\_\_, "An Analytic Derivation of the Efficient Portfolio Frontier," Journal of Financial and Quantitative Analysis, 7 (1972), pp. 1851-1872.
56. \_\_\_\_\_, "Theory of Rational Option Pricing," Bell Journal of Economics and Management Science, 4 (1973), pp. 141-183.
57. \_\_\_\_\_, "An Intertemporal Capital Asset Pricing Model," Econometrica, 41 (1973), pp. 867-887.
58. \_\_\_\_\_, "Theory of Finance from the Perspective of Continuous Time," Journal of Financial and Quantitative Analysis, 10 (1975), pp. 659-674.

74. \_\_\_\_\_, "Increasing Risk II: Its Economic Consequences," Journal of Economic Theory, 3 (1971), pp. 66-84.
75. Rubinstein, M., "The Strong Case for the Generalized Logarithmic Utility Model as the Premier Model of Financial Markets," Journal of Finance, 31 (1976), pp. 551-572.
76. Samuelson, P. A., "Proof that Properly Anticipated Prices Fluctuate Randomly," Industrial Management Review, 6 (1965), pp. 41-49. Reprinted in [83].
77. \_\_\_\_\_, "General Proof That Diversification Pays," Journal of Financial and Quantitative Analysis, 2 (1967), pp. 1-13. Reprinted in [83].
78. \_\_\_\_\_, "Lifetime Portfolio Selection by Dynamic Stochastic Programming," Review of Economics and Statistics, 51 (1969), pp. 239-246. Reprinted in [83].
79. \_\_\_\_\_, "The Fundamental Approximation Theorem of Portfolio Analysis in Terms of Means, Variances, and Higher Moments," Review of Economic Studies, 37 (1970), pp. 537-542. Reprinted in [83].
80. \_\_\_\_\_, "The 'Fallacy' of Maximizing the Geometric Mean in Long Sequences of Investing or Gambling," Proceedings of the National Academy of Sciences, 68 (1971), pp. 2493-2496. Reprinted in [83].
81. \_\_\_\_\_, "Proof that Properly Discounted Present Values of Assets Vibrate Randomly," Bell Journal of Economics and Management Science, 4 (1973), pp. 369-374.
82. \_\_\_\_\_, "St. Petersburg Paradoxes: Defanged, Dissected, and Historically Described," Journal of Economic Literature, 15 (1977), pp. 24-55.
83. \_\_\_\_\_, The Collected Scientific Papers of Paul A. Samuelson, Volume III, R. C. Merton (ed.). Cambridge: M.I.T. Press (1972).
84. Sharpe, W., "Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk," Journal of Finance, 19 (1964), pp. 425-442.
85. \_\_\_\_\_, Portfolio Theory and Capital Markets. New York: McGraw-Hill (1970).
86. Smith, Jr., C. W., "Option Pricing: A Review," Journal of Financial Economics, 3 (1976), pp. 3-52.
87. Stiglitz, J. E., "A Re-Examination of the Modigliani-Miller Theorem," American Economic Review, 59 (1969), pp. 78-93.

59. \_\_\_\_\_, "A Reexamination of the Capital Asset Pricing Model," in Friend and Bicksler (eds.) [23].
60. \_\_\_\_\_, "On the Pricing of Contingent Claims and the Modigliani-Miller Theorem," Journal of Financial Economics, (forthcoming).
61. \_\_\_\_\_ and P. A. Samuelson, "Fallacy of the Log-Normal Approximation to Optimal Portfolio Decision Making Over Many Periods," Journal of Financial Economics, 1 (1974), pp. 67-94.
62. \_\_\_\_\_ and M. G. Subrahmanyam, "The Optimality of a Competitive Stock Market," Bell Journal of Economics and Management Science, 5 (1974), pp. 145-170.
63. Miller, M. H., "Debt and Taxes," Journal of Finance, 32 (1977), pp. 261-276.
64. \_\_\_\_\_ and F. Modigliani, "The Cost of Capital, Corporation Finance, and the Theory of Investment," American Economic Review, 48 (1958), pp. 261-297.
65. Mossin, J., "Equilibrium in a Capital Asset Market," Econometrica, 35 (1966), pp. 768-783.
66. Myers, S. C., "A Time-State-Preference Model of Security Valuation," Journal of Financial and Quantitative Analysis, 3 (1968), pp. 1-33.
67. Pratt, J. W., "Risk Aversion in the Small and in the Large," Econometrica, 32 (1964), pp. 122-136.
68. Radner, R., "Problems in the Theory of Markets Under Uncertainty," American Economic Review, (1970), pp. 454-460.
69. \_\_\_\_\_, "A Note on the Unanimity of Stockholders' Preferences Among Alternative Production Plans: A Reformulation of the Eckern-Wilson Model," Bell Journal of Economics and Management Science, 5 (1974), pp. 181-186.
70. Richard, S., "Optimal Consumption, Portfolio, and Life Insurance Rules for an Uncertain Lived Individual in a Continuous-Time Model," Journal of Financial Economics, 2 (1975), pp. 187-204.
71. Ross, S., "Arbitrage Theory of Capital Asset Pricing," Journal of Economic Theory 8 (1976).
72. \_\_\_\_\_, "Mutual Fund Separation in Financial Theory: The Separating Distributions," Journal of Economic Theory (forthcoming).
73. Rothschild, M. and J. E. Stiglitz, "Increasing Risk I: A Definition," Journal of Economic Theory, 2 (1970), pp. 225-243.

## Bibliography 7

88. \_\_\_\_\_, "On the Optimality of the Stock Market Allocation of Investment," Quarterly Journal of Economics, 86 (1972), pp. 25-60.
89. Tobin, J., "Liquidity Preference as Behavior Towards Risk," Review of Economic Studies, 25 (1958), pp. 68-85.
90. \_\_\_\_\_, "Comment on Borch and Feldstein," Review of Economic Studies, 36 (1969), pp. 13-14.
91. von Neumann, J. and O. Morgenstern, Theory of Games and Economic Behavior, 2nd Edition. Princeton: Princeton University Press (1947).