JOURNAL OF ECONOMIC THEORY 13, 341-360 (1976)

The Arbitrage Theory of Capital Asset Pricing

STEPHEN A. ROSS*

Departments of Economics and Finance, University of Pennsylvania, The Wharton School, Philadelphia, Pennsylvania 19174

Received March 19, 1973; revised May 19, 1976

The purpose of this paper is to examine rigorously the arbitrage model of capital asset pricing developed in Ross [13, 14]. The arbitrage model was proposed as an alternative to the mean variance capital asset pricing model, introduced by Sharpe, Lintner, and Treynor, that has become the major analytic tool for explaining phenomena observed in capital markets for risky assets. The principal relation that emerges from the mean variance model holds that for any asset, i, its (ex ante) expected return

$$E_i = \rho + \lambda b_i , \tag{1}$$

where ρ is the riskless rate of interest, λ is the expected excess return on the market, $E_m - \rho$, and

$$b_i = \sigma_{im}^2/\sigma_m{}^2,$$

is the beta coefficient on the market, where $\sigma_m{}^2$ is the variance of the market portfolio and σ_{im}^2 is the covariance between the returns on the ith asset and the market portfolio. (If a riskless asset does not exist, ρ is the zero-beta return, i.e., the return on all portfolios uncorrelated with the market portfolio.)[1]

The linear relation in (1) arises from the mean variance efficiency of the market portfolio, but on theoretical grounds it is difficult to justify either the assumption of normality in returns (or local normality in Wiener diffusion models) or of quadratic preferences to guarantee such efficiency, and on empirical grounds the conclusions as well as the

* Professor of Economics, University of Pennsylvania. This work was supported by a grant from the Rodney L. White Center for Financial Research at the University of Pennsylvania and by National Science Foundation Grant GS-35780.

[1] See Black [2] for an analysis of the mean variance model in the absence of a riskless asset.

assumptions of the theory have also come under attack.[2] The restrictiveness of the assumptions that underlie the mean variance model have, however, long been recognized, but its tractability and the evident appeal of the linear relation between return, E_i, and risk, b_i, embodied in (1) have ensured its popularity. An alternative theory of the pricing of risky assets that retains many of the intuitive results of the original theory was developed in Ross [13, 14].

In its barest essentials the argument presented there is as follows. Suppose that the random returns on a subset of assets can be expressed by a simple factor model

$$\tilde{x}_i = E_i + \beta_i \tilde{\delta} + \tilde{\epsilon}_i , \tag{2}$$

where $\tilde{\delta}$ is a mean zero common factor, and $\tilde{\epsilon}_i$ is mean zero with the vector $\langle \tilde{\epsilon} \rangle$ sufficiently independent to permit the law of large numbers to hold. Neglecting the noise term, $\tilde{\epsilon}_i$, as discussed in Ross [14] (2) is a statement that the state space tableau of asset returns lies in a two-dimensional space that can be spanned by a vector with elements δ_θ, (where θ denotes the state of the world) and the constant vector, $e \equiv \langle 1,..., 1 \rangle$.

Step 1. Form an arbitrage portfolio, η, of all the n assets, i.e., a portfolio which uses no wealth, $\eta e = 0$. We will also require η to be a well-diversified portfolio with each component, η_i, of order $1/n$ in (absolute) magnitude.

Step 2. By the law of large numbers, for large n the return on the arbitrage portfolio

$$\begin{aligned} \eta \tilde{x} &= \eta E + (\eta \beta)\, \tilde{\delta} + \eta \tilde{\epsilon} \\ &\approx \eta E + (\eta \beta)\, \tilde{\delta}. \end{aligned} \tag{3}$$

In other words the influence on the well-diversified portfolio of the independent noise terms becomes negligible.

Step 3. If we now also require that the arbitrage portfolio, η, be chosen so as to have no systematic risk, then

$$\eta \beta = 0,$$

and from (3)

$$\eta \tilde{x} \approx \eta E.$$

[2] See Blume and Friend [3] for a recent example of some of the empirical difficulties faced by the mean variance model. For a good review of the theoretical and empirical literature on the mean variance model see Jensen [6].

Step 4. Using no wealth, the random return $\eta\tilde{x}$ has now been engineered to be equivalent to a certain return, ηE, hence to prevent arbitrarily large disequilibrium positions we must have $\eta E = 0$. Since this restriction must hold for all η such that $\eta e = \eta\beta = 0$, E is spanned by e and β or

$$E_i = \rho + \lambda\beta_i \tag{4}$$

for constants ρ and λ. Clearly if there is a riskless asset, ρ must be its rate of return. Even if there is not such an asset, ρ is the rate of return on all zero-beta portfolios, α, i.e., all portfolios with $\alpha e = 1$ and $\alpha\beta = 0$. If α is a particular portfolio of interest, e.g., the market portfolio, α_m, with $E_m = \alpha_m E$, (4) becomes

$$E_i = \rho + (E_m - \rho)\,\beta_i\,. \tag{5}$$

Condition (5) is the arbitrage theory equivalent of (1), and if $\tilde{\delta}$ is a market factor return then β_i will approximate b_i. The above approach, however, is substantially different from the usual mean-variance analysis and constitutes a related but quite distinct theory. For one thing, the argument suggests that (5) holds not only in equilibrium situations, but in all but the most profound sort of disequilibria. For another, the market portfolio plays no special role.

There are, however, some weak points in the heuristic argument. For example, as the number of assets, n, is increased, wealth will, in general, also increase. Increasing wealth, though, may increase the risk aversion of some economic agents. The law of large numbers implies, in Step 2, that the noise term, $\eta\tilde{\epsilon}$, becomes negligible for large n, but if the degree of risk aversion is increasing with n these two effects may cancel out and the presence of noise may persist as an influence on the pricing relation. In Section I we will present an example of a market where this occurs. Furthermore, even if the noise term can be eliminated, it is not at all obvious that (5) must hold, since the disequilibrium position of one agent might be offset by the disequilibrium position of another.[3]

In Ross [13], however, it was shown that if (5) holds then it represents an ϵ or quasi-equilibrium. The intent of this paper is to supply the rigorous analysis underlying the stronger stability arguments above. In Section II we will present some weak sufficient conditions to rule out the above exceptions (and the example of Section I) and we will prove a general version of the arbitrage result. Section II also includes a brief argument on the empirical practicality of the results. A mathematical appendix

[3] Green has considered this point in a temporary equilibrium model. Essentially he argues that if subjective anticipations differ too much, then arbitrage possibilities will threaten the existence of equilibrium.

contains some supportive results of a somewhat technical and tangential nature. Section III will briefly summarize the paper and suggest further generalizations.

I. A Counterexample

In this section we will present an example of a market where the sequence of equilibrium pricing relations does not approach the one predicted by the arbitrage theory as the number of assets is increased. The counterexample is valuable because it makes clear what sort of additional assumptions must be imposed to validate the theory.

Suppose that there is a riskless asset and that risky assets are independently and normally distributed as

$$\tilde{x}_i = E_i + \tilde{\epsilon}_i , \tag{6}$$

where

$$E\{\tilde{\epsilon}_i\} = 0,$$

and

$$E\{\tilde{\epsilon}_i^2\} = \sigma^2.$$

The arbitrage argument would imply that in equilibrium all of the independent risk would disappear and, therefore,

$$E_i \approx \rho. \tag{7}$$

Assume, however, that the market consists of a single agent with a von Neumann–Morgenstern utility function of the constant absolute risk aversion form,

$$U(z) = -\exp(-Az). \tag{8}$$

Letting w denote wealth with the riskless asset as the numeraire, and α the portfolio of risky assets (i.e., α_i is the proportion of wealth placed in the ith risky asset) and taking expectations we have

$$\begin{aligned} &E\{U[w(\rho + \alpha[\tilde{x} - \rho \cdot e])]\} \\ &\quad = -\exp(-Aw\rho)\, E\{\exp(-Aw\alpha[\tilde{x} - \rho \cdot e])\} \\ &\quad = -\exp(-Aw\rho)\{\exp(-Aw\alpha[E - \rho \cdot e] + (\sigma^2/2)(Aw)^2(\alpha'\alpha))\}. \end{aligned} \tag{9}$$

The first-order conditions at a maximum are given by

$$\sigma^2(Aw)\, \alpha_i = E_i - \rho. \tag{10}$$

If the riskless asset is in unit supply the budget constraint (Walras' Law for the market) becomes

$$w = \sum_{i=1}^{n} \alpha_i w + 1 = (1/A\sigma^2) \sum_{i=1}^{n} (E_i - \rho) + 1. \tag{11}$$

The interpretation of the budget constraint (11) depends on the particular market situation we are describing. Suppose, first, that we are adding assets which will pay a random total numeraire amount, $\tilde{c}_i$. If p_i is the current numeraire price of the asset then

$$\tilde{x}_i = \tilde{c}_i / p_i .$$

Normalizing all risky assets to be in unit supply we must have

$$p_i = \alpha_i w,$$

and the budget constraint simply asserts that wealth is summed value,

$$w = \sum_{i=1}^{n} p_i + 1.$$

If we let $\bar{c}_i$ denote the mean of $\tilde{c}_i$ and c^2, its variance, then (10) can be solved for p_i as

$$p_i = (1/\rho)\{\bar{c}_i - Ac^2\}.$$

As a consequence, the expected returns,

$$E_i \equiv \bar{c}_i / p_i = \rho\{\bar{c}_i / (\bar{c}_i - Ac^2)\},$$

will be unaffected by changes in the number of assets, n, for $i < n$, and need bear no systematic relation to ρ as n increases. This is a violation of the arbitrage condition, (7). Notice, too, that as long as $\bar{c}_i$ is bounded above Ac^2, wealth and relative risk aversion, Aw, are unbounded in n.

An alternative interpretation of the market situation would be that as n increases the number of risky investment opportunities or activities is being increased, but *not* the number of assets. In this case wealth, w, would simply be the number of units of the riskless asset held and would remain constant as n increased. The quantities $\alpha_i w$ now represent the amount of the riskless holdings put into the ith investment opportunity and for the market as a whole we must have

$$\sum_{i=1}^{n} \alpha_i < 1.$$

Furthermore, if the random technological activities are irreversible, then each $\alpha_i \geqslant 0$. From (10) it follows that

$$E_i - \rho \geqslant 0$$

and

$$\sum_{i=1}^{n} E_i - \rho = \sum_{i=1}^{n} | E_i - \rho | = \sigma^2(Aw) \sum_{i=1}^{n} \alpha_i < \sigma^2 Aw.$$

Hence, as $n \to \infty$, the vector E approaches the constant vector with entries ρ in absolute sum (the l_1 norm) which is a very strong type of approximation. Under this second interpretation, then, the arbitrage condition (7) holds.

An easy way to understand the distinction between these two interpretations is to conceive of the riskless asset as silver dollars, and the risky assets as slot machines. In the first interpretation the slot machines come with a silver dollar in the slot and p_i is the relative price of the ith "primed" machine in terms of silver dollars. In the alternative interpretation, the machines are "unprimed" and we invest $\alpha_i w$ silver dollars in the ith machine. Which of these two senses of a market being "large" is empirically more relevant is a debatable issue, and in the next section we will develop assumptions sufficient to verify the arbitrage result for both cases (and any intermediate ones as well).

II. The Arbitrage Theory

The difficulty with the constant absolute risk aversion example arises because the coefficient of relative risk aversion increases with wealth. This suggests considering risk averse agents for whom the coefficient of relative risk aversion is uniformly bounded,

$$\sup_x \{-(U''(x)\, x/U'(x))\} \leqslant R < \infty. \tag{12}$$

We will refer to such agents as being of Type B (for bounded).

Pratt has shown that given a Type B utility function, U, there exists a monotone increasing convex function, $G(\cdot)$, such that

$$U(x) = G[U(x; R)], \tag{13}$$

where $U(x; R)$ is the utility function with constant relative risk aversion, R. It is well known that

$$U(x; R) = \begin{cases} x^{1-R}/(1 - R) & \text{if } R \neq 1, \\ \log x & \text{if } R = 1. \end{cases} \tag{14}$$

Essentially, then, Type B agents are uniformly less risk averse than some constant relative risk averse agents.

Assume that the returns on the particular subset of assets under consideration are subjectively viewed by agents in the market as being generated by a model of the form

$$\begin{aligned} \tilde{x}_i &= E_i + \beta_{i1}\tilde{\delta}_1 + \cdots + \beta_{ik}\tilde{\delta}_k + \tilde{\epsilon}_i\,, \\ &= E_i + \beta_i\tilde{\delta} + \tilde{\epsilon}_i\,, \end{aligned} \tag{15}$$

where

$$E\{\tilde{\delta}_l\} = E\{\tilde{\epsilon}_i\} = 0,$$

and where the $\tilde{\epsilon}_i$'s are mutually stochastically uncorrelated. We will impose no further restrictions on the form of the multivariate distribution of $(\tilde{\delta}, \tilde{\epsilon})$ beyond the requirement that $(\exists\, \sigma < \infty)$

$$\sigma_i^2 \equiv E\{\tilde{\epsilon}_i^2\} \leqslant \sigma^2. \tag{16}$$

In particular, then, the $\tilde{\delta}_i$ need not be jointly independent or even independent of the $\tilde{\epsilon}_i$'s, they need not possess variances, and none of the random variables need be normally distributed.

A point on notation is also needed. In what follows, α^0 will denote an n-element optimal portfolio for the agent under consideration, i.e., α^0 maximizes $E\{U[w\alpha\tilde{x}]\}$, subject to $\alpha e = 1$. The vector β^l will be the column vector $\langle \beta_{1l}\,,..., \beta_{nl}\rangle'$ and β_i, as above, denotes the row vector $\langle \beta_{i1}\,,..., \beta_{ik}\rangle$. The single letter β will denote the matrix

$$[\beta^1 \vdots \cdots \vdots \beta^k].$$

Assumption 1 (Liability limitations). There exists at least one asset with limited liability in the sense that there is some bound, t, (per unit invested) to the losses for which an agent is liable.

Assumption 1 is satisfied in the real world by a wide variety of assets. We can now prove a key result about Type B agents.

Theorem I. *Consider a Type* B *agent who lives in a world that satisfies Assumption* 1 *and who believes that returns are generated by a model of the form of* (15). *If* $(\exists m < \infty)$ *such that*

$$\alpha^0 E \leqslant m, \tag{17}$$

then $(\exists \rho$ *and a* k *vector,* $\gamma)$ *such that*

$$\sum_{i=1}^{\infty} [E_i - \rho - \beta_i\gamma]^2 < \infty. \tag{18}$$

Proof. The result is independent of the particular wealth sequence $\langle w^n \rangle$ and we must prove it for arbitrary sequences. Assume that $R \neq 1$. We will prove the theorem by constructing a portfolio that bests α^0 when (18) does not hold. First, from (17), concavity and monotonicity

$$
\begin{aligned}
&E\{U[w^n \alpha^0 \tilde{x}]\} \\
&\quad \leqslant U[w^n \alpha^0 E] \\
&\quad \leqslant U[w^n m] \\
&\quad = G[(w^n)^{1-R}\, U(m;\, R)].
\end{aligned}
$$

Now, consider the arbitrage portfolio sequence that solves the associated quadratic problem of minimizing unsystematic (ϵ) risk subject to the constraints of having no systematic (β) risk and attaining an expected return greater than $m + t$: minimize

$$\eta' V \eta,$$

subject to

$$
\begin{aligned}
\eta' e &= 0, \\
\eta' \beta^l &= 0; \qquad l = 1,..., k,
\end{aligned}
\tag{19}
$$

and

$$\eta' E = c > m + t,$$

where V is the covariance matrix of $\langle \tilde{\epsilon} \rangle$ and where t is the maximum liability loss associated with a unit investment in the limited liability asset. Assumption 1 guarantees that t is bounded. We will also assume, without loss of generality, that V is of full rank for all n.[4]

If the constraints are unsolvable for all n, then E must be linearly dependent on e and the columns of β and we are done. Suppose then, that the constraints are solvable for all n sufficiently large and, without loss of generality, let

$$X \equiv [E \,\vdots\, \beta \,\vdots\, e]$$

be of full rank.[5]

[4] Since the $\tilde{\epsilon}_i$ are uncorrelated, V is a diagonal matrix and will be of less than full rank only if some asset has no noise term. If there are two or more such assets the arbitrage argument holds exactly and we can eliminate such assets without loss of generality.

[5] If $[\beta]$ is not of full rank then we can simply eliminate dependent factors. If $[\beta]$ is of full rank, but $[\beta \,\vdots\, e]$ is not, then all assets will have a common factor $\tilde{\xi}$ and we can write (15) as

$$\tilde{x}_i = E_i + \tilde{\xi} + \beta_i \tilde{\delta} + \tilde{\epsilon}_i\,.$$

Now the proof of Theorem I is essentially unaltered, with the common factor, $\tilde{\xi}$ retained in all portfolios.

We will assume that if a sequence of random variables converges to a degenerate law (a constant) in quadratic mean, then the expected utility also converges, and defer a rigorous examination of this point to an appendix. It follows that there must not be any subsequence on which

$$\eta' V \eta \to 0.$$

If such a subsequence existed then

$$E\{U(\eta\tilde{x} - t; R)\} \to U(c - t; R) > U(m; R),$$

and by the convexity of $G(\cdot)$ there would exist an n such that putting all wealth in the limited liability asset and buying the arbitrage portfolio would yield

$$\begin{aligned} E\{U[w^n(\eta\tilde{x} - t)]\} &= E\{G[(w^n)^{1-R}\, U((\eta\tilde{x} - t); R)]\} \\ &\geqslant G[(w^n)^{1-R}\, E\{U((\eta\tilde{x} - t): R)\}] \\ &> G[(w^n)^{1-R}\, U(m; R)], \end{aligned}$$

violating optimality. Hence $(\exists a > 0)$ such that $(\forall n)$

$$\eta' V \eta \geqslant a > 0.$$

Solving (19) we have

$$V\eta = X\lambda,$$

where λ is a $(k + 2)$-vector of multipliers, and applying the constraints of (19) yields

$$[X' V^{-1} X]\, \lambda = \begin{bmatrix} c \\ 0 \end{bmatrix}.$$

It now follows that

$$\begin{aligned} \eta' V \eta &= \lambda' \begin{bmatrix} c \\ 0 \end{bmatrix} \\ &= [c, 0][X' V^{-1} X]^{-1} \begin{bmatrix} c \\ 0 \end{bmatrix} \\ &\geqslant a > 0. \end{aligned}$$

Defining $b \equiv (c, 0)$ we can apply Lemma I in the Appendix to obtain the existence of a^* and $A < \infty$ such that for all n

$$(Xa^*)'(Xa^*) \leqslant A < \infty, \tag{20}$$

where

$$a^* b = c a_1^* = 1$$

or

$$a_1{}^* = 1/c.$$

Defining $(1, -\gamma, -\rho) = ca^*$, (20) becomes the desired result (18).

If $R = 1$, wealth can be factored out of the utility function additively and the proof is nearly identical. Q.E.D.

Theorem I asserts that for a Type B individual, if the optimal expected return is uniformly bounded, then it must be the case that the arbitrage condition

$$\begin{aligned} E_i &\approx \rho + \beta_i \gamma \\ &= \rho + \gamma_1 \beta_{i1} + \cdots + \gamma_k \beta_{ik}, \end{aligned}$$

holds in the approximate sense that the sum of squared deviations is uniformly bounded. This implies, among other things, that as n increases

$$| E_n - \rho - \beta_n \gamma | \to 0. \tag{21}$$

A number of simple corollaries of Theorem I are available. If we adopt the alternative interpretation, suggested in Section I, that $\tilde{x}_i$ is the return on the ith activity, then wealth will be confined to a compact interval if there are a limited number of actual assets. It is easy to see that if wealth is confined to a compact interval on which the utility function is bounded, then Theorem I will hold for any risk averse agent. We also have the following corollary.

COROLLARY I. *Under the conditions of Theorem* I *if there is a riskless asset then* ρ *may be taken to be its rate of return.*

Proof. The return per unit of wealth in the presence of a riskless asset is given by

$$\rho + \alpha(\tilde{x} - \rho),$$

where α is now the portfolio of risky assets. Deleting the constraint that $\eta e = 0$ we can simply repeat the proof of Theorem I with $(E - \rho e)$ in the place of the E vector. Q.E.D.

Corollary I, of course, also extends to the alternative interpretation.

To turn these results into a capital market theory we will assume that there is at least one Type B individual who does not become negligible as the number of assets, n, is increased. The following definition is helpful.

DEFINITION. The agent, a^ν, will be said to be asymptotically negligible if, as the number of assets increases,

$$\omega^\nu \equiv w^\nu / w \to 0,$$

where w^{ν} is the agent's wealth and w is total wealth, i.e.,

$$w \equiv \sum_{\nu} w^{\nu}.$$

For example, an agent will not be asymptotically negligible if the sequence of proportionate quantities of assets the agent is endowed with is bounded away from zero.

ASSUMPTION 2 (Nonnegligibility of Type B agents). There exists at least one Type B agent who believes that returns are generated by a model of the form of (15) and who is not asymptotically negligible.

To permit us to aggregate to a market relation we will make three more assumptions; essentially we must ensure that Theorem I will not be "undone" by the rest of the economy. First we assume that agents hold compatible subjective beliefs.

ASSUMPTION 3 (Homogeneity of expectations). All agents hold the same expectations, E. Furthermore, all agents are risk averse.[6]

ASSUMPTION 4 (Extent of disequilibria). Let ξ_i denote the aggregate demand for the ith asset as a fraction of total wealth. We will assume that only situations with $\xi_i \geqslant 0$ are to be considered.

Notice that Assumption 4 does not rule out the possibility that an asset can be in excess supply; it only implies that the economy as a whole will wish to hold some of it. Assumptions 3 and 4 can be weakened considerably as will be shown below, but for purposes of demonstration we have chosen to leave them in a stronger than necessary form.

Lastly, we need to specify the generating model (15) a bit more.

ASSUMPTION 5 (Boundedness of expectations). The sequence, $\langle E_i \rangle$ is uniformly bounded, i.e.,

$$\| E \| \equiv \sup_i | E_i | < \infty. \tag{22}$$

Assumption 5 will be discussed in Section III.

We can now prove our central result.

[6] The assumption of risk aversion is quite weak since if fair gambles are permitted, any bounded nonconcave portions of agents' utility functions would be irrelevant. See Raiffa [11] or Ross [12] for an elaboration of this point.

THEOREM II. *Given Assumptions* 1 *through* 5, $(\exists \rho, \gamma)$

$$\sum_{i=1}^{\infty} \{E_i - \rho - \beta_i \gamma\}^2 < \infty. \tag{18}$$

Furthermore, if there is a riskless asset, then ρ *is its rate of return.*[7]

Proof. From Theorem I we know that if the conclusion is false then for the Type B agent (on a subsequence)

$$\sum_i \alpha_i{}^0 E_i \to \infty. \tag{23}$$

Let the total fraction of wealth held by the Type B agent be given by ω^0 and by the rest of the economy by $\hat{\omega}$. If $\hat{\alpha}_i$ denotes the fraction of $\hat{\omega}$ held in asset i by the rest of the economy then by Assumption 4

$$\xi_i \equiv \omega^0 \alpha_i{}^0 + \hat{\omega}\hat{\alpha}_i \geqslant 0.$$

By definition,

$$\sum_i \xi_i = 1,$$

hence

$$\begin{aligned} \| E \| &\geqslant \sum_i \xi_i E_i \\ &= \sum_i (\omega^0 \alpha_i{}^0 + \hat{\omega}\hat{\alpha}_i)\, E_i \\ &= \omega^0 \sum_i \alpha_i{}^0 E_i + \hat{\omega} \sum_i \hat{\alpha}_i E_i\,. \end{aligned}$$

From (23) and Assumption 2 the first sum in the last expression is divergent, which together with Assumption 5 (22) implies that

$$\hat{\omega} \sum_i \hat{\alpha}_i E_i \to -\infty.$$

Since

$$\hat{\omega}\hat{\alpha}_i \equiv \sum_{\nu \neq 0} \omega^\nu \alpha_i{}^\nu,$$

where ω^ν is the fraction of wealth held by a^ν, it follows that

$$\begin{aligned} \hat{\omega} \sum_i \hat{\alpha}_i E_i &= \sum_i \sum_{\nu \neq 0} \omega^\nu \alpha_i{}^\nu E_i \\ &= \sum_{\nu \neq 0} \left\{ \sum_i \omega^\nu \alpha_i{}^\nu E_i \right\}, \end{aligned}$$

[7] Theorems I and II and Corollary I can be extended to the case where (15) holds for only a subset of the assets by generalizing the utility function to be a Lebesque dominated sequence of functions conditional on the other assets.

and for some agent, a^ν,

$$\sum_i \omega^\nu \alpha_i{}^\nu E_i \to -\infty,$$

on a subsequence. By Assumptions 1 and 3 this contradicts optimality.

The identification of ρ with the riskless return follows from Corollary 1.

Q.E.D.

Theorem II has a straightforward extension to the alternative interpretation of $\tilde{x}_i$ as the return on activity i. In the extension, though, we can, of course, drop Assumption 2 and obtain (18) from Assumptions 1, 3, 4, and 5 alone.

As was shown in Ross [14] the basic result of Theorem 2 can be written in a number of empirically interesting and intuitively appealing formats. For example, by appropriate normalization it can be shown that

$$E_i - \rho \approx \beta_{i1}(E^1 - \rho) + \cdots + \beta_{ik}(E^k - \rho), \tag{24}$$

where E^l is the return on all portfolios with $\alpha\beta^s = 0$ for $s \neq l$ and $\alpha\beta^l = 1$. The constant ρ is now the return on all $\alpha\beta = 0$, i.e., zero-beta portfolios. Thus, the risk premium on an asset is the β-weighted sum of the factor risk premiums.

While we have formally proven the main result that the sum of squared deviations from the basic pricing relation is bounded above as the number of assets increases, it is worthwhile spending some effort to obtain an empirical estimate of the size of this bound. To do this we will work with a more exact form of our results. Examining the proof of Theorem I and Lemma I in the Appendix, we have found a bound to

$$\begin{aligned}\sum_{i=1}^{n} [E_n - \rho - \beta_n\gamma]^2 &\leqslant c^2 A \\ &= (1/ha)\, c^2,\end{aligned}$$

or, using the exact form of Lemma I, (obtained by leaving the H^n matrices in the sum) we have

$$\sum_{i=1}^{n} (1/\sigma_i{}^2)[E_i - \rho - \beta_i\gamma]^2 \leqslant c^2/a, \tag{25}$$

where c is the return premium on the arbitrage portfolio over a risk free rate ($-t$ in (19)) and a is the lower limit on the variance of an arbitrage portfolio.

If we assume that the market portfolio, as a well-diversified portfolio, cannot be grossly inefficient in a mean variance sense, and if we ignore ex ante–ex post distinctions, then we can use observed market data (see Friend [4] and Myers [9] for the data which follow) to obtain a rough estimate of the bound in (25). Over the period from January 1, 1962 to December 31, 1971 the yearly market return (Standard and Poor's Composite Index) averaged 7.4 % and the risk free rate (prime corporates with 1 year to maturity) averaged 5.1 %, for a market risk premium of

$$c = 2.3\,\%.$$

The sample variance of the market portfolio in this period was $(0.123)^2$, and we will assume that no arbitrage portfolio earning the market risk premium could have had less than one-half the market variance. Hence,

$$a = \tfrac{1}{2}(0.123)^2,$$

and from (25),

$$\sum_{i=1}^{n} (1/\sigma_i^2)[E_i - \rho - \beta_i\gamma]^2 \leqslant 2(0.023)^2/(0.123)^2.$$

The average residual variance in this period from regressions of asset returns on the market portfolio was about $2(0.123)^2$ and using this as a proxy for σ_i^2, the average squared discrepancy is approximately

$$\text{average}(E_i - \rho - \beta_i\gamma)^2 \leqslant (1/n)\, 4(0.023)^2.$$

Taking the number of assets n to be the combined total of listed issues on the NYSE and the Amex on December 31, 1971, about 3000, the average absolute discrepancy is given by

$$\text{average} \mid E_i - \rho - \beta_i\gamma \mid \leqslant 2 \cdot 0.023/3000^{1/2} = 0.00084,$$

or about 1 % of the market return of 7.4 %.

Of course these estimates are very crude and are only intended to be indicative; assets with a high own variance will have a greater latitude for discrepancies than those with low own variances. Most importantly, though, to the extent that there is significant cross-sectional correlation across the $\tilde{\epsilon}_i$ terms, the addition of further factors should reduce the own variance terms, σ_i^2, and improve the estimates.

III. Generalizations and Conclusions

One of the strengths of Theorem II is that it does not require the stringent homogeneity of anticipations of the mean-variance theory. We are now obviously distinguishing between expectations, i.e., the E vector, and anticipations, the whole model (15). If other agents have the same ex ante expectations, but believe returns are generated in a different fashion, then (24) must still hold where β is that of the return generating model believed to hold by the Type B agent. Of course, this is a bit gratuitous since in this model, as in all others, it is necessary to translate the results into observable quantities and the usual ex ante–ex post identity becomes ambiguous with disparate anticipations. Even if all agents agree on (15), however, there is still considerable scope for disagreement on the underlying probability distributions. For example, if $\tilde{\delta}$ represents a market or "GNP" factor, then as long as all agents agree on the impact of this factor on returns, through β_{i1}, they can hold a variety of views on the distribution of $\tilde{\delta}$ without violating the basic arbitrage condition, (24). Similarly, agents can also disagree on the distribution of the idiosyncratic noise terms, $\tilde{\epsilon}_i$, without altering (24). The primary difficulty with the analysis arises when agents differ in their expectations, E^ν. Now the proof of Theorem II must be modified since, unless all E^ν vectors are positive multiples of the same vector, we cannot be assured that the divergence of $\alpha^\nu E^\tau$ to $-\infty$ for $\tau \neq \nu$, implies that $\alpha^\nu E^\nu \to -\infty$. This is a fruitful area for generalizations.

It is also possible to weaken the condition that $\tilde{\epsilon}_i$ be mutually uncorrelated. For example, if the assets can be ordered so that $\tilde{\epsilon}_i$ and $\tilde{\epsilon}_j$ are uncorrelated if $|i-j|$ exceeds a given number, then the analysis is unchanged. In general, any weakening that permits a law of large numbers to hold should be sufficient, although weaker forms of the law would result in weaker approximation norms for the pricing relation (24).

Lastly, it should be emphasized that (24) is much more of an arbitrage relation than an equilibrium condition and may be expected to be quite robust. Assumptions 4 and 5 served only to guarantee that the market return,

$$E_m \equiv \sum_i \xi_i E_i \, , \tag{26}$$

would be uniformly bounded and this will hold in a wide class of disequilibrium situations. Rather then simply assuming that E_m was bounded, we chose to make Assumptions 4 and 5 directly to see how sufficient conditions for a bounded E_m would appear in alternative economic situations. For example, Assumption 4 can be weakened if, instead of having required all $\xi_i \geqslant 0$, we had assumed that $\sum_i |\xi_i|$ was bounded, i.e., we had bounded the sum of the absolute proportions of wealth placed (or shorted) in all assets. This would also be sufficient to bound the market return. In practice, these are very weak conditions and easily satisfied.[8]

In conclusion, we have set forth a rigorous basis for the arbitrage relation and arguments analyzed in Ross [14] (and [13]), and the conditions which are sufficient to support the theory have some intuitive appeal. On a less optimistic note, though, while significantly weakening the assumption that investors have identical (or homogeneous) anticipations, the arbitrage theory still requires essentially identical expectations and agreement on the β coefficients if the identification of ex ante beliefs with ex post realizations is to provide empirically fruitful results. If this assumption is to be fundamentally weakened, this theory (and all others) will require a closer examination of the dynamics by which ex ante beliefs are transformed into ex post observations. Such a study properly lies in the domain of general disequilibrium dynamics and, in particular, should focus on the impact of information on markets. It is one of the most difficult and important areas for future research.

[8] A strong form of Theorem 2 can be obtained by assuming that the weighted sum of subjectively viewed expected portfolio returns

$$\sum_{\nu} \omega^{\nu} \sum_{i} \alpha_i^{\nu} E_i^{\nu} \tag{F1}$$

is uniformly bounded. This would permit us to delete Assumptions 4, 5, and even 3 and, formally at least, would allow heterogeneous expectations. Alternatively, we could replace Assumption 5 with $\| E^{\nu} \| < \infty$, retain Assumption 4 (or the weaker form described in Section III) and drop Assumption 3.

Furthermore, if agents agree on factors and if the actual ex post model generating returns is some convex combination (say wealth weighted, or, for that matter, any uniformly sup norm bounded linear operator) of the individual market ex ante models, then the basic arbitrage condition will be expressible in ex post observables and, as such, will be directly testable. See Ross [14] for a fuller discussion of these issues. None of this, however, is very satisfactory. For one thing, it is not clear what is the force of these boundedness conditions, particularly when the number of agents is typically much larger than the number of marketed assets. As an example, if we have two Type B agents with exactly divergent beliefs (in a sense, which can be made precise in special examples) then they can exactly offset each other. There is now no reason to expect (F1), unlike (26), to be bounded simply because observed ex post return is bounded. For another, we must translate the theory into a statement about observables and this requires relating divergent subjective ex ante expectations to ex post ones via the "right" generating mechanism in a less ad hoc fashion. This is the problem posed in Section III and makes the "strong" version of Theorem II inadequate to stand alone.

Appendix 1

In this appendix we prove the lemma referred to in the proofs of the paper. Define a sequence of $n \times k$ matrices, $\langle X^n \rangle$, by taking the first row, the first two rows, and so on of an infinite matrix with k columns.

Lemma I. *Let $\langle X^n \rangle$ be a sequence of $n \times k$ matrices and let $\langle H^n \rangle$ be a sequence of diagonal matrices with diagonal elements $\langle h_1 \rangle$, $\langle h_1, h_2 \rangle$, and so on where, for some h, $h_i \geqslant h > 0$ for all i. Assume $(\exists b, a)$ $(\forall X^n$ of full rank)*

$$b'[X^{n\prime} H^n X^n]^{-1} b \geqslant a > 0. \tag{A1}$$

It follows that $(\exists a^*$ and $A)$

$$(X^n a^*)'(X^n a^*) \leqslant A < \infty$$

and

$$a^{*\prime} b = 1.$$

Proof. The result is trivial if X^n is of less than full rank for all n. In addition, if X^n is of full rank for some n $(\geqslant k)$ then $X^{\hat{n}}$ is of full rank, $\hat{n} > n$, and we may assume that the sequence $\langle X^n \rangle$ $(n \geqslant k)$ is of full rank for all n. By positive definiteness $X^{n\prime} H^n X^n$ is of full rank and (A1) holds.

Consider the problem:

$$\min (X^n z^n)' H^n (X^n z^n),$$

subject to

$$z^{n\prime} b = 1.$$

The solution is given by

$$z^n = \gamma [X^{n\prime} H^n X^n]^{-1} b,$$

where

$$\begin{aligned} \gamma &= (X^n z^n)' H^n (X^n z^n) \\ &= (b'[X^{n\prime} H^n X^n]^{-1} b)^{-1} \\ &\leqslant 1/a < \infty, \end{aligned}$$

by (A1). Consequently, from the lower bound on $\langle h_i \rangle$ we now obtain

$$(X^n z^n)'(X^n z^n) \leqslant A \equiv 1/ha < \infty.$$

Letting $y^n \equiv X^n z^n$ implies that $y^{n\prime} y^n \leqslant A$. If X is a full rank submatrix of X^n then

$$X z^n = y^n \mid X,$$

where $y^n \mid X$ is the corresponding subvector of y^n, and since $y^n \mid X$ is bounded in the norm it has a convergent subsequence. Letting y^* be its limit we must have $z^n \to a^* = X^{-1}y^*$ on the subsequence. It remains to show that $(\forall n)(X^n a^*)'(X^n a^*) \leqslant A$. Assume to the contrary that for some $\hat{n}$ (and, therefore, all $n > \hat{n}$)

$$(X^{\hat{n}} a^*)'(X^{\hat{n}} a^*) > A.$$

Since $z^n \to a^*$ on a subsequence we would have the contradiction

$$(X^n z^n)'(X^n z^n) \geqslant (X^{\hat{n}} z^n)'(X^{\hat{n}} z^n) > A \qquad \text{for some } n.$$

It follows that $(\forall n)\ (X^n a^*)'(X^n a^*) \leqslant A$. In addition, since $z^{n\prime} b = 1$ for all n we must also have $a^{*\prime} b = 1$. Q.E.D.

Appendix 2

In this appendix we discuss the relationship between convergence in quadratic mean (q.m.) and expected utility. The technical results can be found in Loeve [8] and Billingsley [1].

We can begin with a simple but powerful result. Let $\langle \tilde{X}_n \rangle$ be a sequence of random variables with $E\{\tilde{X}_n\} = 0$, and $\tilde{X}_n \to 0$ (q.m.), i.e., $\sigma^2(\tilde{X}_n) \to 0$.

Proposition. *If $U(\cdot)$ is concave and bounded below (which implies that the domain of $U(\cdot)$ is left bounded), then*

$$E\{U[\rho + \tilde{X}_n]\} \to U(\rho).$$

Proof. By Fatou's lemma

$$\liminf E\{U[\rho + \tilde{X}_n]\} \geqslant U(\rho),$$

but by concavity

$$E\{U[\rho + \tilde{X}_n]\} \leqslant U(\rho),$$

hence

$$\lim E\{U[\rho + \tilde{X}_n]\} = U(\rho).$$

Q.E.D.

A problem arises when $U(\cdot)$ is unbounded from below. About the weakest condition which assures convergence is uniform integrability (U.I.):

$$\lim_{\alpha \to \infty} \sup_n \int_{\Omega_\alpha} | U(\rho + X_n)| \, d\eta_n = 0,$$

$$\Omega_\alpha = \{| U(\rho + X_n)| \geqslant \alpha\},$$

where η_n is the distribution function of X_n.

A number of familiar conditions imply U.I. If the sequence $U(\rho + \tilde{X}_n)$ is bounded below by an integrable function the Lebesque convergence theorem can be invoked or if $(\exists \delta > 0)$

$$\sup_n E\{| U(\rho + X_n)|^{1+\delta}\} < \infty,$$

then the sequence is U.I.

In general, then, the convergence criterion will depend on both the utility function and the random variables. It is possible, however, to find weak sufficient conditions on the random variables alone, by taking advantage of the structure of $\tilde{X}_n$, but the condition that $\tilde{X}_n = (1/n) \sum_i \tilde{\epsilon}_i$; σ_i^2 uniformly bounded and $\tilde{\epsilon}_i$, $\tilde{\epsilon}_j$ independent is not sufficient.[9]

In the text, it is assumed that all sequences satisfy the U.I. condition, and therefore

$$\tilde{X}_n \to a \qquad \text{(q.m.)}$$

will imply that

$$E\{U(\tilde{X}_n)\} \to U(a).$$

References

1. P. Billingsley, "Convergence of Probability Measures," Wiley, New York, 1968.
2. F. Black, Capital market equilibrium with restricted borrowing, *J. Business* **45** (1972), 444–455.
3. M. Blume and I. Friend, A new look at the capital asset pricing model, *J. Finance* (March 1973), 19–33.
4. I. Friend, Rates of return on bonds and stocks, the market price of risk, and the cost of capital, Working Paper No. 23–73, Rodney L. White Center for Financial Research, 1973.
5. J. Green, Preexisting contracts and temporary general equilibrium, *in* "Essays on Economic Behavior under Uncertainty" (Balch, McFadden, and Wir, Eds.), North-Holland, Amsterdam, 1974.
6. M. Jensen, Capital markets: theory and evidence, *Bell. J. Econ. and Management Science* **3** (1972), 357–398.
7. J. Lintner, The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets, *Rev. Econ. Statist.* (February 1965), 30–55.
8. M. Loeve, "Probability Theory," Van Nostrand, Princeton, N. J., 1963.
9. S. Myers, A reexamination of market and industry factors in stock price behavior, *J. Finance* (June 1973), 695–705.
10. J. Pratt, Risk aversion in the small and in the large, *Econometrica* **32** (1964), 122–137.
11. H. Raiffa, "Decision Analysis," Addison-Wesley, Reading, Mass., 1968.

[9] It is not difficult to construct counterexamples by having $U(\cdot)$ go to $-\infty$ rapidly enough as x approaches its lower bound.

12. S. Ross, Comment on "Consumption and Portfolio Choices with Transaction Costs," by E. Zabel and R. Multherjee, *in* "Essays on Economic Behavior under Uncertainty" (Balch, McFadden, and Wir, Eds.), North-Holland, Amsterdam, 1974.
13. S. Ross, Portfolio and capital market theory with arbitrary preferences and distributions—The general validity of the mean-variance approach in large markets, Working Paper No. 12–72, Rodney L. White Center for Financial Research, 1971.
14. S. Ross, Return, risk and arbitrage, *in* "Risk and Return in Finance" (Friend and Bicksler, Eds.), Ballinger, Cambridge, Mass., forthcoming.
15. W. Sharpe, Capital asset prices: A theory of market equilibrium under conditions of risk, *J. Finance* (September 1964), 425–442.
16. J. Treynor, Toward a theory of market value of risky assets, unpublished manuscript, 1961.