Funds, Factors, and Diversification in Arbitrage Pricing Models

Author(s): Gary Chamberlain

Source: *Econometrica*, Sep., 1983, Vol. 51, No. 5 (Sep., 1983), pp. 1305-1323

Published by: The Econometric Society

Stable URL: https://www.jstor.org/stable/1912276

Econometrica, Vol. 51, No. 5 (September, 1983)

FUNDS, FACTORS, AND DIVERSIFICATION IN ARBITRAGE PRICING MODELS

BY GARY CHAMBERLAIN[1]

We present a definition of factor structure that is less restrictive than the one typically used in arbitrage pricing models. Our factor structure restrictions build on the following intuitive distinctions between factor variance and idiosyncratic variance: (i) A well-diversified portfolio contains only factor variance. (ii) If a portfolio is uncorrelated with the well-diversified portfolios, then it contains only idiosyncratic variance; so if a sequence of such portfolios becomes well-diversified, the limiting variance should be zero. Our factor structure restrictions imply Ross' [5] arbitrage pricing formula. We obtain upper and lower bounds on the approximation error in that formula; these bounds may be useful in empirical work. They imply that arbitrage pricing is exact if and only if there is a risky, well-diversified portfolio on the mean-variance frontier. If all mean-variance efficient portfolios are well-diversified, then the well-diversified portfolios provide mutual fund separation. Our factor structure restrictions are satisfied (with K factors) if and only if the covariance matrix of asset returns has only K unbounded eigenvalues as the number of assets increases.

1. INTRODUCTION

CONSIDER A SEQUENCE OF ASSETS, where one dollar invested in the ith asset gives a random return of x_i with expectation $E(x_i) = \mu_i$. In addition, there is a riskless asset s with a return of ρ. The principal implication of the capital asset pricing model of Sharpe [9] and Lintner [3] is that

$$\mu_i = \rho + \tau\beta_i, \tag{1.1}$$

where β_i is the covariance between x_i and the return on the market portfolio.

The theory underlying this result has received the following criticisms: There must be sharp restrictions on either preferences (quadratic utility) or on the distribution of returns (e.g., multivariate normality). Even if these restrictions are satisfied, observing the return on the market portfolio is difficult. Furthermore, investors have endowments of assets that cannot be traded, such as human capital. Finally, the pricing formula in (1.1) is derived from mutual fund separation—all investors divide their wealth between the riskless asset and a single fund, common to all investors, formed from the risky assets. Hence all investors should have perfectly correlated returns on their portfolios, which seems unlikely to be true.

Ross [5, 6] proposed an alternative basis for (1.1). It rests on the distributional restrictions implied by a factor structure:

$$x_i = \mu_i + \beta_{i1} f_1 + \cdots + \beta_{iK} f_K + v_i \qquad (i = 1, 2, \ldots), \tag{1.2}$$

[1] I am grateful to Michael Rothschild for helpful discussions. Financial support was provided by the National Science Foundation, the University of Wisconsin Graduate School, and an Alfred P. Sloan Research Fellowship.

where the idiosyncratic disturbances (v) are uncorrelated with each other and with the factors (f). This factor structure implies that the variance of a portfolio's return can be decomposed into two components: a (market) factor component generated by the f's and an idiosyncratic component generated by the v's—the f's represent economy-wide shocks that affect all asset returns, and v_i represents the uncertainty that is specific to the ith asset. Ross showed that the absence of arbitrage opportunities in equilibrium will imply that (1.1), or its K-factor generalization, is approximately true in the following sense: there exist numbers $\tau_1, \ldots, \tau_K$ such that

$$\gamma \equiv \sum_{i=1}^{\infty} (\mu_i - \rho - \tau_1 \beta_{il} - \cdots - \tau_K \beta_{iK})^2 < \infty. \tag{1.3}$$

In this paper we put sharp upper and lower bounds on γ in (1.3); these bounds imply necessary and sufficient conditions for exact arbitrage pricing ($\gamma = 0$). We also relate this characterization to mutual fund separation. Finally, we show that the assumption that the idiosyncratic components are uncorrelated is unnecessarily strong and can be relaxed; we impose a weaker restriction on the distribution of the asset returns.

Section 2 sets up the Hilbert space framework and summarizes some results from Chamberlain and Rothschild [**1**]. If $p_N = \sum_{i=1}^{N} \alpha_{iN} x_i$ converges (in mean-square) to p, then we refer to p as (the return on) a limit portfolio. If it is not possible to obtain a riskless, positive, return at zero cost, then the linear functional that gives the cost of a portfolio is continuous. It follows from a theorem of Riesz that this linear functional can be represented by a limit portfolio, c. This limit portfolio in turn generates the mean-variance frontier; i.e., every mean-variance efficient portfolio is a linear combination of s and c.

In Section 3 we show how a factor structure allows us to put sharp bounds on the arbitrage pricing formula (1.3). Our procedure is abstract and indirect. First we write down restrictions on the return process that (i) capture the intuitive notion of a K-factor structure and (ii) allow us to establish (1.3). Then in Section 5 we give a sufficient condition for the returns to satisfy these restrictions.

The intuitive distinction between factor variance and idiosyncratic variance is that a well-diversified portfolio contains only factor variance. Our first factor structure restriction builds on this intuition. First we develop a measure of diversification $\| \; \|_2$: if $p = \sum_{i=1}^{N} \alpha_i x_i$, then $\|p\|_2 = (\sum_{i=1}^{N} \alpha_i^2)^{1/2}$. Let $\mathbf{\Sigma}_N$ be the covariance matrix of $(x_1, \ldots, x_N)$, with eigenvalues $\lambda_{1N} \geq \cdots \geq \lambda_{NN}$. We assume that $\mathbf{\Sigma}_N$ does not approach singularity as N increases, so that these eigenvalues are uniformly bounded away from zero: $\lambda_\infty \equiv \inf_N \lambda_{NN} > 0$. This allows us to extend the definition of $\| \; \|_2$ to limit portfolios. We say that a limit portfolio p is well diversified if $\|p\|_2 = 0$.

Our first factor structure restriction is that there be only K (linearly independent) well-diversified limit portfolios. For an asset x_i, the factor loadings, $\beta_{i1}, \ldots, \beta_{iK}$, are the covariances between x_i and these K portfolios.

Our second factor structure restriction is a continuity requirement: If a portfolio is uncorrelated with the factors, then it contains only idiosyncratic variance; so if a sequence of such portfolios becomes well-diversified, then the limiting variance should be zero. Hence if there is a factor structure, then there is a positive constant $\bar{\lambda}$ such that $V(p) \leq \bar{\lambda}\|p\|_2^2$ whenever p is uncorrelated with the well-diversified portfolios.

The main result in Section 3 is that our factor structure restrictions together with a no arbitrage condition imply that (1.3) holds with

$$\lambda_\infty^2 \rho^2 \|c\|_2^2 \leq \gamma \leq \bar{\lambda}^2 \rho^2 \|c\|_2^2. \tag{1.4}$$

A corollary is that there is exact arbitrage pricing ($\gamma = 0$) if and only if all risky, mean-variance efficient portfolios are well-diversified. There is a related result in Connor [**2**]; he has shown that there is exact arbitrage pricing if the supplies of the assets are well-diversified.

Section 4 relates our result to mutual fund separation. Consider the problem of maximizing $E(u(p))$, where $u(\ )$ is a concave utility function and p is constrained to be a limit portfolio formed from $\{s, x_1, x_2, \ldots\}$ that costs one dollar. We show that if the mean-variance frontier is well-diversified, then the solution to this problem is a linear combination of the riskless asset and a well-diversified limit portfolio. Hence the K well-diversified limit portfolios serve as mutual funds which, together with the riskless asset, span the set of portfolios that are optimal for a risk-averse investor.

This mutual fund result has an important qualification. Maximizing $E(u(p))$ is an appropriate objective only if all relevant assets have been included in $\{s, x_1, x_2, \ldots\}$. Part of the appeal of the arbitrage pricing theory is that it does not require us to specify the "market" portfolio, nor does it preclude nontraded assets such as human capital. Our result in (1.4) and its corollary share this advantage. Even if $\{s, x_1, x_2, \ldots\}$ is only a subsequence of the traded assets, we have exact arbitrage pricing for this subsequence if and only if the mean-variance frontier formed from this subsequence is well-diversified.

Section 5 provides sufficient conditions for our factor structure restrictions. The conditions are that only the K largest eigenvalues of $\boldsymbol{\Sigma}_N$ are unbounded as $N \to \infty$ and that $\lambda_\infty > 0$. If λ_{K+1} is a uniform bound on the $(K+1)$th largest eigenvalue, then we can set $\bar{\lambda} = \lambda_{K+1}$ in (1.4), so that

$$\lambda_\infty^2 \rho^2 \|c\|_2^2 \leq \gamma \leq \lambda_{K+1}^2 \rho^2 \|c\|_2^2. \tag{1.5}$$

It is shown in [**1**], using results from Section 5, that these bounded eigenvalue conditions imply that $\{\boldsymbol{\Sigma}_N\}$ has an *approximate K-factor structure*, in the following sense: there exists a sequence $\{\beta_{i1}, \ldots, \beta_{iK}\}_{i=1}^\infty$ such that

$$\boldsymbol{\Sigma}_N = \boldsymbol{B}_N \boldsymbol{B}_N' + \boldsymbol{R}_N \qquad (N = 1, 2, \ldots), \tag{1.6}$$

where the i, j element of $\boldsymbol{B}_N$ is β_{ij} and $\{\boldsymbol{R}_N\}$ is a sequence of positive semi-definite matrices whose eigenvalues are uniformly bounded (by λ_{K+1}) for all N.

Furthermore, this decomposition of $\{\boldsymbol{\Sigma}_N\}$ into $\{\boldsymbol{B}_N\boldsymbol{B}_N'\}$ and $\{\boldsymbol{R}_N\}$ is unique, and we can obtain the β_{ij} from the covariance between x_i and the K well-diversified portfolios. Since $\boldsymbol{R}_N$ need not be diagonal, the idiosyncratic components of the asset returns are allowed to be correlated. We show in Section 5 that our factor-structure restrictions are equivalent to an approximate K-factor structure.

An implication of this paper for empirical work is that it may be fruitful to attempt to measure λ_∞, λ_{K+1}, and $\|c\|_2$.

2. ARBITRAGE AND CONTINUITY

This section sets up the Hilbert space framework and summarizes some of the relevant results from [**1**]. One dollar invested in the ith asset gives a random return of x_i. A portfolio formed by investing α_i in the ith asset has a random return of $\sum_{i=1}^N \alpha_i x_i$; the portfolio is represented by the vector $(\alpha_1, \ldots, \alpha_N)$. Short sales are allowed, so α_i may be negative.

There is an underlying probability space, and $L_2(P)$ denotes the collection of all random variables with finite variances defined on that space. The x_i are assumed to have finite variances, so that the sequence $\{x_i, i = 1, 2, \ldots\}$ is in $L_2(P)$. The means, variances, and covariances of the x_i are denoted by

$$\mu_i = E(x_i), \qquad \sigma_{ii} = V(x_i), \qquad \sigma_{ij} = \operatorname{Cov}(x_i, x_j).$$

We let $\mathscr{F}_N = [x_1, \ldots, x_N]$ denote the span of $x_1, \ldots, x_N$; i.e., the linear subspace consisting of all linear combinations of $x_1, \ldots, x_N$. Let $\mathscr{F} = \bigcup_{N=1}^\infty \mathscr{F}_N$, so that $p \in \mathscr{F}$ is the random return on a portfolio formed from some finite subset of the assets.

It is well-known that $L_2(P)$ is a Hilbert space under the mean-square inner product: the inner product of p and q is

$$E(pq) = \operatorname{Cov}(p,q) + E(p)E(q) \qquad (p, q \in L_2(P)),$$

with $(E(p^2))^{1/2}$ as the associated norm. Since $\mathscr{F}$ is a linear subspace of $L_2(P)$, its closure, $\overline{\mathscr{F}}$, is also a Hilbert space.

Let $\boldsymbol{\Sigma}_N$ denote the covariance matrix of $(x_1, \ldots, x_N)$. We shall assume that $\boldsymbol{\Sigma}_N$ is nonsingular for all N. Hence the return on the finite portfolio $(\alpha_1, \ldots, \alpha_N)$ has zero variance only if the α_i are all zero. The cost of the portfolio $(\alpha_1, \ldots, \alpha_N)$ is $\sum_{i=1}^N \alpha_i$. If $p = \sum_{i=1}^N \alpha_i x_i$ and $q = \sum_{i=1}^N \beta_i x_i$, then "$p = q$" refers to equality in $L_2(P)$; i.e., $E(p-q)^2 = 0$. But then $V(p-q) = 0$ implies that $\alpha_i = \beta_i$. Hence the cost of p, $C(p) = \sum_{i=1}^N \alpha_i$, is well-defined for $p \in \mathscr{F}$. We shall identify $p \in \mathscr{F}$ with its (unique) associated portfolio.

Let $\{p_N\}$ be a sequence of finite portfolios ($p_N \in \mathscr{F}$). We shall say that the market permits no arbitrage opportunities if the following two conditions hold:

Condition (A.i): If $V(p_N) \to 0$ and $C(p_N) \to 0$, then $E(p_N) \to 0$.

Condition (A.ii): If $V(p_N) \to 0$, $C(p_N) \to 1$, and $E(p_N) \to \alpha$, then $\alpha > 0$.

Condition (A.i) states that it is not possible to make an investment that is costless, riskless, and yields a positive return. If Condition (A.ii) does not hold, then the market allows an investor to trade a portfolio that, approximately, costs a dollar and has a riskless, nonpositive return. By selling this portfolio short, it is possible to generate arbitrarily large amounts of cash while incurring no future obligations.

It is shown in [**1**] that Condition (A.ii) implies that the cost functional is continuous on $\mathscr{F}$, so that it can be extended to a continuous linear functional on $\overline{\mathscr{F}}$. Since the cost of p is then well-defined for $p \in \overline{\mathscr{F}}$, we shall refer to these random returns as limit portfolios, or simply portfolios. The mean functional is clearly continuous on $\overline{\mathscr{F}}$. It then follows from Riesz' Theorem [**4**, p. 43] that there exist limit portfolios m and c that generate $E(\ )$ and $C(\ )$:

$$E(p) = E(mp), \qquad C(p) = E(cp)$$

for all $p \in \overline{\mathscr{F}}$.

The limit portfolios m and c generate the mean-variance efficient set in the following sense: Given any $q \in \overline{\mathscr{F}}$, let $p^0 = \alpha m + \beta c$ be the orthogonal projection of q onto the span of m, c. Then p^0 solves the following problem: min $V(p)$ subject to $p \in \overline{\mathscr{F}}$, $E(p) = E(q)$, $C(p) = C(q)$.

We shall say that there is a riskless limit portfolio if there is a $p^* \in \overline{\mathscr{F}}$ with $V(p^*) = 0$ and $E(p^*) \neq 0$. If (A) holds, then $C(p^*) \neq 0$ and we can set $s = p^*/C(p^*)$. Then s costs one dollar and we shall refer to s as a *riskless asset*. It is unique and its return ρ is positive: $\rho = E(s) > 0$.

It will often be convenient to work with a covariance inner product. There are two essential cases. If there is a riskless limit portfolio p^*, then $V^{1/2}(p)$ is not a valid norm on $\overline{\mathscr{F}}$ since $V(p^*) = 0$. So we let $z_i = x_i - \mu_i(p^*/E(p^*))$, $\mathscr{P} = \bigcup_{N=1}^{\infty}[z_1, \ldots, z_N]$, and let $\overline{\mathscr{P}}$ be the (mean-square) closure of $\mathscr{P}$ in $\overline{\mathscr{F}}$. Then $E(z_i) = 0$ implies that the mean-square inner product is actually a covariance inner product on the Hilbert space $\overline{\mathscr{P}}$: $E(pq) = \operatorname{Cov}(p, q)$ for $p, q \in \overline{\mathscr{P}}$.

If there is no riskless limit portfolio, we set $z_i = x_i$, so that $\overline{\mathscr{P}} = \overline{\mathscr{F}}$. It is shown in [**1**] that $\overline{\mathscr{P}}$ is a Hilbert space under the covariance inner product, and that $E(p_N - p)^2 \to 0$ if and only if $V(p_N - p) \to 0$;[2] hence there is no ambiguity in saying that p_N converges to p. If (A) holds, there are limit portfolios m^* and c^* that generate the mean and cost functionals under the covariance inner product: $E(p) = \operatorname{Cov}(m^*, p)$, $C(p) = \operatorname{Cov}(c^*, p)$ for $p \in \overline{\mathscr{P}}$. Furthermore, the span of m^*, c^* coincides with the span of m, c; hence all the mean-variance efficient portfolios are contained in $[m^*, c^*]$.

Our principal analytic tool will be the projection theorem [**4**, p. 42]. If $\mathscr{G}$ is a closed linear subspace of a Hilbert space $\mathscr{H}$, then every $p \in \mathscr{H}$ has a unique decomposition as $p = p_1 + p_2$, where $p_1 \in \mathscr{G}$ and $p_2 \in \mathscr{G}^{\perp}$ (i.e., the inner product of p_2 and q is zero for every $q \in \mathscr{G}$). This direct sum decomposition is denoted by $\mathscr{H} = \mathscr{G} \oplus \mathscr{G}^{\perp}$. The projection theorem is often used together with the fact that

[2] Condition (A) is not needed for this result.

every finite dimensional subspace is closed. For example, if there is a riskless asset s, then every point in $\overline{\mathscr{F}}$ has a unique decomposition as $\alpha s + p$, where $p \in [s]^{\perp}$. Since $E(sz_i) = 0$ for all i, we have $[s]^{\perp} = \overline{\mathscr{P}}$ and $\overline{\mathscr{F}} = [s] \oplus \overline{\mathscr{P}}$.

3. ROSS' THEOREM AND WELL-DIVERSIFIED MEAN-VARIANCE FRONTIERS

We shall say that there is a *strict K-factor structure* if $\boldsymbol{\Sigma}_N$ can be decomposed as follows:

$$\boldsymbol{\Sigma}_N = \boldsymbol{B}_N \boldsymbol{B}_N' + \boldsymbol{D}_N \qquad (N = 1, 2, \dots), \tag{3.1}$$

where $\boldsymbol{B}_N$ is a $N \times K$ matrix whose i, j element is β_{ij} and $\boldsymbol{D}_N$ is a diagonal matrix with uniformly bounded elements: $\boldsymbol{D}_N = \operatorname{diag}\{\theta_1, \dots, \theta_N\}$, $\theta_i \leq \zeta < \infty$ for $i = 1, 2, \dots$. Ross [5] showed that the strict K-factor structure and a no arbitrage condition imply that there exist numbers $\psi, \tau_1, \dots, \tau_K$ such that

$$\gamma \equiv \sum_{i=1}^{\infty} (\mu_i - \psi - \tau_1 \beta_{i1} - \cdots - \tau_K \beta_{iK})^2 < \infty. \tag{3.2}$$

We shall consider a weaker set of distributional restrictions that are still sufficient for this result. Then we shall provide upper and lower bounds on γ.

The factor structure in (3.1) implies that the variance of a portfolio can be decomposed into two components:

$$V\left(\sum_{i=1}^{N} \alpha_i x_i\right) = \boldsymbol{\alpha}_N' \boldsymbol{B}_N \boldsymbol{B}_N' \boldsymbol{\alpha}_N + \sum_{i=1}^{N} \alpha_i^2 \theta_i ,$$

where $\boldsymbol{\alpha}_N' = (\alpha_1, \dots, \alpha_N)$. We shall refer to the first component as factor variance and to the second component as idiosyncratic variance. Consider a sequence of portfolios $\{\sum_{i=1}^N \alpha_{iN} x_i\}$ with $\sum_{i=1}^N \alpha_{iN}^2 \to 0$. Since the θ_i are uniformly bounded, the idiosyncratic variance tends to zero along this sequence, and we shall say that the sequence of portfolios is well-diversified. If a well-diversified sequence of portfolios converges to $p \in \overline{\mathscr{F}}$, then we shall say that p is well-diversified. We shall let $\mathscr{D}$ denote the set of *well-diversified portfolios*.

DEFINITION 1: $\mathscr{D} = \{p \in \overline{\mathscr{F}}$: there exists a sequence $p_N = \sum_{i=1}^N \alpha_{iN} x_i$ with $p_N \to p$, $\sum_{i=1}^N \alpha_{iN}^2 \to 0\}$.

Let $\lambda_{1N} \geq \cdots \geq \lambda_{NN}$ be the eigenvalues of $\boldsymbol{\Sigma}_N$. We shall assume that $\boldsymbol{\Sigma}_N$ does not approach singularity as N increases, so that these eigenvalues are uniformly bounded away from zero: $\lambda_\infty \equiv \inf_N \lambda_{NN} > 0$. If $\{\boldsymbol{\Sigma}_N\}$ has a strict factor structure, then this condition holds if the idiosyncratic variances are uniformly bounded away from zero.

If there is a riskless limit portfolio p^*, then there is a sequence $p_N = \sum_{i=1}^N \alpha_{iN} x_i$ such that $p_N \to p^*$. Since $V(p_N) \geq \lambda_\infty \sum_{i=1}^N \alpha_{iN}^2$ and $V(p_N) \to 0$, we have $\sum_{i=1}^N \alpha_{iN}^2$

$\to 0$. Hence the riskless limit portfolio is well-diversified: $p^* \in \mathscr{D}$. We are more interested, however, in risky, well-diversified portfolios, since we shall be able to interpret those portfolios as factors. So we shall work with $\overline{\mathscr{P}}$, which is a subset of $\bar{F}$ that does not contain a riskless limit portfolio. Recall that $\overline{\mathscr{P}}$ is the mean-square closure of $\bigcup_{N=1}^{\infty}[z_1, \ldots, z_N]$, where $z_i = x_i - \mu_i(p^*/E(p^*))$ if there is a riskless limit portfolio and $z_i = x_i$ if there is not. In both cases $\overline{\mathscr{P}}$ is a Hilbert space under the covariance inner product.

DEFINITION 2: $\mathscr{P}_1 = \{p \in \overline{\mathscr{P}}$: there exists a sequence $p_N = \sum_{i=1}^N \alpha_{iN} z_i$ with $p_N \to p$, $\sum_{i=1}^N \alpha_{iN}^2 \to 0\}$.

It will be useful to have a corresponding measure of diversification.

LEMMA 1: *If* $\lambda_\infty \equiv \inf_N \lambda_{NN} > 0$, *then*:

(i) *There is a nonnegative, real-valued function* ($\| \ \|_2$) *defined on* $\overline{\mathscr{F}}$ *with the following properties*: *for* $\psi \in \mathscr{R}$ *and* $p, q \in \overline{\mathscr{F}}$,

$$\|\psi p\|_2 = |\psi| \, \|p\|_2 ; \qquad \|p + q\|_2 \le \|p\|_2 + \|q\|_2 ;$$

if $\sum_{i=1}^N \alpha_{iN} x_i \to p$ *as* $N \to \infty$, *then*

$$\left(\sum_{i=1}^N \alpha_{iN}^2 \right)^{1/2} \to \|p\|_2 .$$

(ii) *If* $p \in \overline{\mathscr{F}}$, *then* $V(p) \ge \lambda_\infty \|p\|_2^2$; *if* $p_N \in \overline{\mathscr{F}}$ *and* $p_N \to p$, *then* $\|p_N\|_2 \to \|p\|_2$.
(iii) $\mathscr{D} = \{p \in \bar{F} : \|p\|_2 = 0\}$, *which is a linear subspace of* $\overline{\mathscr{F}}$.
(iv) *If* $p_1 \in \mathscr{D}$ *and* $p \in \overline{\mathscr{F}}$, *then* $\|p_1 + p\|_2 = \|p\|_2$.
(v) *If* $p \in \overline{\mathscr{P}}$ *and* $\sum_{i=1}^N \alpha_{iN} z_i \to p$, *then* $(\sum_{i=1}^N \alpha_{iN}^2)^{1/2} \to \|p\|_2$.
(vi) $\mathscr{P}_1 = \{p \in \overline{\mathscr{P}} : \|p\|_2 = 0\}$, *which is a linear subspace of* $\overline{\mathscr{F}}$.

The proof is in the Appendix. Note that special cases of (i) and (v) are

$$\left\| \sum_{i=1}^N \alpha_i x_i \right\|_2 = \left\| \sum_{i=1}^N \alpha_i z_i \right\|_2 = \left(\sum_{i=1}^N \alpha_i^2 \right)^{1/2}$$

A well-diversified portfolio should contain only factor variance. If there are to be only K factors, then there can be only K linearly independent, well-diversified portfolios in $\mathscr{P}_1$. This motivates the following restriction on the distribution of asset returns:

CONDITION (F.i): $\dim \mathscr{P}_1 = K$.

We can then define the factors $f_1, \ldots, f_K$ to be any basis for $\mathscr{P}_1$. It is convenient to choose an orthonormal basis: $\operatorname{Cov}(f_j, f_k) = 0$, $V(f_j) = 1$ $(j, k =$

$1, \ldots, K$; $j \neq k$). Then we can form the orthogonal projection of z_i onto $\mathscr{P}_1$, using the covariance inner product:

$$z_i = \beta_{i1} f_1 + \cdots + \beta_{iK} f_K + e_i \qquad (i = 1, 2, \ldots), \tag{3.3}$$

where $\beta_{ij} = \text{Cov}(z_i, f_j)$ and $\text{Cov}(e_i, f_j) = 0$ $(j = 1, \ldots, K)$.

Now let $\mathscr{P}_2$ be the set of portfolios in $\overline{\mathscr{P}}$ that are uncorrelated with the well-diversified portfolios:

DEFINITION 3: $\mathscr{P}_2 = \{p \in \overline{\mathscr{P}} : \text{Cov}(p, q) = 0 \text{ for all } q \in \mathscr{P}_1\}$.

If $p \in \mathscr{P}_2$, then p is uncorrelated with the factors; we would like to interpret p as containing only idiosyncratic risk, so that a well-diversified sequence of portfolios in $\mathscr{P}_2$ will have a limiting variance of zero. Hence we assume:

CONDITION (F.ii): If $p_N \in \mathscr{P}_2$ and $\| p_N \|_2 \to 0$, then $V(p_N) \to 0$.

Note that (F.ii) is equivalent to the existence of a positive number $\bar{\lambda}$ such that

$$V(p) \leq \bar{\lambda} \| p \|_2^2 \tag{3.4}$$

for all $p \in \mathscr{P}_2$.

In Section 5 we shall show that the following conditions on the $\{\boldsymbol{\Sigma}_N\}$ sequence are sufficient for (F) to hold: $\sup_N \lambda_{KN} = \infty$, $\lambda_{K+1} \equiv \sup_N \lambda_{K+1,N} < \infty$, and $\lambda_\infty \equiv \inf_N \lambda_{NN} > 0$. Furthermore, we can set $\bar{\lambda} = \lambda_{K+1}$ in (3.4). It is shown in **[1]**, using results from Section 5, that these eigenvalue conditions imply that $\{\boldsymbol{\Sigma}_N\}$ has an *approximate K-factor structure*, in the following sense: there exists a sequence $\{\beta_{i1}, \ldots, \beta_{iK}\}_{i=1}^{\infty}$ such that

$$\boldsymbol{\Sigma}_N = \boldsymbol{B}_N \boldsymbol{B}_N' + \boldsymbol{R}_N \qquad (N = 1, 2, \ldots), \tag{3.5}$$

where the i, j element of the $N \times K$ matrix $\boldsymbol{B}_N$ is β_{ij} and $\{\boldsymbol{R}_N\}$ is a sequence of positive semi-definite matrices whose eigenvalues are uniformly bounded for all N. So the sequence of diagonal matrices $\{\boldsymbol{D}_N\}$ in the strict factor structure in (3.1) is replaced by a much more general sequence, which allows for correlation between the idiosyncratic components of the asset returns. It is also shown in **[1]** that as $N \to \infty$, we can obtain the column space of the β_{ij}'s from the eigenvectors of $\boldsymbol{\Sigma}_N$ corresponding to the K largest eigenvalues. Furthermore, the decomposition of $\{\boldsymbol{\Sigma}_N\}$ into $\{\boldsymbol{B}_N \boldsymbol{B}_N'\}$ and $\{\boldsymbol{R}_N\}$ is unique. We show in Section 5 that this decomposition is precisely the one implied by (3.3).

Now we are ready to give our extension of Ross' theorem. In addition to using a less restrictive definition of factor structure, we shall supply sharp bounds on the approximation error in the arbitrage pricing formula. These bounds show that the formula is exact if and only if there is a risky, well-diversified portfolio on the mean-variance frontier.

THEOREM 1: *Suppose that* (A) *and* (F) *hold, and that* $\lambda_\infty > 0$. *Let* $\beta_{ij} = \operatorname{Cov}(x_i, f_j)$.

(i) *If there is a riskless asset, define* $\tau_j = E(f_j) - \rho C(f_j)$; *then*

$$\lambda_\infty^2 \rho^2 \|c\|_2^2 \le \sum_{i=1}^{\infty} (\mu_i - \rho - \tau_1 \beta_{i1} - \cdots - \tau_K \beta_{iK})^2 \le \bar{\lambda}^2 \rho^2 \|c\|_2^2 .$$

Furthermore, if $\sum_{i=1}^{\infty}(\mu_i - \psi - \psi_1 \beta_{i1} - \cdots - \psi_K \beta_{iK})^2 < \infty$, *then* $\psi = \rho$ *and* $\psi_j = \tau_j (j = 1, \ldots, K)$.

(ii) *If there is no riskless asset, define* $\tau_j = E(f_j) - \psi C(f_j)$, *where* ψ *may be any real number*; *then*

$$\lambda_\infty^2 \|m^* - \psi c^*\|_2^2 \le \sum_{i=1}^{\infty} (\mu_i - \psi - \tau_1 \beta_{i1} - \cdots - \tau_K \beta_{iK})^2$$

$$\le \bar{\lambda}^2 \|m^* - \psi c^*\|_2^2 .$$

Furthermore, if $\sum_{i=1}^{\infty}(\mu_i - \psi - \psi_1 \beta_{i1} - \cdots - \psi_K \beta_{iK})^2 < \infty$, *then* $\psi_j = \tau_j$ $(j = 1, \ldots, K)$.

PROOF: If there is a riskless asset, let $a = m - \rho c$ and note that $m = \rho^{-1} s$. Then $a \in \overline{\mathscr{P}}$ since

$$E(a) = E(\rho^{-1} s - \rho c) = 1 - \rho E(\rho^{-1} s c) = 1 - C(s) = 0.$$

Hence $\operatorname{Cov}(a, f_j) = E(a f_j) = E(f_j) - \rho C(f_j)$. If there is no riskless asset, set $a = m^* - \psi c^*$; then $a \in \overline{\mathscr{P}}$ since $\overline{\mathscr{P}} = \overline{\mathscr{F}}$. Form the orthogonal projection of a onto $\mathscr{P}_1$, using the covariance inner product:

$$a = \tau_1 f_1 + \cdots + \tau_K f_K + e,$$

where $e \in \mathscr{P}_2$. Note that $\tau_j = \operatorname{Cov}(a, f_j) = E(f_j) - \varphi C(f_j)$, where $\varphi = \rho$ if there is a riskless asset and $\varphi = \psi$ if there is not. Let $\gamma_i \equiv \operatorname{Cov}(e, x_i)$ and note that

$$\gamma_i = \mu_i - \varphi - \tau_1 \beta_{i1} - \cdots - \tau_K \beta_{iK} .$$

Let $p_N = \sum_{i=1}^{N} \gamma_i z_i$ and decompose $p_N = p_{1N} + p_{2N}$, where $p_{1N} \in \mathscr{P}_1$ and $p_{2N} \in \mathscr{P}_2$. Lemma 1 (iii, vi) implies that $\mathscr{P}_1 \subset \mathscr{D}$. Since p_{1N} is in $\mathscr{P}_1$, Lemma 1(iv) implies that $\|p_N\|_2 = \|p_{2N}\|_2$. Since $f_1, \ldots, f_K$ are $\mathscr{P}_1$, we also have $\|a\|_2 = \|e\|_2$. Hence the Cauchy–Schwarz inequality and (3.4) imply that

$$\left(\sum_{i=1}^{N} \gamma_i^2\right)^2 = (\operatorname{Cov}(e, p_N))^2 = (\operatorname{Cov}(e, p_{2N}))^2$$

$$\le V(e) V(p_{2N}) \le \bar{\lambda}^2 \|e\|_2^2 \|p_{2N}\|_2^2$$

$$= \bar{\lambda}^2 \|a\|_2^2 \|p_N\|_2^2 = \bar{\lambda}^2 \|a\|_2^2 \sum_{i=1}^{N} \gamma_i^2 ;$$

so we have

$$\sum_{i=1}^{N} \gamma_i^2 \leq \bar{\lambda}^2 \|a\|_2^2$$

for all N. In the riskless asset case, $\|a\|_2^2 = \rho^2 \|c\|_2^2$.

To obtain the lower bound, let $e_N = \sum_{i=1}^{N} \alpha_{iN} z_i \to e$ and note that

$$(\mathrm{Cov}(e, e_N))^2 = \left(\sum_{i=1}^{N} \alpha_{iN} \gamma_i \right)^2 \leq \|e_N\|_2^2 \sum_{i=1}^{N} \gamma_i^2 .$$

Taking the limit of both sides of the inequality gives

$$\|e\|_2^2 \sum_{i=1}^{\infty} \gamma_i^2 \geq (V(e))^2 \geq \lambda_\infty^2 (\|e\|_2^2)^2$$

(by Lemma 1(ii)). Hence

$$\sum_{i=1}^{\infty} \gamma_i^2 \geq \lambda_\infty^2 \|e\|_2^2 = \lambda_\infty^2 \|a\|_2^2 .$$

Finally, let $\varphi_i = \mu_i - \psi - \psi_1 \beta_{i1} - \cdots - \psi_K \beta_{iK}$ and suppose that $\sum_{i=1}^{\infty} \varphi_i^2 < \infty$. Let

$$g \equiv m - \psi c - \psi_1 f_1 - \cdots - \psi_K f_K .$$

If there is a riskless asset, then $f_j \in \mathscr{P}_1$ implies $E(f_j) = 0$, so that $E(x_i f_j) = \mathrm{Cov}(x_i, f_j) = \beta_{ij}$ and $E(g x_i) = \varphi_i$. Let $q_N = \sum_{i=1}^{N} \alpha_{iN} x_i \to q \in \mathscr{D}$. Then

$$(E(q_N g))^2 = \left(\sum_{i=1}^{N} \alpha_{iN} \varphi_i \right)^2 \leq \|q_N\|_2^2 \sum_{i=1}^{N} \varphi_i^2 .$$

Taking the limit of both sides of the inequality gives

$$(E(qg))^2 \leq \|q\|_2^2 \sum_{i=1}^{\infty} \varphi_i^2 = 0.$$

Hence $g \in \mathscr{D}^{\perp} = \mathscr{P}_2$. ($\overline{\mathscr{F}} = [s] \oplus \mathscr{P}_1 \oplus \mathscr{P}_2 = \mathscr{D} \oplus \mathscr{P}_2$.) Hence $E(g) = 0$, so that

$$0 = E(m - \psi c) = 1 - \psi E(\rho^{-1} s c) = 1 - \psi \rho^{-1}$$

implies $\psi = \rho$. Then $g = m - \rho c - \sum_{j=1}^{K} \psi_j f_j \in \mathscr{P}_2$ implies that $\sum_{j=1}^{K} \psi_j f_j$ is the (covariance) orthogonal projection of $m - \rho c$ onto $\mathscr{P}_1$. Hence $\psi_j = \tau_j$. If there is no riskless asset, let $g \equiv m^* - \psi c^* - \sum_{j=1}^{K} \psi_j f_j$ and note that $\varphi_i = \mathrm{Cov}(g, x_i)$. As above, we can show that $g \in \mathscr{P}_2$, so that $\sum_{j=1}^{K} \psi_j f_j$ is the (covariance) orthogonal projection of $m^* - \psi c^*$ onto $\mathscr{P}_1$. Hence $\psi_j = \tau_j$. *Q.E.D.*

The following result is an immediate implication of Theorem 1.

COROLLARY 1: *Suppose that* (A) *and* (F) *hold, and that* $\lambda_\infty > 0$. *Let* $\beta_{ij} = \operatorname{Cov}(x_i, f_j)$.

(i) *If there is a riskless asset, then there exist numbers* $\psi, \psi_1, \dots, \psi_K$ *such that*

$$\mu_i = \psi + \psi_1 \beta_{i1} + \cdots + \psi_K \beta_{iK} \qquad (i = 1, 2, \dots)$$

if and only if $\psi = \rho$ *and* c *is well-diversified.*

(ii) *If there is no riskless asset, then for a given* ψ, *there exist numbers* $\psi_1, \dots, \psi_K$ *such that*

$$\mu_i = \psi + \psi_1 \beta_{i1} + \cdots + \psi_K \beta_{iK} \qquad (i = 1, 2, \dots)$$

if and only if $m^* - \psi c^*$ *is well-diversified.*

DEFINITION 4: There is a *well-diversified mean-variance frontier* if m and c are well-diversified.

DEFINITION 5: Let $\beta_{ij} = \operatorname{Cov}(x_i, f_j)$. There is *exact arbitrage pricing* if there exist numbers $\psi, \psi_1, \dots, \psi_K$ such that

$$\mu_i = \psi + \psi_1 \beta_{i1} + \cdots + \psi_K \beta_{iK} \qquad (i = 1, 2, \dots).$$

If there is a riskless asset, then $m = \rho^{-1} s$ is well-diversified. Hence c is well-diversified if and only if the mean-variance frontier is well-diversified. So in the riskless asset case, exact arbitrage pricing is equivalent to a well-diversified mean-variance frontier. If there is no riskless asset, then the Corollary asserts that we have exact arbitrage pricing provided that some portfolio on the mean-variance frontier is well-diversified. It is not necessary that both m and c be well-diversified. But a well-diversified mean-variance frontier is always a sufficient condition for exact arbitrage pricing, whether there is or is not a riskless asset.

4. MUTUAL FUND SEPARATION

We have shown that a well-diversified mean-variance frontier implies exact arbitrage pricing. We now want to establish conditions under which a well-diversified frontier also implies mutual fund separation. The set of K well-diversified portfolios will constitute the funds which span the set of risky portfolios that are optimal for any risk-averse investor.

Ross [7] provides a characterization of mutual fund separation for markets with a finite set of assets. If x_0 is a riskless asset with a return of ρ, then the essential restrictions on the distribution of asset returns are that

$$x_i = \mu_i + \beta_{i1}^* f_1^* + \cdots + \beta_{iK}^* f_K^* + e_i^*, \tag{4.1}$$

$$\mu_i = \rho + \psi_1 \beta_{i1}^* + \cdots + \psi_K \beta_{iK}^*, \tag{4.2}$$

$$E(e_i^* \mid f_1^*, \dots, f_K^*) = 0 \qquad (i = 1, \dots, N), \tag{4.3}$$

and there exist linearly independent vectors $\boldsymbol{\alpha}_k' = (\alpha_{1k}, \ldots, \alpha_{Nk})$ such that

$$\sum_{i=1}^{N} \alpha_{ik} e_i^* = 0 \qquad (k = 1, \ldots, K). \tag{4.4}$$

Then the portfolios $\sum_{i=1}^{N} \alpha_{ik} x_i$ contain only "factor" variance. It is a fairly direct implication of Jensen's inequality that these K portfolios constitute mutual funds which, together with the riskless asset, span the set of portfolios that are optimal for any risk-averse investor.

In the finite asset case, the existence of portfolios that do not contain e_i^* variance is not well motivated. But in our context, this restriction in (4.4) corresponds precisely to the elimination of idiosyncratic variance in a well-diversified portfolio. Equation (4.1) corresponds to our approximate K-factor structure in (3.3). The disturbances e_i^* are not restricted to be uncorrelated; in fact they are linearly dependent. The only restriction we place on the e_i in (3.3) is that linear combinations of them converge to zero if the linear combinations form a well-diversified sequence. Equation (4.2) corresponds to exact arbitrage pricing, which holds in our model if the mean-variance frontier is well-diversified. We almost have a counterpart to (4.3), since $\mathrm{Cov}(f_k, e_i) = 0$ in (3.3). But we need to strengthen this to mean independence.

CONDITION (M): If $p_1 \in \mathscr{P}_1$ and $p_2 \in \mathscr{P}_2$, then $E(p_2 | p_1) = E(p_2)$.

Now it follows from Jensen's inequality that our factors are funds if the mean-variance frontier is well-diversified.[3]

THEOREM 2: *Suppose that* (A), (F), *and* (M) *hold, that* $\lambda_\infty > 0$, *and that the mean-variance frontier is well-diversified.*

(i) *If there is a riskless asset s, then given any portfolio $\alpha s + p$ ($p \in \overline{\mathscr{P}}$), there is an $f \in \mathscr{P}_1$ such that $C(f) = C(p)$ and $E(u(\alpha s + f)) \geq E(u(\alpha s + p))$ for every concave utility function u.*

(ii) *If there is no riskless asset, then given any portfolio $p \in \overline{\mathscr{P}}$, there is an $f \in \mathscr{P}_1$ such that $C(f) = C(p)$ and $E(u(f)) \geq E(u(p))$ for every concave utility function u.*

PROOF: $p = f + p_2$, where $f \in \mathscr{P}_1$ and $p_2 \in \mathscr{P}_2$. If there is a riskless asset s, then $c = \psi s + c_1$, where $c_1 \in \mathscr{P}_1$; $E(p) = 0$ for all $p \in \overline{\mathscr{P}}$, so that $E(p_2) = 0$, which implies that

$$C(p_2) = E(cp_2) = \mathrm{Cov}(c, p_2) = \mathrm{Cov}(c_1, p_2) = 0.$$

If there is no riskless asset, then $c^* \in \mathscr{P}_1$, so that $C(p_2) = \mathrm{Cov}(c^*, p_2) = 0$. In

[3] There is a related result, based on a different argument, in Ross [**8**]. He works with a strict factor structure, as in (3.1), and uses a quadratic approximation to u to argue that exact arbitrage pricing implies mutual fund separation, in a certain approximate sense.

either case,

$$C(p) = C(f) + C(p_2) = C(f).$$

If there is a riskless asset, then $E(p_2) = 0$. If there is no riskless asset, then $m^* \in \mathscr{P}_1$ implies that $E(p_2) = \mathrm{Cov}(m^*, p_2) = 0$. In either case, (M) implies that

$$E(p_2 \mid f) = E(p_2) = 0.$$

In what follows, simply set $\alpha = 0$ if there is no riskless asset:

$$E(\alpha s + p \mid f) = \alpha s + f + E(p_2 \mid f) = \alpha s + f;$$

hence Jensen's inequality gives

$$E(u(\alpha s + p) \mid f) \leq u(E(\alpha s + p \mid f)) = u(\alpha s + f),$$

so that

$$E(u(\alpha s + p)) = E(E(u(\alpha s + p) \mid f)) \leq E(u(\alpha s + f)). \qquad Q.E.D.$$

If there is a riskless asset, then Corollary 1 shows that a well-diversified mean-variance frontier is equivalent to exact arbitrage pricing. Then it follows from Theorem 2 that exact arbitrage pricing implies mutual fund separation. In the one-factor case, we have

$$\mu_i = \rho + \tau\beta_i \qquad (i = 1, 2, \ldots), \tag{4.5}$$

where $\beta_i = \mathrm{Cov}(x_i, f)$ and f is the return on any risky, mean-variance efficient portfolio; i.e., $f = \alpha_1 s + \alpha_2 c$ ($\alpha_2 \neq 0$). The mutual fund property implies that the return on the jth investor's portfolio has the form $\alpha_{1j} s + \alpha_{2j} c$, and so the return on the market portfolio is also a linear combination of s and c. Hence we can set f equal to the return on the market portfolio, and (4.5) is precisely the capital asset pricing formula.

This synthesis of the two models appears to suggest that the arbitrage pricing formula can work very well, with $K = 1$, only if everyone is holding a mean-variance efficient portfolio. This mutual fund property is the key element in the Sharpe–Lintner model, but it is also one of its objectionable features; I find it implausible that all investors should have perfectly correlated returns on their portfolios.

In fact this objection need not apply to the arbitrage pricing version of (4.5). Theorem 2 is based on maximizing expected utility over portfolios formed from $\{x_1, x_2, \ldots\}$; this is appropriate only if all relevant assets have been included in $\{x_1, x_2, \ldots\}$. Just as in the capital asset pricing model, there is an implicit assumption that $\{x_1, x_2, \ldots\}$ provides an exhaustive list of the risks that the investor faces. Theorem 1 is more robust. It requires an infinite set of assets, but it allows $\{x_1, x_2, \ldots\}$ to be merely a subsequence. We may have omitted various assets, some of which are not traded. Nevertheless, if the mean-variance frontier based on $\{x_1, x_2, \ldots\}$ is well-diversified, then we shall have exact arbitrage pricing for this subset of the assets. It need not be true, however, that this subsequence of the assets enters all portfolios through the same K mutual funds.

5. THE BOUNDED EIGENVALUE CONDITION

Recall that $\lambda_{1N} \geq \cdots \geq \lambda_{NN}$ are the eigenvalues of $\boldsymbol{\Sigma}_N$.

THEOREM 3: *Suppose that* $\sup_N \lambda_{KN} = \infty$, $\lambda_{K+1} \equiv \sup_N \lambda_{K+1,N} < \infty$, *and* $\lambda_\infty \equiv \inf_N \lambda_{NN} > 0$. *Then* (F) *holds and we can set* $\bar{\lambda} = \lambda_{K+1}$ *in* (3.4).

The proof uses a series of lemmas. These lemmas are based exclusively on the Hilbert space $\overline{\mathscr{P}}$, using a covariance inner product and a variance norm:

$$(p,q) = \operatorname{Cov}(p,q), \qquad \|p\| = (V(p))^{1/2}$$

for $p,q \in \overline{\mathscr{P}}$. Recall that $z_i = x_i - \mu_i(p^*/E(p^*))$ if there is a riskless limit portfolio p^*, and $z_i = x_i$ if there is not. In both cases, $\mathscr{P} = \bigcup_{N=1}^{\infty}[z_1, \ldots, z_N]$ is a dense subset of $\overline{\mathscr{P}}$ (under the variance norm) and $\operatorname{Cov}(z_i, z_j) = \operatorname{Cov}(x_i, x_j) = \sigma_{ij}$. The proofs of the lemmas will invoke properties of $\| \ \|_2$; these properties are all contained in Lemma 1.

Let the spectral decomposition of $\boldsymbol{\Sigma}_N$ be

$$\boldsymbol{\Sigma}_N = \sum_{j=1}^{N} \lambda_{jN} \boldsymbol{t}_{jN} \boldsymbol{t}_{jN}',$$

where the eigenvectors $\boldsymbol{t}_{jN}$ satisfy $\boldsymbol{t}_{jN}'\boldsymbol{t}_{jN} = 1$, $\boldsymbol{t}_{jN}'\boldsymbol{t}_{kN} = 0$ $(j,k = 1, \ldots, N;\ j \neq k)$.

LEMMA 2: *If* $\lambda_{K+1} < \infty$ *and* $\lambda_\infty > 0$, *then* $\dim \mathscr{P}_1 \leq K$.

PROOF: Suppose that $\dim \mathscr{P}_1 > K$ and choose $p_j \in \mathscr{P}_1$ such that $(p_j, p_k) = 0$, $\|p_j\| = 1$ $(j,k = 1, \ldots, K+1;\ j \neq k)$. Let $p_{jN} = \boldsymbol{\alpha}_{jN}' z_N \rightarrow p_j$, where $\boldsymbol{\alpha}_{jN}' = (\alpha_{1jN}, \ldots, \alpha_{NjN})$ and $z_N' = (z_1, \ldots, z_N)$. Then

$$(p_{jN}, p_{kN}) = \boldsymbol{\alpha}_{jN}' \boldsymbol{\Sigma}_{1N} \boldsymbol{\alpha}_{kN} + \boldsymbol{\alpha}_{jN}' \boldsymbol{\Sigma}_{2N} \boldsymbol{\alpha}_{kN},$$

where

$$\boldsymbol{\Sigma}_{1N} = \sum_{j=1}^{K} \lambda_{jN} \boldsymbol{t}_{jN} \boldsymbol{t}_{jN}', \qquad \boldsymbol{\Sigma}_{2N} = \boldsymbol{\Sigma}_N - \boldsymbol{\Sigma}_{1N}.$$

Note that

$$(\boldsymbol{\alpha}_{jN}' \boldsymbol{\Sigma}_{2N} \boldsymbol{\alpha}_{kN})^2 \leq (\boldsymbol{\alpha}_{jN}' \boldsymbol{\Sigma}_{2N} \boldsymbol{\alpha}_{jN})(\boldsymbol{\alpha}_{kN}' \boldsymbol{\Sigma}_{2N} \boldsymbol{\alpha}_{kN})$$

$$\leq \lambda_{K+1}^2 \|p_{jN}\|_2^2 \|p_{kN}\|_2^2 \rightarrow 0$$

as $N \to \infty$. Hence $\boldsymbol{\alpha}'_{jN}\boldsymbol{\Sigma}_{1N}\boldsymbol{\alpha}_{kN} \to (p_j, p_k)$. Let $\boldsymbol{A}_N$ be the $N \times (K+1)$ matrix whose jth column is $\boldsymbol{\alpha}_{jN}$. Then $\boldsymbol{A}'_N\boldsymbol{\Sigma}_{1N}\boldsymbol{A}_N \to \boldsymbol{I}_{K+1}$, an identity matrix of order $K+1$. But $\det(\boldsymbol{A}'_N\boldsymbol{\Sigma}_{1N}\boldsymbol{A}_N) = 0$, since rank $\boldsymbol{\Sigma}_{1N} = K$. This contradiction implies that $\dim \mathscr{P}_1 \leq K$. *Q.E.D.*

DEFINITION 6: $\mathscr{P}_2^* = \{p \in \overline{\mathscr{P}} : \sum_{i=1}^{\infty}(p, z_i)^2 < \infty\}$.

LEMMA 3: *If* $\lambda_\infty > 0$, *then* $\mathscr{P}_1 = \mathscr{P}_2^{*\perp}$.

PROOF: Let $p_N = \boldsymbol{\alpha}'_N z_N \to p \in \mathscr{P}_1$. Given a $q \in \mathscr{P}_2^*$, set $\gamma_i = (q, z_i)$ and $\gamma'_N = (\gamma_1, \dots, \gamma_N)$. Then

$$(p_N, q)^2 = (\boldsymbol{\alpha}'_N \gamma_N)^2 \leq \|p_N\|_2^2 \sum_{i=1}^{N} \gamma_i^2 \to \|p\|_2^2 \sum_{i=1}^{\infty} \gamma_i^2 = 0.$$

Hence $(p, q) = \lim(p_N, q) = 0$, and so $\mathscr{P}_1 \subset \mathscr{P}_2^{*\perp}$.

Now take an arbitrary $p \in \mathscr{P}_2^{*\perp}$ and show that $\|p\|_2 = 0$. Let $p_N = \sum_{i=1}^{\infty} \alpha_{iN} z_i \to p$, where $\alpha_{iN} = 0$ for $i > N$. Then $\|p_N - p_M\| \to 0$ as $N, M \to \infty$. Hence $\|p_N - p_M\|^2 \geq \lambda_\infty \|p_N - p_M\|_2^2$ implies that $\sum_{i=1}^{\infty}(\alpha_{iN} - \alpha_{iM})^2 \to 0$ as $N, M \to \infty$. Since the square-summable sequences form a complete linear space (l_2), there exists a sequence $\{\gamma_i\}$ with $\sum_{i=1}^{\infty} \gamma_i^2 < \infty$ and $\sum_{i=1}^{\infty}(\alpha_{iN} - \gamma_i)^2 \to 0$ as $N \to \infty$. Define a linear functional on $\mathscr{P}$ by $G(z_i) = \gamma_i$. Then if $q = \sum_{i=1}^{N} \beta_i z_i \in \mathscr{P}$, we have

$$(G(q))^2 \leq \left(\sum_{i=1}^{N} \beta_i^2\right)\left(\sum_{i=1}^{N} \gamma_i^2\right) \leq \lambda_\infty^{-1}\|q\|^2 \sum_{i=1}^{\infty} \gamma_i^2,$$

so that $G(\ )$ is continuous at $q = 0$ and hence continuous on $\mathscr{P}$ [**4**, p. 9]. Then $G(\ )$ can be extended to a continuous linear functional on $\overline{\mathscr{P}}$ [**4**, p. 9]. It follows from Riesz' Theorem [**4**, p. 43] that there exists a $q \in \overline{\mathscr{P}}$ such that $(q, z_i) = \gamma_i$; in fact $q \in \mathscr{P}_2^*$ since $\sum_{i=1}^{\infty} \gamma_i^2 < \infty$. Since $p_N = \sum_{i=1}^{N} \alpha_{iN} z_i \to p$ and $p \in \mathscr{P}_2^{*\perp}$, we have

$$(p_N, q) = \sum_{i=1}^{N} \alpha_{iN} \gamma_i \to (p, q) = 0.$$

Since

$$\sum_{i=1}^{\infty} (\alpha_{iN} - \gamma_i)^2 = \sum_{i=1}^{N} \alpha_{iN}^2 - 2\sum_{i=1}^{N} \alpha_{iN}\gamma_i + \sum_{i=1}^{\infty} \gamma_i^2 \to 0$$

as $N \to \infty$, we have

$$\|p_N\|_2^2 = \sum_{i=1}^{N} \alpha_{iN}^2 \to 0.$$

Hence $\|p\|_2 = 0$ and $\mathscr{P}_2^{*\perp} \subset \mathscr{P}_1$. *Q.E.D.*

Let $z_N' = (z_1, \ldots, z_N)$ and define

$$r_{jN} = (t_{jN}' z_N)/\lambda_{jN}^{1/2}, \tag{5.1}$$

so that $(r_{jN}, r_{kN}) = 0$, $\|r_{jN}\| = 1$ $(j, k = 1, \ldots, N;\ j \neq k)$.

LEMMA 4: *If* $\sup_N \lambda_{KN} = \infty$, $\lambda_{K+1} < \infty$, *and* $\lambda_\infty > 0$, *then* $\|p\|^2 \leq \lambda_{K+1}\|p\|_2^2$ *for* $p \in \mathscr{P}_2$.

PROOF: Let $p_N = \alpha_N' z_N \to p \in \mathscr{P}_2^*$, and let $\gamma_i = (p, z_i)$, $\gamma_N' = (\gamma_1, \ldots, \gamma_N)$. If $j \leq K$,

$$(p, r_{jN})^2 = (\gamma_N' t_{jN})^2/\lambda_{jN} \leq \gamma_N'\gamma_N/\lambda_{jN}$$

$$\leq \sum_{i=1}^{\infty} \gamma_i^2/\lambda_{jN} \to 0$$

as $N \to \infty$. Since $r_{1N}, \ldots, r_{NN}$ provide an orthonormal basis for $[z_1, \ldots, z_N]$, we have

$$\begin{aligned} \|p_N\|^2 &= \sum_{j=1}^{N} (p_N, r_{jN})^2 = \sum_{j=1}^{K} (p_N, r_{jN})^2 + \sum_{j=K+1}^{N} \lambda_{jN} (\alpha_N' t_{jN})^2 \\ &\leq \sum_{j=1}^{K} (p_N, r_{jN})^2 + \lambda_{K+1} \sum_{j=1}^{N} (\alpha_N' t_{jN})^2 \\ &= \sum_{j=1}^{K} (p_N, r_{jN})^2 + \lambda_{K+1} \alpha_N' \alpha_N . \end{aligned}$$

Since $\alpha_N'\alpha_N \to \|p\|_2^2$, taking limits as $N \to \infty$ of both sides of the inequality gives $\|p\|^2 \leq \lambda_{K+1}\|p\|_2^2$ for $p \in \mathscr{P}_2^*$. Define $\overline{\mathscr{P}}_2^*$ to be the closure of $\mathscr{P}_2^*$ and let $p_n \to p \in \overline{\mathscr{P}}_2^*$ as $n \to \infty$, where $p_n \in \mathscr{P}_2^*$. Then $\|p_n\|^2 \leq \lambda_{K+1}\|p_n\|_2^2$, and taking limits as $n \to \infty$ gives $\|p\|^2 \leq \lambda_{K+1}\|p\|_2^2$ for $p \in \overline{\mathscr{P}}_2^*$. $\mathscr{P}_2^*$ is a linear subspace, since the square-summable sequences form a linear space. Since Lemma 3 shows that $\mathscr{P}_1 = \mathscr{P}_2^{*\perp}$, we can complete the proof by noting that [**4**, p. 64]

$$\mathscr{P}_2 = \mathscr{P}_1^{\perp} = (\mathscr{P}_2^{*\perp})^{\perp} = \overline{\mathscr{P}}_2^* . \qquad Q.E.D.$$

Lemma 2 shows that $\mathscr{P}_1$ has finite dimension if $\lambda_{K+1} < \infty$ and $\lambda_\infty > 0$. Then $\mathscr{P}_1$ is closed and we can uniquely decompose $p \in \overline{\mathscr{P}}$ into $p = p_1 + p_2$, where $p_1 \in \mathscr{P}_1$ and $p_2 \in \mathscr{P}_2$. The operator Q defined by $Qp = p_1$ gives the orthogonal projection of p onto $\mathscr{P}_1$. Since $r_{1N}, \ldots, r_{KN}$ are orthonormal, $Q_N p = \sum_{j=1}^{K}(p, r_{jN})r_{jN}$ gives the orthogonal projection onto the subspace of $\overline{\mathscr{P}}$ spanned by $r_{1N}, \ldots, r_{KN}$.

LEMMA 5: *If* $\sup_N \lambda_{KN} = \infty$, $\lambda_{K+1} < \infty$, *and* $\lambda_\infty > 0$, *then* (i) $\lim_{N\to\infty} Q_N p = Qp$ $(p \in \overline{\mathscr{P}})$; (ii) $\dim \mathscr{P}_1 = K$.

PROOF: Let $r_{jN} = r_{1jN} + r_{2jN}$, where $r_{1jN} \in \mathscr{P}_1$ and $r_{2jN} \in \mathscr{P}_2$. Lemma 4 and Lemma 1 show that

$$\|r_{2jN}\|^2 \leq \lambda_{K+1}\|r_{2jN}\|_2^2 = \lambda_{k+1}\|r_{jN}\|_2^2$$
$$= \lambda_{K+1}/\lambda_{jN} \to 0$$

as $N \to \infty$ $(j \leq K)$. Since $(p, r_{jN})^2 \leq \|p\|^2$, we have $\sum_{j=1}^K (p, r_{jN}) r_{2jN} \to 0$ as $N \to \infty$. Let $\boldsymbol{H}_N$ be the $K \times K$ matrix whose j,k element is (r_{1jN}, r_{1kN}). Then $r_{2jn} \to 0$ $(j \leq K)$ implies that $\boldsymbol{H}_N \to \boldsymbol{I}_K$ as $N \to \infty$. Hence there is an N^* such that $r_{11N}, \ldots, r_{1KN}$ are linearly independent for $N \geq N^*$. Since $\dim \mathscr{P}_1 \leq K$, we have $\dim \mathscr{P}_1 = K$ and $\{r_{1jN}, j = 1, \ldots, K\}$ is a basis for $\mathscr{P}_1$.

Choose some $N \geq N^*$ and let $\boldsymbol{C} = (c_{jk})$ be a $K \times K$ nonsingular matrix such that $\boldsymbol{C}\boldsymbol{H}_N\boldsymbol{C}' = \boldsymbol{I}_K$. Let $s_j = \sum_{k=1}^K c_{jk} r_{1kN}$. Then $\{s_1, \ldots, s_K\}$ is an orthonormal basis for $\mathscr{P}_1$ and

$$Qp = \sum_{j=1}^K (p, s_j) s_j = \sum_{j,k=1}^K (p, r_{1jN}) h_N^{jk} r_{1kN},$$

where h_N^{jk} is the j,k element of $\boldsymbol{H}_N^{-1}$. Since $\boldsymbol{H}_N \to \boldsymbol{I}_K$ and $r_{2jn} \to 0$ $(j \leq K)$ as $N \to \infty$, we have $Q_N p \to Qp$. *Q.E.D.*

PROOF OF THEOREM 3: (F.i) and (F.ii) follow from Lemmas 4 and 5. Lemma 4 shows that we can set $\bar{\lambda} = \lambda_{K+1}$ in (3.4). *Q.E.D.*

We shall conclude by relating our Condition (F) to an approximate K-factor structure, as defined in (3.5). Suppose that $\lambda_\infty > 0$. If (F) holds, then (3.3) implies

$$\boldsymbol{\Sigma}_N = \boldsymbol{B}_N \boldsymbol{B}_N' + \boldsymbol{R}_N \qquad (N = 1, 2, \ldots), \tag{5.2}$$

where the i, j element of the $N \times K$ matrix $\boldsymbol{B}_N$ is β_{ij} and the i, j element of $\boldsymbol{R}_N$ is $\mathrm{Cov}(e_i, e_j)$. Let $\boldsymbol{\alpha}' = (\alpha_1, \ldots, \alpha_N)$, $\boldsymbol{e}_N' = (e_1, \ldots, e_N)$, and $\boldsymbol{z}_N' = (z_1, \ldots, z_N)$. Since $e_i \in \mathscr{P}_2$, (F) and Lemma 1 imply that

$$\boldsymbol{\alpha}' \boldsymbol{R}_N \boldsymbol{\alpha} = V(\boldsymbol{\alpha}' \boldsymbol{e}_N) \leq \bar{\lambda} \|\boldsymbol{\alpha}' \boldsymbol{e}_N\|_2^2$$
$$= \bar{\lambda} \|\boldsymbol{\alpha}' \boldsymbol{z}_N\|_2^2 = \bar{\lambda} \boldsymbol{\alpha}' \boldsymbol{\alpha};$$

hence the eigenvalues of $\boldsymbol{R}_N$ are uniformly bounded by $\bar{\lambda}$. Thus (F) implies that $\{\boldsymbol{\Sigma}_N\}$ has an approximate K-factor structure if $\lambda_\infty > 0$.

Now let K be the smallest integer such that $\{\boldsymbol{\Sigma}_N\}$ has an approximate K-factor structure. It is shown in [**1**, Proposition 4] that an approximate K-factor structure implies that $\lambda_{K+1} < \infty$. If $\lambda_\infty > 0$, then $\sup_N \lambda_{KN} = \infty$, for otherwise there is an approximate K'-factor structure with $K' < K$ [**1**, Theorem 4]. Hence our Theorem 3 implies that (F) holds. We conclude that (F) is equivalent to an approximate K-factor structure if $\lambda_\infty > 0$.

University of Wisconsin, Madison

Manuscript received June, 1981; final revision received September, 1982.

APPENDIX

PROOF OF LEMMA 1: (i) Let

$$l_2 = \left\{ \alpha = (\alpha_1, \alpha_2, \dots) : \sum_{i=1}^{\infty} \alpha_i^2 < \infty \right\}$$

and define $\|\alpha\|^* = (\sum_{i=1}^{\infty} \alpha_i^2)^{1/2}$. It is well-known that l_2 together with the norm $\|\ \|^*$ form a complete linear space. Let $p_N = \sum_{i=1}^{\infty} \alpha_{iN} x_i$ and $q_N = \sum_{i=1}^{\infty} \beta_{iN} x_i$ be in $\mathscr{F}$, so that $\alpha_{iN} = \beta_{iN} = 0$ for i greater than some integer N^*. Let $\alpha_N = (\alpha_{1N}, \alpha_{2N}, \dots)$ and $\beta_N = (\beta_{1N}, \beta_{2N}, \dots)$. Note that $V(p_N) \geq \lambda_\infty \|\alpha_N\|^{*2}$. If $p_N \to p \in \overline{\mathscr{F}}$, then $V(p_N - p_M) \to 0$ as $N, M \to \infty$. Hence $V(p_N - p_M) \geq \lambda_\infty \|\alpha_N - \alpha_M\|^{*2}$ implies that $\|\alpha_N - \alpha_M\|^* \to 0$ as $N, M \to \infty$. Since l_2 is complete, there is a point $\gamma \in l_2$ such that $\|\alpha_N - \gamma\|^* \to 0$. If q_N also converges to p, then $V(p_N - q_N) \to 0$ implies $\|\alpha_N - \beta_N\|^* \to 0$, and

$$\begin{aligned} \|\alpha_N - \beta_N\|^* &= \|(\alpha_N - \gamma) + (\gamma - \beta_N)\|^* \\ &\geq |(\|\alpha_N - \gamma\|^* - \|\beta_N - \gamma\|^*)| \end{aligned}$$

implies that $\|\beta_N - \gamma\|^* \to 0$. Since $\|\beta_N - \gamma\|^* \geq |(\|\beta_N\|^* - \|\gamma\|^*)|$, it follows that $\|\beta_N\|^* \to \|\gamma\|^*$, and we can define $\|p\|_2 = \|\gamma\|^*$.

If $p_N \to p$, then

$$\begin{aligned} \|\psi p\|_2 &= \lim \|\psi \alpha_N\|^* = |\psi| \lim \|\alpha_N\|^* \\ &= |\psi|\ \|p\|_2 ; \end{aligned}$$

if $p_N \to p$ and $q_N \to q$, then

$$\begin{aligned} \|p + q\|_2 &= \lim \|\alpha_N + \beta_N\|^* \leq \lim(\|\alpha_N\|^* + \|\beta_N\|^*) \\ &= \|p\|_2 + \|q\|_2 . \end{aligned}$$

(ii) If $p_N \in \mathscr{F}$, $p \in \overline{\mathscr{F}}$, and $p_N \to p$, then $V(p_N) \geq \lambda_\infty \|p_N\|_2^2$ and taking the limit of both sides of the inequality gives $V(p) \geq \lambda_\infty \|p\|_2^2$ for $p \in \overline{\mathscr{F}}$. Now suppose that $p_N \in \overline{\mathscr{F}}$ and $p_N \to p$. Then $V(p_N - p) \to 0$ implies $\|p_N - p\|_2^2 \to 0$, so that $\|p_N\|_2 \to \|p\|_2$.

(iii) From the definition of $\mathscr{D}$, we have $\|p\|_2 = 0$ for $p \in \mathscr{D}$. If $p \in \overline{\mathscr{F}}$, then there is a sequence $p_N = \sum_{i=1}^{N} \alpha_{iN} x_i$ with $p_N \to p$. If $\|p\|_2 = 0$, then $\Sigma_{i=1}^{N} \alpha_{iN}^2 \to 0$, so that $p \in \mathscr{D}$. If $\psi \in \mathscr{R}$ and $p, q \in \mathscr{D}$, then

$$0 \leq \|p + \psi q\|_2 \leq \|p\|_2 + |\psi|\ \|q\|_2 = 0$$

implies that $p + \psi q \in \mathscr{D}$; hence $\mathscr{D}$ is a linear subspace.

(iv) $$\|p_1 + p\|_2 \leq \|p_1\|_2 + \|p\|_2 = \|p\|_2$$

by (iii);

$$\begin{aligned} \|p\|_2 &= \|(p_1 + p) - p_1\|_2 \leq \|p_1 + p\|_2 + \|p_1\|_2 \\ &= \|p_1 + p\|_2 ; \end{aligned}$$

hence $\|p_1 + p\|_2 = \|p\|_2$.

(v) If p^* is a riskless limit portfolio, then $p^* \in \mathscr{D}$. Hence (iv) implies that

$$\left\| \sum_{i=1}^{N} \alpha_{iN} z_i \right\|_2 = \left\| \sum_{i=1}^{N} \alpha_{iN} x_i \right\|_2 = \left(\sum_{i=1}^{N} \alpha_{iN}^2 \right)^{1/2}.$$

Then the result follows from (ii).

(vi) If $p \in \overline{\mathscr{P}}$, then there is a sequence $p_N = \sum_{i=1}^{N} \alpha_{iN} z_i$ with $p_N \to p$. If $p \in \mathscr{P}_1$, then $\sum_{i=1}^{N} \alpha_{iN}^2 \to 0$ for some such sequence. Then (v) implies that $\|p\|_2 = 0$. If $\|p\|_2 = 0$ and $\sum_{i=1}^{N} \alpha_{iN} z_i \to p$, then (v) implies that $\sum_{i=1}^{N} \alpha_{iN}^2 \to 0$, so that $p \in \mathscr{P}_1$. $\mathscr{P}_1$ is a linear subspace since it is the intersection of the linear subspaces $\mathscr{D}$ and $\overline{\mathscr{P}}$. *Q.E.D.*

REFERENCES

[1] Chamberlain, G., and M. Rothschild: "Arbitrage, Factor Structure, and Mean-Variance Analysis on Large Asset Markets," *Econometrica*, 51(1983), 1281–1304.

[2] Connor, G.: "Asset Prices in a Well-Diversified Economy," Yale University Technical Paper No. 47, 1980.

[3] Lintner, J.: "The Valuation of Risky Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets," *Review of Economics and Statistics*, 47(1965), 13–37.

[4] Reed, M., and B. Simon: *Methods of Modern Mathematical Physics. I: Functional Analysis*. New York: Academic Press, 1972.

[5] Ross, S. A.: "The Arbitrage Theory of Capital Asset Pricing," *Journal of Economic Theory*, 13(1976), 341–360.

[6] ———: "Return, Risk, and Arbitrage," in *Risk and Return in Finance*, ed. by I. Friend and J. L. Bicksler. Cambridge, Massachusetts: Ballinger, 1977.

[7] ———: "Mutual Fund Separation in Financial Theory—The Separating Distributions," *Journal of Economic Theory*, 17(1978), 254–286.

[8] ———: "On the General Validity of the Mean-Variance Approach in Large Markets," in *Financial Economics: Essays in Honor of Paul Cootner*, ed. by W. Sharpe and C. Cootner. Englewood Cliffs, N.J.: Prentice-Hall, 1982.

[9] Sharpe, W. F.: "Capital Asset Prices: A Theory of Market Equilibrium Under Conditions of Risk," *Journal of Finance*, 19(1964), 425–442.