ECONOMETRICA
JOURNAL OF THE ECONOMETRIC SOCIETY

Arbitrage, Factor Structure, and Mean-Variance Analysis on Large Asset Markets
Author(s): Gary Chamberlain and Michael Rothschild
Source: *Econometrica*, Sep., 1983, Vol. 51, No. 5 (Sep., 1983), pp. 1281-1304
Published by: The Econometric Society
Stable URL: https://www.jstor.org/stable/1912275

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at https://about.jstor.org/terms

The Econometric Society is collaborating with JSTOR to digitize, preserve and extend access to *Econometrica*

JSTOR

This content downloaded from
194.34.233.158 on Wed, 17 Jul 2024 19:28:52 UTC
All use subject to https://about.jstor.org/terms

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

ARBITRAGE, FACTOR STRUCTURE, AND MEAN-VARIANCE ANALYSIS ON LARGE ASSET MARKETS

BY GARY CHAMBERLAIN AND MICHAEL ROTHSCHILD[1]

We examine the implications of arbitrage in a market with many assets. The absence of arbitrage opportunities implies that the linear functionals that give the mean and cost of a portfolio are continuous; hence there exist unique portfolios that represent these functionals. These portfolios span the mean-variance efficient set. We resolve the question of when a market with many assets permits so much diversification that risk-free investment opportunities are available.

Ross [**12, 14**] showed that if there is a factor structure, then the mean returns are approximately linear functions of factor loadings. We define an *approximate factor structure* and show that this weaker restriction is sufficient for Ross' result. If the covariance matrix of the asset returns has only K unbounded eigenvalues, then there is an approximate factor structure and it is unique. The corresponding K eigenvectors converge and play the role of factor loadings. Hence only a principal component analysis is needed in empirical work.

1. INTRODUCTION

TWO OF THE MOST SIGNIFICANT DEVELOPMENTS in finance have been the formulation of the capital asset pricing model (CAPM) and the working out of the implications of arbitrage beginning with the Modigliani–Miller Theorem and culminating in the theory of option pricing. While the principle that competitive markets do not permit profitable arbitrage opportunities to remain unexploited seems unexceptionable, the same cannot be said for the crucial assumptions of the CAPM. Few believe that asset returns are well described by their first two moments or that some well-defined set of marketable assets contains most of the investment opportunities available to individual investors. Casual observation is sufficient to refute one of the main implications of the CAPM—that everyone holds the market portfolio. Nonetheless, the CAPM seems to do a good job of explaining relationships among asset prices. Ross [**12, 14**] has argued that the apparent empirical success of the CAPM is due to three assumptions which are more plausible than the assumptions needed to derive the CAPM. These assumptions are first, that there are many assets; second, that the market permits no arbitrage opportunities; and third, that asset returns have a factor structure with a small number of factors.[2] Ross presents a heuristic argument which suggests that on a market with an infinite number of assets, there are sufficiently many

[1] We are indebted to Bob Anderson, Steve Ross, Jose Scheinkman, and the editors for helpful suggestions, and to the National Science Foundation and the University of Wisconsin Graduate School for research support. Chamberlain held an Alfred P. Sloan Research Fellowship and Rothschild held the Oskar Morgenstern Distinguished Fellowship at Mathematica while some of the work on this paper was done. The research reported here is part of the NBER's program in Financial Markets and Monetary Economics. Any opinions expressed are those of the authors and not those of the National Bureau of Economic Research.

[2] That is, the covariance matrix of asset returns may be written as the sum of a diagonal matrix and a matrix of short rank as in equation (1.2) below.

This content downloaded from 194.34.233.158 on Wed, 17 Jul 2024 19:28:52 UTC
All use subject to https://about.jstor.org/terms

riskless portfolios that prices of assets are determined by an arbitrage requirement—riskless portfolios which require no net investment should not have a positive return. Asset prices are linear functions of factor loadings. Although Ross' heuristics cannot be made rigorous, he does prove that lack of arbitrage implies that asset prices are approximately linear functions of factor loadings, and Chamberlain **[3]** and Connor **[4]** have given conditions under which the conclusions of Ross' heuristic argument are precisely true.[3] Nonetheless, all of Ross' investigations of the implications of the absence of arbitrage opportunities take place in the context of a factor structure. Furthermore, Ross' definition of a factor structure is sufficiently stringent that it is unlikely that any large asset market has, by his definition, a usefully small number of factors.

This paper has two purposes: The first is to examine the implications of the absence of arbitrage opportunities on a market with many assets which does not necessarily have a factor structure. We show in Sections 2 and 3 that an asset market with countably many assets has a natural Hilbert space structure which makes it easy to examine the implications of the no arbitrage condition. Our second goal is to define an *approximate factor structure*—a concept which is weaker than the standard strict factor structure which Ross uses. We show in Sections 4 and 5 that this is an appropriate concept for investigating the relationship between factor loadings and asset prices.

In Section 2 we introduce our model of the asset market. We consider a market on which a countable number of assets are traded. As is customary in investigations of this sort, we take a given price system and ask if it could possibly be an equilibrium price system. Since prices are fixed we normalize by assuming each asset costs one dollar. For a dollar an investor may purchase a random return with a specified distribution.

The assets on the market may be arranged in a sequence. The first two moments of the joint distribution of returns of the first N assets are a mean vector $\boldsymbol{\mu}_N$ and a covariance matrix $\boldsymbol{\Sigma}_N$. In this paper we often look at what happens to various objects (such as the mean-variance efficiency frontier or the eigenvalues of $\boldsymbol{\Sigma}_N$) as N increases to infinity. Such limits have meaning, in part, because our model of the asset market may be embedded in a Hilbert space. In Section 2 we list some of the basic facts about Hilbert space which we use.

Section 3 defines the absence of arbitrage opportunities and explores the implications of the definition. Our definition, essentially the same as Ross', is that it should not be possible to form a portfolio which is riskless, costless, and earns a positive return. If prevailing prices permit such a portfolio to be formed, investors, at least those whose preferences satisfy some weak conditions, will want to buy arbitrarily large amounts of that portfolio; consequently the prevailing prices cannot be equilibrium prices.

[3] Chamberlain requires that there be a mean-variance efficient portfolio which is so well-diversified that it contains only factor variance and no idiosyncratic variance. Connor requires that the supply of the assets be well-diversified.

This content downloaded from
194.34.233.158 on Wed, 17 Jul 2024 19:28:52 UTC
All use subject to https://about.jstor.org/terms

There is a close link between the absence of arbitrage opportunities and mean-variance analysis. If the asset market permits arbitrage opportunities then investors do not have to choose between mean and variance. They can for a given price acquire portfolios which have arbitrarily high expected returns and arbitrarily low variances. If market prices do not permit arbitrage, investors must choose between mean and variance. An object of considerable interest on an asset market without arbitrage opportunities is the mean-variance efficient set. This is the set of all portfolios for which variance is at a minimum subject to constraints on cost and expected return. One of the reasons the mean-variance efficient set is of such interest is that Roll [**10**] and Ross [**13**] have shown that the CAPM is equivalent to the statement that the market portfolio is mean-variance efficient. We show that on a market with an infinite number of assets the mean-variance efficient set is the same kind of object as on a market with a finite number of assets. In each case the mean-variance efficient set is contained in a particular two-dimensional subspace.[4]

For portfolios of a given cost which are efficient, there is a linear tradeoff between mean and standard deviation. We call the slope of this tradeoff δ. The constant δ will play an important role in our analysis of factor structure; δ is also the distance, in a certain norm, between the vector of mean returns from each asset and a vector of ones.

Our model of the asset market assumes that all of the assets on the market are risky. We investigate the question of whether investors, by allocating their purchases among many assets, can create a portfolio that is riskless, costs a dollar, and has a positive return. If the answer to this question is yes, then we say there is a riskless asset. It is commonly believed that if all assets are affected by the same random event, the market will not allow investors to diversify risks so effectively that they can create a riskless portfolio with a positive return. Our necessary and sufficient condition for the existence of a riskless asset sharpens this intuition. A riskless asset will exist unless the sequence of covariance matrices has the same structure as it would have if there were a random event which affected the returns of all assets in precisely the same way. If there is a riskless asset, then the mean-variance efficiency frontier must be a straight line in mean-standard deviation space—not the curve that is usually drawn.

Sections 4 and 5 explore the relationship between factor structure and asset pricing. We say the asset market has a strict K-factor structure if the return on the ith asset is generated by

$$x_i = \mu_i + \beta_{i1} f_1 + \cdots + \beta_{iK} f_K + v_i, \tag{1.1}$$

where μ_i is the mean return on asset i and the factors f_k are uncorrelated with the idiosyncratic disturbances v_i, which are in turn uncorrelated with each other. An implication of (1.1) is that the covariance matrix may be decomposed into a

[4] See [**2** and **10**] for the case with a finite number of assets.

This content downloaded from
194.34.233.158 on Wed, 17 Jul 2024 19:28:52 UTC
All use subject to https://about.jstor.org/terms

matrix of rank K and a diagonal matrix. That is, for any N,

$$\mathbf{\Sigma}_N = \boldsymbol{B}_N \boldsymbol{B}_N' + \boldsymbol{D}_N, \tag{1.2}$$

where $\boldsymbol{B}_N$ is the $N \times K$ matrix of factor loadings and $\boldsymbol{D}_N$ is a diagonal matrix. Ross proved that if (1.1) holds, then asset means are approximately linear functions of factor loadings. If there is one factor ($K = 1$) and a riskless asset with a return of ρ, then Ross' conclusion may be stated as

$$\mu_i \approx \rho + \tau_1 \beta_{i1},$$

which is almost indistinguishable from the CAPM pricing equation.[5] In Section 4 we extend this result by showing that the same conclusion holds if there is a sequence $\{\beta_{i1}, \ldots, \beta_{iK}\}_{i=1}^{\infty}$ such that for any N,

$$\mathbf{\Sigma}_N = \boldsymbol{B}_N \boldsymbol{B}_N' + \boldsymbol{R}_N, \tag{1.3}$$

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 matrices with uniformly bounded eigenvalues.

If condition (1.3) is satisfied, then we say that the market has an *approximate K-factor structure*. In Section 5 we characterize approximate factor structures. The idea which decompositions like (1.1) and (1.2) are meant to convey is that for all practical purposes the stochastic structure of asset returns is determined by a small number (in this case K) of things; everything else is inessential and may be ignored. Our characterization captures this notion. Since the rank of $\boldsymbol{B}_N \boldsymbol{B}_N'$ in (1.3) is no more than K, the $K + 1$st eigenvalue of $\mathbf{\Sigma}_N$ is smaller than the largest eigenvalue of $\boldsymbol{R}_N$ and is thus bounded. An asset market has an approximate K-factor structure if and only if exactly K of the eigenvalues of the covariance matrices $\mathbf{\Sigma}_N$ increase without bound and all other eigenvalues are bounded.[6]

The concept of approximate factor structure is useful for exploring the theoretical relationship between asset prices and factor loadings. It should also prove to be a useful tool for examining this relationship empirically. If there is an approximate factor structure, then mean returns are approximately linear functions of the β's. The approximation error (that is, the sum of squared deviations) is bounded by the product of the constant δ^2 and the $K + 1$st largest eigenvalue of $\mathbf{\Sigma}_N$. The eigenvectors corresponding to the exploding eigenvalues converge to factor loadings (in the sense that one can use the eigenvectors to approximate the matrix $\boldsymbol{B}_N \boldsymbol{B}_N'$ of (1.3) arbitrarily well). Furthermore, the approximate factor structure is unique in the following sense: Suppose that there is a nested sequence

[5] Since all assets cost a dollar, a formula which explains the mean return of the ith asset *is* an asset pricing formula; it determines the mean return per dollar spent on asset i. If there is a riskless asset with rate of return ρ, then $\mu_i - \rho$ is the risk premium which investors get if they buy asset i.

[6] We assume also that the smallest eigenvalue of $\mathbf{\Sigma}_N$ is uniformly bounded away from zero for all N.

This content downloaded from
194.34.233.158 on Wed, 17 Jul 2024 19:28:52 UTC
All use subject to https://about.jstor.org/terms

of $N \times K$ matrices $\{G_N\}$ such that[7]

$$\boldsymbol{\Sigma}_N = \boldsymbol{G}_N \boldsymbol{G}_N' + \boldsymbol{W}_N$$

and the eigenvalues of $\{W_N\}$ are uniformly bounded. Then $\boldsymbol{G}_N \boldsymbol{G}_N' = \boldsymbol{B}_N \boldsymbol{B}_N'$ and $\boldsymbol{W}_N = \boldsymbol{R}_N$.

These results suggest that extracting the eigenvectors of $\boldsymbol{\Sigma}_N$ is as good a way as any of finding approximate factor structures. Thus, principal component analysis, which is computationally and conceptually simpler than factor analysis, is an appropriate technique for finding an approximate factor structure.[8] A common objection to principal component analysis is that it is arbitrary to examine the eigenvectors of $\boldsymbol{\Sigma}_N$ relative to an identity matrix rather than relative to some other positive-definite matrix—one which is in some sense more natural for the problem at hand. We show that for the problem of investigating the approximate factor structure of an asset market this objection is groundless. Since the approximate factor structure is unique, all positive-definite matrices lead to the same approximate factor structure.

2. THE HILBERT SPACE SETTING

We examine a market in which there are an infinite number of assets. 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

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

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:

$$(p, q) = E(pq) = \mathrm{Cov}(p, q) + E(p)E(q),$$

[7] Let $\{G_N\}$ be a sequence of matrices where G_N has r_N rows and c_N columns. Then $\{G_N\}$ is a nested sequence if the r_N by c_N upper left-hand submatrix of G_{N+1} is G_N. Clearly $\{B_N\}$ and $\{\Sigma_N\}$ (of both (1.2) and (1.3)) are nested.

[8] Roll and Ross [**11**] have used factor analysis in their empirical work on arbitrage pricing.

This content downloaded from
194.34.233.158 on Wed, 17 Jul 2024 19:28:52 UTC
All use subject to https://about.jstor.org/terms

with the associated norm:

$$\|p\| = \left(E(p^2)\right)^{1/2} = \left(V(p) + (E(p))^2\right)^{1/2},$$

for $p, q \in L_2(P)$. Since $\mathscr{F}$ is a linear subspace of $L_2(P)$, its closure, $\overline{\mathscr{F}}$, is also a Hilbert space. If $p \in \overline{\mathscr{F}}$, then there is a sequence $\{p_N\}$ in $\mathscr{F}$ with $E((p_N - p)^2) \to 0$ as $N \to \infty$. So there are finite portfolios whose random returns are arbitrarily good approximations to p.

Let $\boldsymbol{x}_N' = (x_1, \ldots, x_N)$ and let $\boldsymbol{\Sigma}_N$ be the covariance matrix of $\boldsymbol{x}_N$. 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., $E((p-q)^2) = 0$. If $p = q$, $V(p - q) = 0$, so 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 often identify $p \in \mathscr{F}$ with its (unique) associated portfolio.

Note that $C(\ )$ is a linear functional on $\mathscr{F}$. In Section 3 we shall extend the definition of $C(\ )$ to $\overline{\mathscr{F}}$, and we shall relate the linear functionals $E(\ )$ and $C(\ )$ to the mean-variance frontier. This will require the following two basic properties of a Hilbert space:[9]

Projection Theorem: *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., $(p_2, q) = 0$ for every $q \in \mathscr{G}$).*

Riesz Representation Theorem: *If L is a continuous linear functional on a Hilbert space $\mathscr{H}$, then there is a unique $q \in \mathscr{H}$ such that $L(p) = (q, p)$ for every $p \in \mathscr{H}$.*

The projection theorem is often used together with the fact that every finite dimensional subspace is closed. We shall also use the following two elementary properties of linear functionals:

If $\mathscr{G}$ is a linear subspace of a Hilbert space $\mathscr{H}$, then a linear functional L is continuous on $\mathscr{G}$ if and only if $L(p_N) \to 0$ for every sequence $\{p_N\}$ in $\mathscr{G}$ that converges to zero.

If $\mathscr{G}$ is a linear subspace of a Hilbert space $\mathscr{H}$ and if the linear functional L is continuous on $\mathscr{G}$, then there is a unique continuous linear functional on the closure of $\mathscr{G}$ that coincides with L on $\mathscr{G}$.

[9] See, for example [**9**, Chapters I and II].

This content downloaded from
194.34.233.158 on Wed, 17 Jul 2024 19:28:52 UTC
All use subject to https://about.jstor.org/terms

3. ARBITRAGE OPPORTUNITIES AND MEAN-VARIANCE EFFICIENCY

3.1. *Lack of Arbitrage Opportunities*

We now consider what it means for there to be no arbitrage opportunities on the asset market. By defining x_i as the return available for one dollar, we have assumed prices are determined. These prices can be equilibrium prices if no trader would want to make an infinitely large trade. We define the absence of arbitrage opportunities in terms of conditions which, if they failed, would make some risk-averse traders want to take infinitely large positions.

Let $\{p_N\}$ be a sequence of finite portfolios ($p_N \in \mathscr{F}$). Then 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) simply states that it is not possible to make an investment that is costless, riskless, and yields a positive return. Ross **[12]** has shown that if Condition (A.i) fails, many (but not all) risk-averse traders will want to take infinitely large positions. A similar argument justifies Condition (A.ii). Suppose that (A.ii) does not hold; that is, suppose that the market allows investors to trade a portfolio that, approximately, costs a dollar and has a riskless, nonpositive return. Then investors face no budget constraints; by selling this portfolio short they can generate arbitrarily large amounts of cash which can be used to purchase investments or, in a complete model, for current consumption, while incurring no future obligations. In fact, if $\alpha < 0$, then investors could consume infinite amounts both now and in the future without risk.

3.2. *Mean-Variance Efficiency*

Roll **[10]** and Ross **[13]** have shown that the empirical content of the capital asset pricing model is contained in the observation that the market portfolio is on the mean-variance efficiency frontier. If arbitrage opportunities exist on an infinite market, then there is no tradeoff between mean and variance; there exist costless finite portfolios with arbitrarily large means and arbitrarily small variances. If (A) does hold, there is a well-defined tradeoff between mean and variance. The mean-variance efficient set has the same structure in our infinite market as on any finite market. In each case it lies in the subspace generated by the (limit) portfolios that represent the linear functionals $E(\ )$ and $C(\ )$.

To prove this, we must show first that $E(\ )$ and $C(\ )$ are continuous. Clearly $E(\ )$ is continuous since $\|p\|^2 = V(p) + (E(p))^2$. Thus if $\|p_N\| \to 0$, then $E(p_N) \to 0$. The continuity of $C(\ )$ follows from Condition (A.ii).[10] Suppose $\|p_N\| \to 0$

[10]The relationship between arbitrage and the continuity of price functionals is examined in a more general setting by Kreps **[6]**.

This content downloaded from
194.34.233.158 on Wed, 17 Jul 2024 19:28:52 UTC
All use subject to https://about.jstor.org/terms

but $C(p_N)$ does not converge to zero. Then there is an $\epsilon > 0$ and a subsequence $\{p'_N\}$ with $|C(p'_N)| \geq \epsilon$. Let $q_N = p'_N / C(p'_N)$. Then along the subsequence we have $C(q_N) = 1$ and

$$\|q_N\| = \|p'_N\| / |C(p'_N)| \leq \|p'_N\| / \epsilon \to 0.$$

Thus $E(q_N)$ converges to zero, which contradicts Condition (A.ii). This contradiction proves the following proposition.

PROPOSITION 1: *Condition* (A.ii) *implies that* $C(\ )$ *is continuous.*

Hence we can extend $C(\ )$ to a continuous linear functional on $\overline{\mathscr{F}}$. Since the cost of p is now well-defined when $p \in \overline{\mathscr{F}}$, we shall refer to these random returns as limit portfolios. It follows from Riesz' theorem that there exist limit portfolios m and c in $\overline{\mathscr{F}}$ that represent $E(\ )$ and $C(\ )$ in the sense that

$$E(p) = (m, p), \qquad C(p) = (c, p)$$

for all $p \in \overline{\mathscr{F}}$. The following theorem shows that the mean-variance efficient set is generated by m and c.

THEOREM 1: *Suppose that* (A.ii) *holds. 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* $p \in \overline{\mathscr{F}}$, $E(p) = E(q)$, $C(p) = C(q)$.

PROOF: Since $q = p^0 + e$, where $e \in [m, c]^{\perp}$, we have $E(q) = E(p^0)$ and $C(q) = C(p^0)$. Let p be any limit portfolio satisfying $E(p) = E(q)$ and $C(p) = C(q)$. Then since $(p - p^0) \perp p^0$, $\|p\|^2 = \|p^0\|^2 + \|p - p^0\|^2$. Thus, $E(p) = E(p^0)$ implies that

$$V(p) - V(p^0) = \|p\|^2 - \|p^0\|^2 = \|p - p^0\|^2 \geq 0. \qquad \textit{Q.E.D.}$$

COROLLARY 1: *Suppose that* (A) *holds and define*

$$\delta = \sup |E(p)| / V^{1/2}(p)$$

subject to $p \in \mathscr{F}$, $C(p) = 0$, $p \neq 0$. *Define*

$$\psi = (m, c)/(c, c), \qquad h = m - \psi c.$$

If $h \neq 0$, *then* $C(h) = 0$, $V(h) > 0$, *and*

$$\delta = |E(h)| / V^{1/2}(h); \tag{3.1}$$

if $h = 0$, *then* $\delta = 0$.

PROOF: If $h = 0$, then $E(p) = 0$ whenever $C(p) = 0$, and so $\delta = 0$. Suppose that $h \neq 0$. In that case, $E(h) \neq 0$, for otherwise $(m, h) = 0$, $(c, h) = 0$, and

This content downloaded from
194.34.233.158 on Wed, 17 Jul 2024 19:28:52 UTC
All use subject to https://about.jstor.org/terms

$h \in [m,c]$ imply that $h = 0$. By (A.i), if $E(p) \neq 0$ and $C(p) = 0$, then $V(p) > 0$. If $C(p) = 0$, $p \neq 0$, and

$$|E(p)|/V^{1/2}(p) > |E(h)|/V^{1/2}(h),$$

then $p^* = (E(h)/E(p))p$ has $E(p^*) = E(h)$, $C(p^*) = C(h) = 0$, and $V(p^*) < V(h)$. This contradicts Theorem 1 and completes our proof. *Q.E.D.*

The parameter δ gives the slope of the tradeoff between mean and risk (measured by standard deviation) along the efficient frontier; $[h]$ is the linear subspace of costless portfolios which are efficient. An investor can increase risk in an efficient manner by adding a hedge portfolio from $[h]$ to his holdings. Another way of making this point is to observe that if $p, q \in [m,c]$ with $C(p) = C(q) = 1$, then $p - q \in [h]$ and so (3.1) implies

$$|E(p) - E(q)| = \delta V^{1/2}(p-q). \tag{3.2}$$

We shall see that δ plays an important role in our treatment of factor models.

3.3. *Riskless Asset*

In this section we shall examine the implications of the existence of a riskless limit portfolio.

DEFINITION 1: 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 Condition (A.ii) holds, then $C(p^*)$ is well-defined and $C(p^*) = 0$ violates Condition (A.i). Hence we can set $s = p^*/C(p^*)$. We shall refer to s as a *riskless asset*. If there is an $s' \in \overline{\mathscr{F}}$ with $C(s') = 1$ and $V(s') = 0$, then $C(s - s') = 0$, $V(s - s') = 0$, and Condition (A.i) implies $E(s - s') = 0$. So $s = s'$ and the riskless asset is unique. Let $\rho = E(s)$ be the return on the riskless asset; Condition (A.ii) implies that $\rho > 0$.

Note that $(s/\rho, p) = E(p)$ for all $p \in \overline{\mathscr{F}}$; hence $m = s/\rho$. If $p \in [m,c]$ and $C(p) = 1$, then setting $q = s$ in (3.2) gives the following tradeoff between mean and risk along the efficient frontier:

$$|E(p) - \rho| = \delta V^{1/2}(p). \tag{3.3}$$

Thus, if there is a riskless asset, the frontier of the efficient set (in (μ, σ) space) is a straight line.

We now develop a necessary and sufficient condition for the existence of a riskless asset. We also show that if there is no riskless limit portfolio, then the covariance is a natural inner product for the space $\overline{\mathscr{F}}$. We use this construction in Section 5. Suppose that there is no riskless limit portfolio. Then $E(m) \neq 1$; for

This content downloaded from
194.34.233.158 on Wed, 17 Jul 2024 19:28:52 UTC
All use subject to https://about.jstor.org/terms

otherwise

$$E(m) = (m,m) = V(m) + (E(m))^2$$

implies that $V(m) = 0$, a contradiction. If $E(m) \neq 1$, then

$$E(p) = (m, p) = \mathrm{Cov}(m, p) + E(m)E(p)$$

implies that

$$E(p) = \mathrm{Cov}(m, p)/(1 - E(m)) = \mathrm{Cov}(m^*, p),$$

where $m^* = m/(1 - E(m))$. So we can generate the mean functional from the covariance with m^*. If (A) holds, we can also use covariance to generate the cost functional:

$$C(p) = (c, p) = \mathrm{Cov}(c, p) + E(c)E(p)$$
$$= \mathrm{Cov}(c, p) + E(c)\mathrm{Cov}(m^*, p) = \mathrm{Cov}(c^*, p),$$

where

$$c^* = c + E(c)m^*.$$

Now we have

$$\|p\|^2 = V(p) + (\mathrm{Cov}(m^*, p))^2 \leq (1 + V(m^*))V(p) \qquad (p \in \overline{\mathscr{F}}),$$

so that $V(p) = 0$ implies $p = 0$. Hence Cov(,) is an inner product and $V^{1/2}(\ )$ is a norm. Furthermore, if $\{p_N\} \in \overline{\mathscr{F}}$ and $V(p_N - p_M) \to 0$ as $N, M \to \infty$, then $\|p_N - p_M\| \to 0$ and, since $\overline{\mathscr{F}}$ is complete under the mean-square norm, there is a $p \in \overline{\mathscr{F}}$ with $\|p_N - p\| \to 0$. Hence $V(p_N - p) \to 0$ so that $\overline{\mathscr{F}}$ is complete under the variance norm, and $\overline{\mathscr{F}}$ together with the covariance inner product forms a Hilbert space. Condition (A) is not needed for this result.

Note that the span of m^*, c^* is identical to the span of m, c. Let $\psi^* = \mathrm{Cov}(m^*, c^*)/V(c^*)$ and $h^* = m^* - \psi^* c^*$. Then $h^* \in [h]$ since $h^* \in [m, c]$ and $C(h^*) = 0$. If $h \neq 0$, then $h^* \neq 0$ and

$$\delta = |E(h^*)|/V^{1/2}(h^*) = V^{1/2}(h^*). \tag{3.4}$$

We can use the covariance inner product to characterize the existence of a riskless asset. If (A) holds and there is no riskless asset, then we can form the (covariance) orthogonal projection of x_i onto c^*:

$$x_i = \tau c^* + w_i \qquad (i = 1, 2, \ldots),$$

where $\tau = \mathrm{Cov}(x_i, c^*)/V(c^*) = 1/V(c^*)$ and $C(w_i) = \mathrm{Cov}(c^*, w_i) = 0$. Let $\boldsymbol{l}_N$ be an $N \times 1$ vector of ones and let $\boldsymbol{w}_N' = (w_1, \ldots, w_N)$. Then

$$\boldsymbol{\Sigma}_N = \tau \boldsymbol{l}_N \boldsymbol{l}_N' + V(\boldsymbol{w}_N) \qquad (N = 1, 2, \ldots), \tag{3.5}$$

This content downloaded from
194.34.233.158 on Wed, 17 Jul 2024 19:28:52 UTC
All use subject to https://about.jstor.org/terms

so that $\{\boldsymbol{\Sigma}_N\}$ has an equicorrelated component. We show now that (3.5) is also a sufficient condition for there to be no riskless asset.

PROPOSITION 2: *Suppose that* (A) *holds. Then there is no riskless asset if and only if there is a* $\varphi > 0$ *such that*

$$\boldsymbol{\Sigma}_N - \varphi \boldsymbol{l}_N \boldsymbol{l}_N'$$

is positive semi-definite for $N = 1, 2, \ldots$.

PROOF: We have already seen that this condition is necessary. Suppose that the condition holds and that there is a riskless asset s. Then there is a sequence $p_N = \boldsymbol{\alpha}_N' \boldsymbol{x}_N$ with $V(p_N) = \boldsymbol{\alpha}_N' \boldsymbol{\Sigma}_N \boldsymbol{\alpha}_N \to 0$ and $C(p_N) = \boldsymbol{\alpha}_N' \boldsymbol{l}_N \to 1$. But

$$\boldsymbol{\alpha}_N' \boldsymbol{\Sigma}_N \boldsymbol{\alpha}_N - \varphi (\boldsymbol{\alpha}_N' \boldsymbol{l}_N)^2 \geq 0$$

implies that $\boldsymbol{\alpha}_N' \boldsymbol{l}_N \to 0$. This contradiction completes the proof. *Q.E.D.*

The equicorrelated component condition is quite stringent; it is not enough for the assets $\{x_1, x_2, \ldots\}$ all to be positively correlated with the same factor. Suppose, for example, that

$$\begin{aligned} x_i &= \alpha + \beta f + v_i \qquad && (i \text{ odd}), \\ &= f + v_i && (i \text{ even}), \end{aligned}$$

where $\alpha > 0$, $0 < \beta < 1$, and all of the v_i are zero-mean, uncorrelated random variables with uniformly bounded variances. If we invest $1/N$ in the first N odd assets and $-\beta/N$ in the first N even assets, then net investment is $1 - \beta$ and the random return is

$$\alpha + N^{-1} \sum_{i=1}^{N} (v_{2i-1} - \beta v_{2i}),$$

which converges in mean-square to α. Thus, there is a riskless asset.

3.4. *A Construction of* δ

We give a construction of δ that will be used in our treatment of factor models. First we need to develop some concepts and results from least squares theory which we will use again. Define

$$(3.6) \qquad \|\boldsymbol{y}\|_{\boldsymbol{Q}} = (\boldsymbol{y}' \boldsymbol{Q} \boldsymbol{y})^{1/2}, \qquad \|\boldsymbol{y} - [\boldsymbol{A}]\|_{\boldsymbol{Q}} = \inf_{\boldsymbol{\tau}} \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{\tau}\|_{\boldsymbol{Q}},$$

where $\boldsymbol{y}$ is a $N \times 1$ vector, $\boldsymbol{Q}$ is a $N \times N$ positive semi-definite matrix, and $\boldsymbol{A}$ is a $N \times J$ matrix. Let $\boldsymbol{x}_N' = (x_1, \ldots, x_N)$ and $\boldsymbol{\mu}_N' = (\mu_1, \ldots, \mu_N)$.

This content downloaded from
194.34.233.158 on Wed, 17 Jul 2024 19:28:52 UTC
All use subject to https://about.jstor.org/terms

LEMMA 1: *There is a* $\boldsymbol{\tau}^*$ *that achieves the infinum in* (3.6). *If* $\boldsymbol{A}$ *has full column rank and* $\boldsymbol{Q}$ *is positive definite, then* $\boldsymbol{\tau}^* = (\boldsymbol{A}'\boldsymbol{Q}\boldsymbol{A})^{-1}\boldsymbol{A}'\boldsymbol{Q}\boldsymbol{y}$ *and*

$$\begin{aligned}\|\boldsymbol{y}-[\boldsymbol{A}]\|_{\boldsymbol{Q}}^2 &= (\boldsymbol{y}-\boldsymbol{A}\boldsymbol{\tau}^*)'\boldsymbol{Q}(\boldsymbol{y}-\boldsymbol{A}\boldsymbol{\tau}^*)\\ &= \boldsymbol{y}'\big(\boldsymbol{Q}-\boldsymbol{Q}\boldsymbol{A}(\boldsymbol{A}'\boldsymbol{Q}\boldsymbol{A})^{-1}\boldsymbol{A}'\boldsymbol{Q}\big)\boldsymbol{y}.\end{aligned}$$

PROOF: There is an $N \times N$ matrix $\boldsymbol{C}$ such that $\boldsymbol{Q} = \boldsymbol{C}\boldsymbol{C}'$. Let $\boldsymbol{z} = \boldsymbol{C}'\boldsymbol{y}$ and $\boldsymbol{G} = \boldsymbol{C}'\boldsymbol{A}$, so that

$$\|\boldsymbol{y}-\boldsymbol{A}\boldsymbol{\tau}\|_{\boldsymbol{Q}}^2 = \|\boldsymbol{z}-\boldsymbol{G}\boldsymbol{\tau}\|_{\boldsymbol{I}}^2.$$

If we use $\boldsymbol{a}'\boldsymbol{b}$ as the inner product of $\boldsymbol{a}$ and $\boldsymbol{b}$, then the projection theorem asserts that there is a vector $\hat{\boldsymbol{z}} = \boldsymbol{G}\boldsymbol{\tau}^*$ such that $\boldsymbol{G}'(\boldsymbol{z}-\hat{\boldsymbol{z}}) = \boldsymbol{0}$. Hence

$$\begin{aligned}\|\boldsymbol{z}-\boldsymbol{G}\boldsymbol{\tau}\|_{\boldsymbol{I}}^2 &= \|\boldsymbol{z}-\hat{\boldsymbol{z}}\|_{\boldsymbol{I}}^2 + \|\hat{\boldsymbol{z}}-\boldsymbol{G}\boldsymbol{\tau}\|_{\boldsymbol{I}}^2\\ &\geq \|\boldsymbol{z}-\hat{\boldsymbol{z}}\|_{\boldsymbol{I}}^2.\end{aligned}$$

In the full rank case, we have

$$\boldsymbol{\tau}^* = (\boldsymbol{G}'\boldsymbol{G})^{-1}\boldsymbol{G}'\boldsymbol{z} = (\boldsymbol{A}'\boldsymbol{Q}\boldsymbol{A})^{-1}\boldsymbol{A}'\boldsymbol{Q}\boldsymbol{y}. \qquad \textit{Q.E.D.}$$

PROPOSITION 3: *Suppose that* (A) *holds and define*

$$\delta_N = \sup|E(p)|/V^{1/2}(p)$$

subject to $p \in \mathscr{F}_N$, $C(p)=0$, $p \neq 0$. *Then*

$$\delta_N = \|\boldsymbol{\mu}_N - [\boldsymbol{l}_N]\|_{\boldsymbol{\Sigma}_N^{-1}}$$

and $\{\delta_N\}$ *is a nondecreasing sequence which converges to* δ *as* $N \to \infty$. *If there is a riskless asset with a return of* ρ, *then*

$$\{\|\boldsymbol{\mu}_N - \rho\boldsymbol{l}_N\|_{\boldsymbol{\Sigma}_N^{-1}}\}$$

is a nondecreasing sequence which also converges to δ *as* $N \to \infty$.

PROOF: Choose a nonsingular matrix $\boldsymbol{C}$ such that $\boldsymbol{C}'\boldsymbol{C} = \boldsymbol{\Sigma}_N$. Let $\tau = (\boldsymbol{\mu}_N'\boldsymbol{\Sigma}_N^{-1}\boldsymbol{l}_N)/(\boldsymbol{l}_N'\boldsymbol{\Sigma}_N^{-1}\boldsymbol{l}_N)$. If $p = \boldsymbol{\alpha}'\boldsymbol{x}_N$ and $C(p)=0$, then $\boldsymbol{l}_N'\boldsymbol{\alpha} = 0$ and

$$\begin{aligned}(3.7)\qquad |E(p)| &= |(\boldsymbol{\mu}_N - \tau\boldsymbol{l}_N)'\boldsymbol{\alpha}| = |(\boldsymbol{\mu}_N - \tau\boldsymbol{l}_N)'\boldsymbol{C}^{-1}\boldsymbol{C}\boldsymbol{\alpha}|\\ &\leq \|\boldsymbol{\mu}_N - [\boldsymbol{l}_N]\|_{\boldsymbol{\Sigma}_N^{-1}} V^{1/2}(p),\end{aligned}$$

by Lemma 1 and the Cauchy–Schwarz inequality; equality holds in (3.7) if $p = (\boldsymbol{\mu}_N - \tau\boldsymbol{l}_N)'\boldsymbol{\Sigma}_N^{-1}\boldsymbol{x}_N$. Hence $\delta_N = \|\boldsymbol{\mu}_N - [\boldsymbol{l}_N]\|_{\boldsymbol{\Sigma}_N^{-1}}$. $\{\delta_N\}$ is nondecreasing since

This content downloaded from
194.34.233.158 on Wed, 17 Jul 2024 19:28:52 UTC
All use subject to https://about.jstor.org/terms

$\mathscr{F}_N \subset \mathscr{F}_{N+1}$. The limit of $\{\delta_N\}$ as $N \to \infty$ is $\sup|E(p)|/V^{1/2}(p)$ subject to $p \in \mathscr{F}$, $C(p) = 0$, $p \neq 0$; this is the definition of δ.

If there is a riskless asset s, define $z_i = x_i - s$ and note that $E(z_i) = \mu_i - \rho$, $C(z_i) = 0$. Let $\mathscr{G}_N$ be the span of $\{z_1, \ldots, z_N\}$ and $\mathscr{G} = \bigcup_{N=1}^{\infty} \mathscr{G}_N$. Then

$$\gamma_N \equiv \left((\boldsymbol{\mu}_N - \rho \boldsymbol{l}_N)' \boldsymbol{\Sigma}_N^{-1} (\boldsymbol{\mu}_N - \rho \boldsymbol{l}_N) \right)^{1/2} = \sup|E(p)|/V^{1/2}(p)$$

subject to $p \in \mathscr{G}_N$, $p \neq 0$. $\{\gamma_N\}$ is nondecreasing since $\mathscr{G}_N \subset \mathscr{G}_{N+1}$. The limit of $\{\gamma_N\}$ as $N \to \infty$ is $\sup|E(p)|/V^{1/2}(p)$ subject to $p \in \mathscr{G}$, $p \neq 0$. Since $p \in \mathscr{G}$ implies that $p \in \overline{\mathscr{F}}$ and $C(p) = 0$, we have $\gamma = \lim \gamma_N \leq \delta$. Since $\delta_N \leq \gamma_N$, we have $\delta \leq \gamma$. Hence $\gamma = \delta$. *Q.E.D.*

4. FACTOR STRUCTURE AND ROSS' THEOREM

4.1. *Strict Factor Structure*

The phenomenon that a factor structure tries to capture is that the covariance matrix $\boldsymbol{\Sigma}_N$ can be approximated by a simpler, lower dimensional structure. We shall say that there is a strict K-factor structure if the return on the ith asset is generated by

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

where the factors f_k are uncorrelated with the idiosyncratic disturbances v_i, which in turn are uncorrelated with each other. We assume that $V(v_i) \leq \zeta < \infty$ for all i. Let $\boldsymbol{B}_N$ be the $N \times K$ matrix whose i, j element is β_{ij}. Only the column space of $\boldsymbol{B}_N$ is well-defined, since we can form new factors by taking linear combinations of the f_k. A convenient normalization specifies that the factors are uncorrelated with each other, with zero mean and unit variance. Then $\boldsymbol{\Sigma}_N$ may be decomposed as follows:

$$\boldsymbol{\Sigma}_N = \boldsymbol{B}_N \boldsymbol{B}_N' + \boldsymbol{D}_N \qquad (N = 1, 2, \ldots), \tag{4.2}$$

where $\boldsymbol{D}_N$ is a diagonal matrix whose elements are uniformly bounded by ζ for all N. Of course rank $(B_N B_N') \leq K$.

The following theorem is due to Ross [**12**].[11]

THEOREM 2: *Suppose that* (A) *holds and that there is a strict factor structure, as in* (4.2). *Then there exist numbers* $\tau_0, \tau_1, \ldots, \tau_K, \gamma$ *such that*

$$\sum_{i=1}^{\infty} (\mu_i - \tau_0 - \tau_1 \beta_{i1} - \cdots - \tau_K \beta_{iK})^2 \leq \gamma < \infty.$$

If there is a riskless asset with a return of ρ, *then we may set* $\tau_0 = \rho$.

[11] See Huberman [**5**] for an alternative proof.

This content downloaded from
194.34.233.158 on Wed, 17 Jul 2024 19:28:52 UTC
All use subject to https://about.jstor.org/terms

Theorem 2 is a special case of Theorem 3, whose proof is given below. The theorem states that if there is a strict factor structure, then the absence of arbitrage opportunities implies that the vector of mean returns is approximately a linear function of the factor loadings. Suppose that there is a riskless asset. Since μ_i is the mean return available for one dollar, we can interpret $(\tau_1 \beta_{i1} + \cdots + \tau_K \beta_{iK})$ as the risk premium on asset i. Thus if there is a strict factor structure, Ross' theorem implies that an asset's risk premium is determined by its factor loadings in a particularly simple way. If there is but a single factor, then

$$\mu_i \approx \rho + \tau \beta_i ,$$

which is almost the capital asset pricing formula, with factor loadings playing the role of beta.

The assumption that a strict factor structure holds with a small number of factors seems overly strong. Suppose, for example, that $x_i = \mu_i + \beta_i f + w_i$, $\mathrm{Cov}(f, w_i) = 0$, where the w_i are "almost" uncorrelated: $\mathrm{Cov}(w_i, w_j) = 0$ if $|i-j| > 1$. Then we must let the number of factors grow without limit in order to maintain a strict factor structure as $N \to \infty$. We shall present a weaker condition that is still sufficient for Ross' theorem to hold.

4.2. *Approximate Factor Structure*

The eigenvalues of the diagonal matrix $\boldsymbol{D}_N$ in (4.2) are simply the diagonal elements. Since $V(v_i) \leq \zeta$, the eigenvalues of $\boldsymbol{D}_N$ are uniformly bounded as $N \to \infty$. We shall use this condition to define an approximate K-factor structure. Given a symmetric matrix $\boldsymbol{C}$, let $g_j(\boldsymbol{C})$ denote its jth largest eigenvalue.

DEFINITION 2: The sequence $\{\boldsymbol{\Sigma}_N\}$ has an *approximate K-factor structure* if there exists a sequence $\{\beta_{i1}, \ldots, \beta_{iK}\}_{i=1}^{\infty}$ such that

$$\boldsymbol{\Sigma}_N = \boldsymbol{B}_N \boldsymbol{B}_N' + \boldsymbol{R}_N \qquad (N = 1, 2, \ldots), \tag{4.3}$$

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 with

$$\bar{\lambda} \equiv \sup_N g_1(\boldsymbol{R}_N) < \infty .$$

In the example given above, suppose that $x_i = \mu_i + \beta_i f + w_i$, where $V(w_i) = \varphi_1$, $\mathrm{Cov}(w_i, w_j) = \varphi_2$ if $|i-j| = 1$, and $\mathrm{Cov}(w_i, w_j) = 0$ if $|i-j| > 1$. Then the covariance matrix of $(w_1, \ldots, w_N)$ has uniformly bounded eigenvalues as $N \to \infty$ [**1**, Theorem 6.5.3], and so $\{\boldsymbol{\Sigma}_N\}$ has an approximate 1-factor structure.

The following theorem shows that an approximate factor structure is sufficient for Ross' result.

This content downloaded from
194.34.233.158 on Wed, 17 Jul 2024 19:28:52 UTC
All use subject to https://about.jstor.org/terms

THEOREM 3: *Suppose that* (A) *holds and that* $\{\boldsymbol{\Sigma}_N\}$ *has an approximate K-factor structure, as in* (4.3). *Then there exist numbers* $\tau_0, \tau_1, \ldots, \tau_K$ *such that*

$$\sum_{i=1}^{\infty} (\mu_i - \tau_0 - \tau_1 \beta_{i1} - \cdots - \tau_K \beta_{iK})^2 \leq \bar{\lambda}\delta^2. \tag{4.4}$$

If there is a riskless asset with a return of ρ, *then we may set* $\tau_0 = \rho$.

PROOF: We shall say $\boldsymbol{A} \leq \boldsymbol{B}$ if $\boldsymbol{B} - \boldsymbol{A}$ is a positive semi-definite matrix. Let $\boldsymbol{I}$ be an identity matrix.

$$\boldsymbol{\Sigma}_N = \boldsymbol{B}_N \boldsymbol{B}_N' + \boldsymbol{R}_N \leq \bar{\lambda}(\boldsymbol{C}\boldsymbol{C}' + \boldsymbol{I}),$$

where $\boldsymbol{C} = \bar{\lambda}^{-1/2}\boldsymbol{B}_N$. We can assume without loss of generality that $\boldsymbol{B}_N$ has full column rank for N sufficiently large; otherwise we can throw away some columns.

$$\boldsymbol{\Sigma}_N^{-1} \geq \bar{\lambda}^{-1}(\boldsymbol{C}\boldsymbol{C}' + \boldsymbol{I})^{-1} = \bar{\lambda}^{-1}\big(\boldsymbol{I} - \boldsymbol{C}(\boldsymbol{C}'\boldsymbol{C} + \boldsymbol{I})^{-1}\boldsymbol{C}'\big)$$
$$\geq \bar{\lambda}^{-1}\big(\boldsymbol{I} - \boldsymbol{C}(\boldsymbol{C}'\boldsymbol{C})^{-1}\boldsymbol{C}'\big) = \bar{\lambda}^{-1}\big(\boldsymbol{I} - \boldsymbol{B}_N(\boldsymbol{B}_N'\boldsymbol{B}_N)^{-1}\boldsymbol{B}_N'\big).$$

By Proposition 2,

$$\begin{aligned} \delta^2 &\geq \| \boldsymbol{\mu}_N - [\boldsymbol{l}_N] \|^2_{\boldsymbol{\Sigma}_N^{-1}} \\ &\geq \bar{\lambda}^{-1} \min_{\alpha_0} (\boldsymbol{\mu}_N - \alpha_0 \boldsymbol{l}_N)'\big(\boldsymbol{I} - \boldsymbol{B}_N(\boldsymbol{B}_N'\boldsymbol{B}_N)^{-1}\boldsymbol{B}_N'\big)(\boldsymbol{\mu}_N - \alpha_0 \boldsymbol{l}_N) \\ &= \bar{\lambda}^{-1} \min_{\alpha_0} \min_{\boldsymbol{\alpha}} \| \boldsymbol{\mu}_N - \alpha_0 \boldsymbol{l}_N - \boldsymbol{B}_N \boldsymbol{\alpha} \|^2_{\boldsymbol{I}} \end{aligned} \tag{4.5}$$

(by Lemma 1). Let $\alpha_0 = \tau_{0N}$, $\boldsymbol{\alpha} = \boldsymbol{\tau}_N$ solve this minimization problem. Let $\boldsymbol{G}_N$ be the matrix $(\boldsymbol{l}_N, \boldsymbol{B}_N)$ and let γ_N' be the row vector $(\tau_{0N}, \boldsymbol{\tau}_N')$. We can assume without loss of generality that $\boldsymbol{G}_J$ has full column rank for some J; for if $\boldsymbol{l}_N$ is in the column space of $\boldsymbol{B}_N$ for all N, then we can drop $\boldsymbol{l}_N$, setting $\tau_{0N} = 0$. For $N \geq J$ we have

$$\begin{aligned} \bar{\lambda}^{1/2}\delta &\geq \| \boldsymbol{\mu}_N - \boldsymbol{G}_N \gamma_N \|_{\boldsymbol{I}} \geq \| \boldsymbol{\mu}_J - \boldsymbol{G}_J \gamma_N \|_{\boldsymbol{I}} \\ &\geq \| \boldsymbol{G}_J \gamma_N \|_{\boldsymbol{I}} - \| \boldsymbol{\mu}_J \|_{\boldsymbol{I}} \\ &\geq \varphi^{1/2} \| \gamma_N \|_{\boldsymbol{I}} - \| \boldsymbol{\mu}_J \|_{\boldsymbol{I}}, \end{aligned}$$

where φ is the smallest eigenvalue of $\boldsymbol{G}_J'\boldsymbol{G}_J$; $\varphi > 0$ since $\boldsymbol{G}_J'\boldsymbol{G}_J$ is positive-definite. Hence $\{\gamma_N\}$ is a uniformly bounded sequence and has a convergent subsequence: $\gamma_{N(j)} \to \gamma$ as $j \to \infty$. For any $k \leq N(j)$ we have

$$\| \boldsymbol{\mu}_k - \boldsymbol{G}_k \gamma_{N(j)} \|^2_{\boldsymbol{I}} \leq \bar{\lambda}\delta^2,$$

This content downloaded from
194.34.233.158 on Wed, 17 Jul 2024 19:28:52 UTC
All use subject to https://about.jstor.org/terms

and taking the limit as $j \to \infty$ gives

$$\| \boldsymbol{\mu}_k - \boldsymbol{G}_k \gamma \|_{\boldsymbol{I}}^2 \leq \bar{\lambda} \delta^2.$$

Since this holds for all k, (4.4) follows with $(\tau_0, \boldsymbol{\tau}') = \gamma'$.

If there is a riskless asset, then we can use Proposition 2 to replace (4.5) by

$$\delta^2 \geq \| \boldsymbol{\mu}_N - \rho \boldsymbol{l}_N \|_{\boldsymbol{\Sigma}_N^{-1}}^2;$$

then essentially the same argument gives (4.4) with $\tau_0 = \rho$. *Q.E.D.*

5. A CHARACTERIZATION OF APPROXIMATE FACTOR STRUCTURES

We would like to have a simple condition on the $\{\boldsymbol{\Sigma}_N\}$ sequence that implies an approximate K-factor structure, and we would like to know how to construct the factor loadings (risk premia) from $\{\boldsymbol{\Sigma}_N\}$. If an approximate factor structure does exist, we would like to know whether the decomposition of $\{\boldsymbol{\Sigma}_N\}$ into $\{\boldsymbol{B}_N \boldsymbol{B}_N'\}$ and $\{\boldsymbol{R}_N\}$ is unique. We shall show that the relevant condition is that only K of the eigenvalues of $\{\boldsymbol{\Sigma}_N\}$ are unbounded as $N \to \infty$. Furthermore, there is a unique sequence $\{\boldsymbol{B}_N \boldsymbol{B}_N'\}$ that gives the approximate factor structure, and it can be obtained from the eigenvectors of $\{\boldsymbol{\Sigma}_N\}$ corresponding to the K largest eigenvalues.

We show first that if there is an approximate K-factor structure, then only K of the eigenvalues can be unbounded.

PROPOSITION 4: *Suppose that $\{\boldsymbol{\Sigma}_N\}$ has an approximate K-factor structure as in* (4.3). *Define*

$$\lambda_{K+1} \equiv \sup_N g_{K+1}(\boldsymbol{\Sigma}_N). \tag{5.1}$$

Then λ_{K+1} is finite.

PROOF: It follows from [**8**, Exercise 1.f.1.9] and rank $(\boldsymbol{B}_N \boldsymbol{B}_N') \leq K$ that

$$\begin{aligned} g_{K+1}(\boldsymbol{\Sigma}_N) &\leq g_{K+1}(\boldsymbol{B}_N \boldsymbol{B}_N') + g_1(\boldsymbol{R}_N) \\ &= g_1(\boldsymbol{R}_N); \end{aligned}$$

so $\lambda_{K+1} \leq \bar{\lambda} < \infty$. *Q.E.D.*

Now suppose that $\lambda_{K+1} < \infty$. 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 $\boldsymbol{t}_{jN}' \boldsymbol{t}_{jN} = 1$, $\boldsymbol{t}_{jN}' \boldsymbol{t}_{kN} = 0$ $(j, k = 1, \ldots, N;\ j \neq k)$. Let t_{ijN} be the ith element of $\boldsymbol{t}_{jN}$. Order the eigenvalues so that $\lambda_{1N} \geq \lambda_{2N} \geq \cdots$

This content downloaded from
194.34.233.158 on Wed, 17 Jul 2024 19:28:52 UTC
All use subject to https://about.jstor.org/terms

$\geq \lambda_{NN}$. Then we can decompose

$$\boldsymbol{\Sigma}_N = \boldsymbol{B}_N^* \boldsymbol{B}_N^{*\prime} + \boldsymbol{R}_N^*,$$

where the jth column of the $N \times K$ matrix $\boldsymbol{B}_N^*$ is $\lambda_{jN}^{1/2} \boldsymbol{t}_{jN}$ and

$$\boldsymbol{R}_N^* = \sum_{j=K+1}^{N} \lambda_{jN} \boldsymbol{t}_{jN} \boldsymbol{t}_{jN}^{\prime}.$$

This decomposition cannot correspond to our definition of an approximate K-factor structure since $\{\boldsymbol{B}_N^*\}$ is not a nested sequence—instead of simply adding rows as N increases, the entire matrix changes, so that the i, j element of $\boldsymbol{B}_N^*$ depends upon N. But we do have the following version of Theorem 3.

THEOREM 3′: *Suppose that* (A) *holds and that* $\lambda_{K+1} < \infty$. *Let* $\beta_{ijN} = \lambda_{jN}^{1/2} t_{ijN}$. *Then there exists a sequence* $\{\tau_{0N}, \tau_{1N}, \ldots, \tau_{KN}\}_{N=1}^{\infty}$ *such that*

$$\text{(5.2)} \qquad \sum_{i=1}^{N} (\mu_i - \tau_{0N} - \tau_{1N}\beta_{i1N} - \cdots - \tau_{KN}\beta_{iKN})^2 \leq \lambda_{K+1}\delta^2$$

for $N = 1, 2, \ldots$. *If there is a riskless asset with a return of* ρ, *then we may set* $\tau_{0N} = \rho$.

The proof follows that of Theorem 3 to show $\| \boldsymbol{\mu}_N - [\boldsymbol{l}_N, \boldsymbol{B}_N^*] \|_{\boldsymbol{I}}^2 \leq \lambda_{K+1}\delta^2$; the nesting property of $\{\boldsymbol{B}_N\}$ is not used in that argument. In the riskless asset case, the nesting property is not used in showing that $\| \boldsymbol{\mu}_N - \rho \boldsymbol{l}_N - [\boldsymbol{B}_N] \|_{\boldsymbol{I}}^2 \leq \lambda_{K+1}\delta^2$.

A weakness in (5.2) is that we do not know whether the column space of the β_{ijN} converges as $N \to \infty$; hence the risk premium that we assign to asset i may keep changing as we include more assets in $\boldsymbol{\Sigma}_N$. Theorem 4 will establish that in fact we do have convergence.

THEOREM 4: *Suppose that* $\sup_N \lambda_{KN} = \infty$, $\lambda_{K+1} < \infty$, *and* $\lambda_\infty \equiv \inf_N \lambda_{NN} > 0$. *Define* $\boldsymbol{B}_{NM}$ *to be the* $N \times K$ *matrix whose jth column contains the first* N *elements of* $\lambda_{jM}^{1/2} \boldsymbol{t}_{jM}$ $(M \geq N)$. *Then*

(i) $\{\boldsymbol{\Sigma}_N\}$ *has an approximate K-factor structure*; *i.e., there exists a sequence* $\{\beta_{i1}, \ldots, \beta_{iK}\}_{i=1}^{\infty}$ *such that*

$$\boldsymbol{\Sigma}_N = \boldsymbol{B}_N \boldsymbol{B}_N^{\prime} + \boldsymbol{R}_N \qquad (N = 1, 2, \ldots),$$

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 by* λ_{K+1} *for all* N.

(ii) *For any* N,

$$\lim_{M \to \infty} \boldsymbol{B}_{NM} \boldsymbol{B}_{NM}^{\prime} = \boldsymbol{B}_N \boldsymbol{B}_N^{\prime}.$$

This content downloaded from 194.34.233.158 on Wed, 17 Jul 2024 19:28:52 UTC
All use subject to https://about.jstor.org/terms

(iii) *The approximate K-factor structure is unique*; *i.e., suppose that there is a sequence* $\{\gamma_{i1}, \ldots, \gamma_{iK}\}_{i=1}^{\infty}$ *such that*

$$\boldsymbol{\Sigma}_N = \boldsymbol{G}_N \boldsymbol{G}_N' + \boldsymbol{W}_N \qquad (N = 1, 2, \ldots),$$

where the i, j element of the $N \times K$ matrix $\boldsymbol{G}_N$ is γ_{ij}, and $\{\boldsymbol{W}_N\}$ is a sequence of positive semi-definite matrices whose eigenvalues are uniformly bounded for all N; *then*

$$\boldsymbol{G}_N \boldsymbol{G}_N' = \boldsymbol{B}_N \boldsymbol{B}_N', \qquad \boldsymbol{W}_N = \boldsymbol{R}_N.$$

COROLLARY 2: *Suppose that* (A) *holds together with the assumptions of Theorem* 4. *Then there exist numbers $\tau_0, \tau_1, \ldots, \tau_K$ such that*

$$\sum_{i=1}^{\infty} (\mu_i - \tau_0 - \tau_1 \beta_{i1} - \cdots - \tau_K \beta_{iK})^2 \leq \lambda_{K+1} \delta^2.$$

If there is a riskless asset with a return of ρ, then we may set $\tau_0 = \rho$.

The Corollary is an immediate implication of Theorems 3 and 4. The proof of the Theorem requires a covariance inner product. We have seen (in Section 3.3) that $\overline{\mathscr{F}}$ is a Hilbert space under the covariance inner product when there is no riskless limit portfolio. If there is a riskless limit portfolio p^*, then $V^{1/2}(\ )$ is not a valid norm on $\overline{\mathscr{F}}$ since $V(p^*) = 0$. So we let $z_i = x_i - \mu_i(p^*/E(p^*))$ and define $\mathscr{P}_N$ to be the span of $\{z_1, \ldots, z_N\}$. $\overline{\mathscr{P}}$ is defined as the mean-square closure in $\overline{\mathscr{F}}$ of $\bigcup_{N=1}^{\infty} \mathscr{P}_N$. Then $\overline{\mathscr{P}}$ is a Hilbert space under the mean-square inner product, which is actually a covariance inner product on $\overline{\mathscr{P}}$ since $E(z_i) = 0$. If there is no riskless limit portfolio, we simply set $z_i = x_i$, so that $\overline{\mathscr{P}} = \overline{\mathscr{F}}$. $\overline{\mathscr{P}}$ is still a Hilbert space under the covariance inner product. All references to orthogonality, norms, and convergence in $\overline{\mathscr{P}}$ will be with respect to the covariance inner product and the variance norm.

Before proving the Theorem, we shall need some definitions and a lemma. Let $z_N' = (z_1, \ldots, z_N)$ and set $r_{jN} = (t_{jN}' z_N)/\lambda_{jN}^{1/2}$, so that $V(r_{jN}) = 1$, $\mathrm{Cov}(r_{jN}, r_{kN}) = 0$ $(j, k = 1, \ldots, K; j \neq k)$. Then the orthogonal projection of $p \in \overline{\mathscr{P}}$ onto the subspace spanned by $\{r_{1N}, \ldots, r_{KN}\}$ is given by

$$Q_N p = \sum_{k=1}^{K} \mathrm{Cov}(p, r_{kN}) r_{kN}. \tag{5.3}$$

The following result is proved in [**3**, Lemmas 1, 4, and 5].

LEMMA 2: *Suppose that* $\sup_N \lambda_{KN} = \infty$, $\lambda_{K+1} < \infty$, *and* $\lambda_\infty \equiv \inf_N \lambda_{NN} > 0$. *Then the following results hold*:

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

$$\|\alpha p\|_2 = |\alpha| \, \|p\|_2; \qquad \|p + q\|_2 \leq \|p\|_2 + \|q\|_2;$$

This content downloaded from
194.34.233.158 on Wed, 17 Jul 2024 19:28:52 UTC
All use subject to https://about.jstor.org/terms

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

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

(ii) $\mathscr{P}_1 \equiv \{p\in\overline{\mathscr{P}}:\ \|p\|_2 = 0\}$ *is a* K*-dimensional linear subspace of* $\overline{\mathscr{P}}$.
(iii) *If* $p_1\in\mathscr{P}_1$ *and* $p\in\overline{\mathscr{P}}$, *then* $\|p_1+p\|_2 = \|p\|_2$.
(iv) *If* $p\in\overline{\mathscr{P}}$ *and* $\mathrm{Cov}(p,q)=0$ *for all* $q\in\mathscr{P}_1$, *then* $V(p)\le\lambda_{K+1}\|p\|_2^2$.
(v) *Let* Qp *be the orthogonal projection of* $p\in\overline{\mathscr{P}}$ *onto* $\mathscr{P}_1$; *then*

$$\lim_{N\to\infty} Q_Np = Qp.$$

Note that a special case of (i) is $\|\sum_{i=1}^{N}\alpha_i z_i\|_2 = (\sum_{i=1}^{N}\alpha_i^2)^{1/2}$. The seminorm $\|\ \|_2$ is a measure of diversification, and $\mathscr{P}_1$ can be interpreted as the space of well-diversified portfolios. This is developed in [**3**].

Proof of Theorem 4: Let $f_1,\dots,f_K$ be an orthonormal basis for $\mathscr{P}_1$ and set $\beta_{ij}=\mathrm{Cov}(z_i,f_j)$. Then

$$z_i = \beta_{i1}f_1 + \cdots + \beta_{iK}f_K + e_i \qquad (i=1,2,\dots) \tag{5.4}$$

gives the orthogonal projection of z_i onto $\mathscr{P}_1$; $\mathrm{Cov}(f_j,e_i)=0$ $(j=1,\dots,K)$.

(i) Let the i,j element of $\boldsymbol{B}_N$ be β_{ij} and let the i,j element of $\boldsymbol{R}_N$ be $\mathrm{Cov}(e_i,e_j)$. Since $\mathrm{Cov}(z_i,z_j)=\sigma_{ij}$, (5.4) implies that

$$\boldsymbol{\Sigma}_N = \boldsymbol{B}_N\boldsymbol{B}_N' + \boldsymbol{R}_N \qquad (N=1,2,\dots).$$

Let $\boldsymbol{e}_N' = (e_1,\dots,e_N)$. Since $e_i\in\mathscr{P}_1^{\perp}$, Lemma 2 implies that

$$\boldsymbol{\alpha}'\boldsymbol{R}_N\boldsymbol{\alpha} = V(\boldsymbol{\alpha}'\boldsymbol{e}_N)\le\lambda_{K+1}\|\boldsymbol{\alpha}'\boldsymbol{e}_N\|_2^2 = \lambda_{K+1}\|\boldsymbol{\alpha}'\boldsymbol{z}_N\|_2^2 = \lambda_{K+1}\boldsymbol{\alpha}'\boldsymbol{\alpha}.$$

Hence the eigenvalues of $\boldsymbol{R}_N$ are uniformly bounded by λ_{K+1}.

(ii)
$$\mathrm{Cov}(z_i,Q_Mz_j) = \sum_{k=1}^{K}\mathrm{Cov}(z_j,r_{kM})\mathrm{Cov}(z_i,r_{kM}) = \sum_{k=1}^{K}\lambda_{kM}t_{ikM}t_{jkM};$$

$$\mathrm{Cov}(z_i,Qz_j) = \mathrm{Cov}\left(\sum_{k=1}^{K}\beta_{ik}f_k + e_i,\ \sum_{k=1}^{K}\beta_{jk}f_k\right) = \sum_{k=1}^{K}\beta_{ik}\beta_{jk}.$$

Lemma 2 asserts that $Q_Mz_j\to Qz_j$ as $M\to\infty$. Hence

$$\lim_{M\to\infty}\sum_{k=1}^{K}\beta_{ikM}\beta_{jkM} = \sum_{k=1}^{K}\beta_{ik}\beta_{jk},$$

where $\beta_{ikM}=\lambda_{kM}^{1/2}t_{ikM}$ is the i,k element of $\boldsymbol{B}_{NM}$ $(i,j=1,\dots,N;\ k=1,\dots,K)$.

(iii) Let $\boldsymbol{S}_M$ be the $M\times K$ matrix whose jth column is $\lambda_{jM}^{-1/2}\boldsymbol{t}_{jM}$ $(M=1,2,\dots)$. Let $\boldsymbol{D}_M$ be the $K\times K$ diagonal matrix with diagonal elements

This content downloaded from
194.34.233.158 on Wed, 17 Jul 2024 19:28:52 UTC
All use subject to https://about.jstor.org/terms

$\lambda_{1M}, \ldots, \lambda_{KM}$. Let $\bar{\lambda}$ be a uniform upper bound on the eigenvalues of $\boldsymbol{W}_M$ for all M. Recall that $\boldsymbol{A} \leq \boldsymbol{B}$ if $\boldsymbol{B} - \boldsymbol{A}$ is a positive semi-definite matrix.

$$\boldsymbol{S}_M' \boldsymbol{W}_M \boldsymbol{S}_M \leq \bar{\lambda} \boldsymbol{S}_M' \boldsymbol{S}_M = \bar{\lambda} \boldsymbol{D}_M^{-1} \to \boldsymbol{0},$$

and

$$\boldsymbol{S}_M' \boldsymbol{W}_M \boldsymbol{W}_M \boldsymbol{S}_M \leq \bar{\lambda}^2 \boldsymbol{S}_M' \boldsymbol{S}_M = \bar{\lambda}^2 \boldsymbol{D}_M^{-1} \to \boldsymbol{0}$$

as $M \to \infty$, so that $\boldsymbol{S}_M' \boldsymbol{W}_M \boldsymbol{S}_M \to \boldsymbol{0}$ and $\boldsymbol{S}_M' \boldsymbol{W}_M \boldsymbol{W}_M \boldsymbol{S}_M \to \boldsymbol{0}$, in the sense that each element of these matrices converges to zero. Let $\boldsymbol{C}_M$ be the $K \times K$ matrix $\boldsymbol{G}_M' \boldsymbol{S}_M$ and let $\boldsymbol{I}_K$ be the $K \times K$ identity matrix. Then

$$\boldsymbol{I}_K = \boldsymbol{S}_M' \boldsymbol{\Sigma}_M \boldsymbol{S}_M = \boldsymbol{S}_M' \boldsymbol{G}_M \boldsymbol{G}_M' \boldsymbol{S}_M + \boldsymbol{S}_M' \boldsymbol{W}_M \boldsymbol{S}_M$$

implies that $\boldsymbol{C}_M' \boldsymbol{C}_M \to \boldsymbol{I}_K$ as $M \to \infty$. Hence the elements of $\boldsymbol{C}_M$ are uniformly bounded for all M, and $\{\boldsymbol{C}_M\}$ has a convergent subsequence: $\boldsymbol{C}_{M(j)} \to \boldsymbol{C}$ as $j \to \infty$. $\boldsymbol{C}'\boldsymbol{C} = \boldsymbol{I}_K$ implies that $\boldsymbol{C}' = \boldsymbol{C}^{-1}$ and so $\boldsymbol{C}\boldsymbol{C}' = \boldsymbol{I}_K$.

Recall that $\boldsymbol{B}_{NM}$ is the $N \times K$ matrix whose jth column contains the first N elements of $\lambda_{jM}^{1/2} \boldsymbol{t}_{jM}$ $(M \geq N)$. Then

$$\begin{aligned} \boldsymbol{B}_{MM} &= \boldsymbol{\Sigma}_M \boldsymbol{S}_M = (\boldsymbol{G}_M \boldsymbol{G}_M' + \boldsymbol{W}_M) \boldsymbol{S}_M \\ &= \boldsymbol{G}_M \boldsymbol{C}_M + \boldsymbol{W}_M \boldsymbol{S}_M \end{aligned}$$

implies that

$$\boldsymbol{B}_{NM} = \boldsymbol{G}_N \boldsymbol{C}_M + \boldsymbol{H}_{NM},$$

where $\boldsymbol{H}_{NM}$ is the $N \times K$ matrix that contains the first N rows of $\boldsymbol{W}_M \boldsymbol{S}_M$.

$$\boldsymbol{H}_{NM}' \boldsymbol{H}_{NM} \leq \boldsymbol{S}_M' \boldsymbol{W}_M \boldsymbol{W}_M \boldsymbol{S}_M \to \boldsymbol{0}$$

implies that $\boldsymbol{H}_{NM} \to \boldsymbol{0}$ as $M \to \infty$ and so $\boldsymbol{H}_{NM} \boldsymbol{H}_{NM}' \to \boldsymbol{0}$. Since $\{\boldsymbol{C}_M\}$ is uniformly bounded, $\boldsymbol{G}_N \boldsymbol{C}_M \boldsymbol{H}_{NM}' \to \boldsymbol{0}$ as $M \to \infty$. Hence part (ii) of the Theorem gives

$$\begin{aligned} \boldsymbol{B}_N \boldsymbol{B}_N' &= \lim_{M \to \infty} \boldsymbol{B}_{NM} \boldsymbol{B}_{NM}' \\ &= \boldsymbol{G}_N \Big(\lim_{j \to \infty} \boldsymbol{C}_{M(j)} \boldsymbol{C}_{M(j)}' \Big) \boldsymbol{G}_N' = \boldsymbol{G}_N \boldsymbol{G}_N'. \end{aligned}$$

It follows that

$$\boldsymbol{W}_N = \boldsymbol{\Sigma}_N - \boldsymbol{G}_N \boldsymbol{G}_N' = \boldsymbol{\Sigma}_N - \boldsymbol{B}_N \boldsymbol{B}_N' = \boldsymbol{R}_N. \qquad Q.E.D.$$

We would like to relate our results to conventional factor analysis. Recall that a strict one-factor model specifies

$$\boldsymbol{\Sigma}_N = \boldsymbol{b}_N \boldsymbol{b}_N' + \boldsymbol{D}_N \qquad (N = 1, 2, \ldots), \tag{5.5}$$

where $\boldsymbol{b}_N' = (\beta_1, \ldots, \beta_N)$ and $\boldsymbol{D}_N$ is a diagonal matrix with $V(v_i)$ as the ith

This content downloaded from
194.34.233.158 on Wed, 17 Jul 2024 19:28:52 UTC
All use subject to https://about.jstor.org/terms

diagonal element; $V(v_i) \leq \zeta < \infty$ for all i. Given some weak restrictions on $\{\boldsymbol{b}_N\}$ and $\{\boldsymbol{D}_N\}$, we shall show that $\{\boldsymbol{\Sigma}_N\}$ satisfies the assumptions of Theorem 4. Proposition 4 implies that $\sup_N \lambda_{2N} < \infty$. If $\varphi \equiv \inf_i V(v_i) > 0$, then $\boldsymbol{\alpha}'\boldsymbol{\Sigma}_N\boldsymbol{\alpha} \geq \varphi\boldsymbol{\alpha}'\boldsymbol{\alpha}$ implies that $\inf_N \lambda_{NN} > 0$. Since $\boldsymbol{b}_N'\boldsymbol{\Sigma}_N\boldsymbol{b}_N \geq (\boldsymbol{b}_N'\boldsymbol{b}_N)^2$, we have $\lambda_{1N} \to \infty$ if $\sum_{i=1}^{\infty} \beta_i^2 = \infty$. So given these restrictions, Theorem 4 implies that there is an approximate one-factor structure; since it is unique, it must coincide with the strict factor structure in (5.5). If $\beta_1 \neq 0$, the convergence part of Theorem 4 gives

$$\lim_{M \to \infty} t_{i1M}/t_{11M} = \beta_i/\beta_1 \qquad (i = 1, 2, \ldots),$$

where $\boldsymbol{t}_{1M}' = (t_{11M}, \ldots, t_{M1M})$ is the eigenvector of $\boldsymbol{\Sigma}_M$ corresponding to the largest eigenvalue.

Hence we can obtain the factor loadings of the strict factor structure from the first eigenvector of $\{\boldsymbol{\Sigma}_N\}$. In conventional factor analysis, the factor loadings are obtained from a different eigenvalue problem. It follows from (5.5) that

$$\boldsymbol{\Sigma}_N \boldsymbol{s} = \theta \boldsymbol{D}_N \boldsymbol{s},$$

where $\boldsymbol{s} = \boldsymbol{D}_N^{-1}\boldsymbol{b}_N$ and $\theta = 1 + \boldsymbol{b}_N'\boldsymbol{D}_N^{-1}\boldsymbol{b}_N$. So $\boldsymbol{s}$ is an eigenvector of $\boldsymbol{\Sigma}_N$ relative to $\boldsymbol{D}_N$ (or an ordinary eigenvector of $\boldsymbol{D}_N^{-1}\boldsymbol{\Sigma}_N$). In empirical factor analysis, there is a sample counterpart to this population result. Given a sample covariance matrix $\hat{\boldsymbol{\Sigma}}_N$ from a strict K-factor structure, the maximum likelihood estimator of $\boldsymbol{B}_N$ (under normality assumptions) can be obtained from the first K eigenvectors of $\hat{\boldsymbol{\Sigma}}_N$ relative to $\hat{\boldsymbol{D}}_N$, where $\hat{\boldsymbol{D}}_N$ is the maximum likelihood estimator of $\boldsymbol{D}_N$ [7, p. 27]. Much of the work in maximum likelihood factor analysis is in the computation of $\boldsymbol{D}_N$. Our results provide a rigorous justification for principal component analysis, which is computationally simpler than factor analysis since $\boldsymbol{D}_N$ is set equal to an identity matrix. Furthermore, the arbitrage pricing interpretation of the principal components holds under much weaker assumptions than a strict factor structure.

A common objection to principal component analysis is that it is arbitrary to take the eigenvectors of $\boldsymbol{\Sigma}_N$ relative to an identity matrix, instead of using some other positive-definite matrix $\boldsymbol{\Omega}_N$. In the case of a strict factor model, for example, it seems more natural to set $\boldsymbol{\Omega}_N = \boldsymbol{D}_N$, which gives conventional factor analysis. We have just argued, however, that factor analysis and principal component analysis are asymptotically equivalent, if there is a strict factor structure. We shall show in Corollary 3 that there is a much stronger result, which only requires an approximate factor structure. Under weak restrictions on $\{\boldsymbol{\Omega}_N\}$, taking the eigenvectors of $\{\boldsymbol{\Sigma}_N\}$ relative to $\{\boldsymbol{\Omega}_N\}$ gives the same asymptotic factor loadings as principal component analysis.

Let $\{\boldsymbol{\Omega}_N\}$ be a nested sequence of positive-definite matrices, with eigenvalues uniformly bounded away from 0 and ∞: $\varphi_\infty \equiv \inf_N g_N(\boldsymbol{\Omega}_N) > 0$, $\varphi_1 \equiv \sup_N g_1(\boldsymbol{\Omega}_N) < \infty$; the i, j element of $\boldsymbol{\Omega}_N$ is ω_{ij}. There exists a nonsingular matrix $\boldsymbol{S}_N$ such that

$$\boldsymbol{S}_N'\boldsymbol{\Sigma}_N\boldsymbol{S}_N = \boldsymbol{\Theta}_N, \qquad \boldsymbol{S}_N'\boldsymbol{\Omega}_N\boldsymbol{S}_N = \boldsymbol{I}_N,$$

This content downloaded from
194.34.233.158 on Wed, 17 Jul 2024 19:28:52 UTC
All use subject to https://about.jstor.org/terms

where $\boldsymbol{\Theta}_N$ is a diagonal matrix with diagonal elements $\theta_{1N} \geq \cdots \geq \theta_{NN}$. Let $\boldsymbol{s}_{jN}$ be the jth column of $\boldsymbol{S}_N$. Then

$$\boldsymbol{\Sigma}_N \boldsymbol{S}_N = \boldsymbol{S}_N'^{-1} \boldsymbol{\Theta}_N = \boldsymbol{\Omega}_N \boldsymbol{S}_N \boldsymbol{\Theta}_N ,$$

so that $\boldsymbol{s}_{jN}$ is an eigenvector of $\boldsymbol{\Sigma}_N$ relative to $\boldsymbol{\Omega}_N$.

We can use these eigenvectors to obtain an alternative arbitrage pricing formula. Since

$$\boldsymbol{\Sigma}_N = \boldsymbol{\Omega}_N \boldsymbol{S}_N \boldsymbol{\Theta}_N \boldsymbol{S}_N' \boldsymbol{\Omega}_N ,$$

we have

$$\boldsymbol{\Sigma}_N = \boldsymbol{B}_N^* \boldsymbol{B}_N^{*\prime} + \boldsymbol{R}_N^* ,$$

where $\boldsymbol{B}_N^*$ is the $N \times K$ matrix whose jth column is $\boldsymbol{t}_{jN}^* = \theta_{jN}^{1/2} \boldsymbol{\Omega}_N \boldsymbol{s}_{jN}$ and

$$\boldsymbol{R}_N^* = \sum_{j=K+1}^{N} \theta_{jN} \boldsymbol{\Omega}_N \boldsymbol{s}_{jN} \boldsymbol{s}_{jN}' \boldsymbol{\Omega}_N .$$

Assume that $\lambda_{K+1} < \infty$; we shall see below that $\theta_{K+1,N} \leq \varphi_\infty^{-1} \lambda_{K+1}$. Hence

$$\boldsymbol{R}_N^* \leq \varphi_\infty^{-1} \lambda_{K+1} \boldsymbol{\Omega}_N \left(\sum_{j=1}^{N} \boldsymbol{s}_{jN} \boldsymbol{s}_{jN}' \right) \boldsymbol{\Omega}_N$$

$$= \varphi_\infty^{-1} \lambda_{K+1} \boldsymbol{\Omega}_N \leq \varphi_\infty^{-1} \lambda_{K+1} \varphi_1 \boldsymbol{I}_N .$$

Then, just as with Theorem 3′, we can follow the proof of Theorem 3 to obtain

$$\sum_{i=1}^{N} (\mu_i - \tau_{0N} - \tau_{1N} \beta_{i1N}^* - \cdots - \tau_{KN} \beta_{iKN}^*)^2 \leq \varphi_\infty^{-1} \lambda_{K+1} \varphi_1 \delta^2, \tag{5.6}$$

for $N = 1, 2, \ldots ,$ where β_{ijN}^* is the i, j element of $\boldsymbol{B}_N^*$.

The objection that principal component analysis is arbitrary is relevant in our context only if the column space of $\boldsymbol{B}_N^*$ still depends upon $\{\boldsymbol{\Omega}_N\}$ in the limit as $N \to \infty$. The following result shows that it does not.

COROLLARY 3: *Suppose that* $\{\boldsymbol{\Sigma}_N\}$ *satisfies the assumptions of Theorem* 4 *and that* $\varphi_1 \equiv \sup_N g_1(\boldsymbol{\Omega}_N) < \infty$, $\varphi_\infty \equiv \inf_N g_N(\boldsymbol{\Omega}_N) > 0$. *Let* $\{\boldsymbol{B}_N \boldsymbol{B}_N'\}$ *be defined as in Theorem* 4 *and define* $\boldsymbol{B}_{NM}^*$ *to be the* $N \times K$ *matrix whose jth column contains the first N elements of* $\theta_{jM}^{1/2} \boldsymbol{\Omega}_M \mathbf{s}_{jM}$ $(M \geq N)$. *Then*

$$\lim_{M \to \infty} \boldsymbol{B}_{NM}^* \boldsymbol{B}_{NM}^{*\prime} = \boldsymbol{B}_N \boldsymbol{B}_N' \qquad (N = 1, 2, \ldots),$$

which does not depend upon $\{\boldsymbol{\Omega}_N\}$.

PROOF: Recall that a sequence of matrices $\{\boldsymbol{A}_N\}$ is nested if the i, j element of $\boldsymbol{A}_N$ does not depend upon N. We can recursively form a nested sequence $\{\boldsymbol{C}_N\}$

This content downloaded from
194.34.233.158 on Wed, 17 Jul 2024 19:28:52 UTC
All use subject to https://about.jstor.org/terms

of upper-triangular, nonsingular matrices such that $C_N'\Omega_N C_N = I_N$ for all N; the i, j element of C_N is γ_{ij}, with $\gamma_{ij} = 0$ for $i > j$. Let $\bar{\Sigma}_N = C_N'\Sigma_N C_N$ and $\bar{T}_N = C_N^{-1}S_N$. Then $\bar{\Sigma}_N \bar{T}_N = \bar{T}_N \Theta_N$, so that θ_{jN} is an eigenvalue of $\bar{\Sigma}_N$, with $\bar{t}_{jN} = C_N^{-1}s_{jN}$ as the corresponding eigenvector. Note that $\bar{t}_{jN}'\bar{t}_{jN} = 1$, $\bar{t}_{jN}'\bar{t}_{kN} = 0$ $(j,k = 1, \ldots, N; j \neq k)$. We need to show that $\{\bar{\Sigma}_N\}$ satisfies the assumptions of Theorem 4. $\{\bar{\Sigma}_N\}$ is a sequence of positive-definite, nested matrices since the C_N are nonsingular, upper-triangular, and form a nested sequence. Let G be a $N \times (k-1)$ matrix and let α be a $N \times 1$ vector; it follows from [**8**, 1f.2.iii] that

$$\theta_{kN} = \inf_{G} \sup_{G'\alpha = 0} \frac{\alpha'\bar{\Sigma}_N \alpha}{\alpha'\alpha}.$$

The substitution $H = C_N'^{-1}G$, $\beta = C_N \alpha$ gives

$$\theta_{kN} = \inf_{H} \sup_{H'\beta = 0} \left(\frac{\beta'\Sigma_N\beta}{\beta'\beta} \right) \left(\frac{\beta'\beta}{\beta'\Omega_N\beta} \right).$$

Since

$$\lambda_{kN} = \inf_{H} \sup_{H'\beta = 0} \frac{\beta'\Sigma_N\beta}{\beta'\beta}$$

and $\varphi_\infty \leq \beta'\Omega_N\beta/\beta'\beta \leq \varphi_1$, we have

$$\varphi_1^{-1}\lambda_{kN} \leq \theta_{kN} \leqslant \varphi_\infty^{-1}\lambda_{kN} \qquad (k = 1, \ldots, N).$$

Hence $\sup_N \theta_{KN} = \infty$, $\sup_N \theta_{K+1,N} < \infty$, and $\inf_N \theta_{NN} > 0$.

So we can apply Theorem 4 to conclude that $\{\bar{\Sigma}_N\}$ has an approximate K-factor structure:

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

and

$$\lim_{M \to \infty} \bar{B}_{NM} \bar{B}_{NM}' = \bar{B}_N \bar{B}_N',$$

where $\bar{B}_{NM}$ is the $N \times K$ matrix whose jth column contains the first N elements of $\theta_{jM}^{1/2}\bar{t}_{jM}$ $(M \geq N)$. Hence

$$\Sigma_N = \left(C_N'^{-1}\bar{B}_N\right)\left(C_N'^{-1}\bar{B}_N\right)' + C_N'^{-1}\bar{R}_N C_N^{-1}. \tag{5.7}$$

Note that $\{C_N'^{-1}\bar{B}_N\}$ is a sequence of nested matrices, since $\{\bar{B}_N\}$ is nested and $\{C_N'^{-1}\}$ is nested and lower-triangular. Since

$$g_1\left(C_N'^{-1}\bar{R}_N C_N^{-1}\right) \leq \varphi_1 g_1\left(\bar{R}_N\right),$$

the uniform upper bound on $g_1(\bar{R}_N)$ implies that the eigenvalues of $C_N'^{-1}\bar{R}_N C_N^{-1}$ are uniformly bounded for all N, which implies that (5.7) gives an approximate

This content downloaded from
194.34.233.158 on Wed, 17 Jul 2024 19:28:52 UTC
All use subject to https://about.jstor.org/terms

K-factor structure for $\{\boldsymbol{\Sigma}_N\}$. Then the uniqueness result in Theorem 4 shows that

$$\boldsymbol{C}_N'^{-1}\overline{\boldsymbol{B}}_N\overline{\boldsymbol{B}}_N'\boldsymbol{C}_N^{-1} = \boldsymbol{B}_N\boldsymbol{B}_N'.$$

Since $\boldsymbol{C}_M'^{-1}$ is lower-triangular and $\boldsymbol{C}_M'^{-1}\overline{\boldsymbol{B}}_{MM} = \boldsymbol{B}_{MM}^*$, we have $\boldsymbol{B}_{NM}^* = \boldsymbol{C}_N'^{-1}\overline{\boldsymbol{B}}_{NM}$. Hence

$$\boldsymbol{B}_N\boldsymbol{B}_N' = \boldsymbol{C}_N'^{-1}\left(\lim_{M\to\infty}\overline{\boldsymbol{B}}_{NM}\overline{\boldsymbol{B}}_{NM}'\right)\boldsymbol{C}_N^{-1} = \lim_{M\to\infty}\boldsymbol{B}_{NM}^*\boldsymbol{B}_{NM}^{*\prime}. \qquad Q.E.D.$$

University of Wisconsin, Madison
and
National Bureau of Economic Research

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

REFERENCES

[1] ANDERSON, T. W.: *The Statistical Analysis of Time Series.* New York: John Wiley and Sons, 1971.

[2] CHAMBERLAIN, G.: "A Characterization of the Distributions that Imply Mean-Variance Utility Functions," *Journal of Economic Theory*, 29(1983), 185–201.

[3] ———: "Funds, Factors, and Diversification in Arbitrage Pricing Models," *Econometrica*, 51(1983), 1305–1323.

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

[5] HUBERMAN, G.: "A Simple Approach to Arbitrage Pricing Theory," *Journal of Economic Theory*, 28(1982), 183–191.

[6] KREPS, D. M.: "Arbitrage and Equilibrium in Economies with Infinitely Many Commodities," *Journal of Mathematical Economics*, 8(1981), 15–35.

[7] LAWLEY, D. N., AND A. E. MAXWELL: *Factor Analysis as a Statistical Method.* New York: Elsevier, 1971.

[8] RAO, C. R.: *Linear Statistical Inference and Its Applications*, 2nd edition. New York: John Wiley and Sons, 1973.

[9] REED, M., AND B. SIMON: *Methods of Modern Mathematical Physics. I: Functional Analysis.* New York: Academic Press, 1972.

[10] ROLL, R.: "A Critique of the Asset Pricing Theory's Tests. Part I: On Past and Potential Testability of the Theory," *Journal of Financial Economics*, 4(1977), 129–176.

[11] ROLL, R., AND S. A. ROSS: "An Empirical Investigation of the Arbitrage Pricing Theory," *Journal of Finance*, 5(1980), 1073–1103.

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

[13] ———: "The Capital Asset Pricing Model (CAPM), Short-Sale Restrictions, and Related Issues," *Journal of Finance*, 32(1977), 177–183.

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

This content downloaded from
194.34.233.158 on Wed, 17 Jul 2024 19:28:52 UTC
All use subject to https://about.jstor.org/terms