Discussion Paper #1140R

# FACTOR ANALYSIS AND ARBITRAGE PRICING IN LARGE ASSET ECONOMIES*

Nabil I. AL-NAJJAR**

May 1994
Revised: August 1997

(*) Forthcoming in *Journal of Economic Theory*. CMSEMS Discussion Paper no. 1140R.

(**) Department of Managerial Economics and Decision Sciences, J. L. Kellogg Graduate School of Management, Northwestern University, 2001 Sheridan Road, Evanston, IL, 60208.

[Phone: 847-491-5426; Fax: 847-467-1220; e-mail: al-najjar@nwu.edu]

I am especially indebted to Greg Greiff for his comments and many discussions about the ideas presented in this paper. I also thank an Associate Editor, a referee, Torben Andersen, Mike Hemler, and seminar participants at Notre Dame (Finance), Queen's (Finance), Laval and Toronto for their comments. Work on the first version of this paper (May 1994) was partially funded by a grant from the Social Sciences and Humanities Research Council of Canada. Any remaining errors and shortcomings are my own.

**Abstract:**

The paper develops a framework for factor analysis and arbitrage pricing in a large asset economy modeled as one with a continuum of assets. It is shown that the assumptions of absence of arbitrage opportunities and that returns have a strict factor structure imply exact factor-pricing for a full measure of assets. If finite subsets of assets are interpreted as independent random draws from the underlying economy, then there is probability one that, in a finite sample of assets, every asset is exactly factor-priced. I further show that approximate factor structures exist in general and that they can be chosen optimally according to a measure of their explanatory power. Factor structures in the present model are also robust to asset repackaging and to the use of proxies to approximate the true factors.

*Journal of Economic Literature* Classification Numbers: G1, G12, C14.

Suggested Running Head: *Factor Analysis and Arbitrage Pricing in Large Economies*

# 1. INTRODUCTION

Factor models simplify the study of complex correlation patterns in large populations by dividing individual risks into systematic economy-wide components, and individual-specific idiosyncratic components. By identifying common factor-risks and providing a simple relationship relating them to individual risks, factor models proved to be a useful tool with a wide range of applications.

One important application of factor models is the Arbitrage Pricing Theory (APT) proposed by Ross [14].[1] The APT builds on the intuition that a large economy offers investors the opportunity to eliminate idiosyncratic risk through diversification of asset holdings. The absence of arbitrage opportunities then implies that an expected excess return (or a risk premium) will be paid only to compensate for bearing non-diversifiable systematic risk. Assets can therefore be 'factor-priced' in the sense that any excess return can be explained as a linear combination of the factors' risk premia weighted by the asset's exposures to factor risks. This intuition is traditionally formalized using a model of an economy with an infinite number of assets $T = \{1, 2, \ldots\}$. Call the difference between the actual excess return of an asset and the excess return predicted by the APT's factor-pricing formula the asset's 'pricing error'. The main result of the APT states that the sum of squared pricing errors is finite, so, most assets have small pricing errors.

The present paper provides an alternative framework in which the space of assets is indexed by $T = [0, 1]$ instead of the traditional approach employing an infinite sequence. Within this framework, pricing results in the spirit of the APT are derived. One such result is that, outside a set of measure zero, every asset is exactly factor-priced. If we interpret finite subsets of assets as independent random samples drawn from the underlying economy, then, with probability one, all assets in such samples have zero pricing errors. These results differ from the traditional APT's conclusion that 'most' assets have 'small' pricing errors, a conclusion which is consistent with all assets being incorrectly priced.

The pricing results are derived under the standard assumptions of the pure-arbitrage version of the APT, requiring only a strict factor structure for asset returns and the absence of arbitrage opportunities (equivalently, continuity of the pricing function). As with the usual APT, the idea is to determine the restrictions on asset pricing relationships derived from the no-arbitrage assumption alone - without imposing further conditions on market equilibrium, investor preferences, or the distributional properties of returns.

Of course, interpreting finite subsets of assets as independent random draws is not intended as a descriptively accurate account of how asset pricing theories are tested in practice. Rather, it is an idealization of how an outside observer might test the APT's pricing restriction using information contained in a finite sample of assets that highlights the strong pricing restrictions imposed by the APT's assumptions. The pricing results are subject to two other caveats.[2] First, just as in the usual APT, the pricing errors in a particular finite subset of assets can be arbitrarily large. The pricing result asserts something about the magnitude of pricing errors *on average* rather than in any particular sample. Formally, the sampling result is formulated as a probability-one statement on the space of randomly and independently drawn samples. Second, the results assume that the pricing function is fixed independently of the samples drawn.

The paper also provides an analysis of factor structures in large economies, with applications not necessarily limited to the APT. The main concept developed is that of the *explanatory power* of a set of candidate factors, which may be viewed as a continuum analogue of $R^2$ in Statistics. The idea is to view candidate factors as a set of regressors and compute the percentage of total variations in asset returns explained by them. This provides a formal criterion for ranking the explanatory power of alternative sets of factors, a criterion that can help in evaluating the gain from including additional factors and in formulating a trade-off between parsimony and completeness of factor representations.

Using the criterion of explanatory power, a procedure of optimal sequential factor selection is introduced and we study its asymptotic properties. It is shown that optimal approximate factor structures exist and can be selected to reflect the primitives of asset returns. If the economy has a strict factor structure, then the 'true' factor space is unique and can be computed using this sequential procedure. Furthermore, the explanatory power of a factor space changes continuously with that space in the sense small misspecification of the true factors lead to a correspondingly small loss in explanatory power. This is important because the true factor space is not known in practice, so proxies containing estimation errors must be relied on instead.[3]

Why does the choice of an index set (continuum vs. infinite sequence) matter? Intuitively, the reason is that the main conclusions of the APT and factor models involve comparisons of the relative size of various subsets of assets. Examples of such statements are: *most* assets are priced correctly; a *typical* asset can be factor priced; and, a factor is significant if it accounts for a significant part of the variation in *many* assets. These statements can be given a clear meaning in finite assets economies by a measure of the relative weight of subsets of assets to reflect what the modeler intuitively has in mind (*e.g.*, assigning assets equal weights). It therefore seems reasonable to expect abstract infinite models to also have a measure of relative weights reflecting the structure implied by the finite environments they are meant to represent. It is not at all obvious how to develop such a measure for a model with an infinite sequence of assets. To illustrate the difficulties involved, consider the following example:

*Example 1:* Let $\{\tilde{\eta}_m\}$ be an i.i.d. sequence of random variables with zero mean and unit variance and consider the sequence of assets:

$$
\begin{array}{llll}
\tilde{r}_1 = \tilde{\eta}_1, & & & \\
\tilde{r}_2 = \tilde{\eta}_1, & \tilde{r}_3 = \tilde{\eta}_2, & & \\
\tilde{r}_4 = \tilde{\eta}_1, & \tilde{r}_5 = \tilde{\eta}_2, & \tilde{r}_6 = \tilde{\eta}_3, & \\
\tilde{r}_7 = \tilde{\eta}_1, & \tilde{r}_8 = \tilde{\eta}_2, & \tilde{r}_9 = \tilde{\eta}_3, & \tilde{r}_{10} = \tilde{\eta}_4, \\
\ldots & & &
\end{array}
$$

This economy has no approximate $K$-factor structure for any finite $K$ (in the sense of Chamberlain and Rothchild). The problem in this example is that there is no obvious way to assign weights to assets to use in ranking the 'factors' $\tilde{\eta}_1, \tilde{\eta}_2, \ldots$. Concretely, this means that it is difficult to answer questions like: Is $\tilde{\eta}_1$ more significant in explaining asset returns than, say, $\tilde{\eta}_{2^{100}}$? If we take a large sample of assets, in what ratio would we expect $\tilde{\eta}_1$ and $\tilde{\eta}_{2^{100}}$ to be represented? And, in what sense does the infinite sequence economy reflect properties of large finite asset economies $E_n = \{\tilde{r}_1, \ldots, \tilde{r}_n\}$? In fact, the fundamentals of two finite economies $E_n$ and $E_{n'}$ may be very different from each other when $n'$ is much larger than $n$, making it difficult to see how either one relates to the infinite sequence economy.

These problems cannot be resolved by defining a probability measure on the sequence space because any such measure will assign nearly unit mass to the first $n$ assets for large enough $n$. The tail of the sequence, which presumably holds a significant part of the defining features of the economy, is left with negligible weight and hence under-represented. For example, it is difficult to construct a meaningful sampling procedure from $T = \{1, 2, \ldots\}$ such that all assets have equal chance of being represented. In the APT literature, this difficulty led to the reliance on asymptotic statements which hold as the number of assets increases to infinity, but with few implications for finite subsets of fixed size.

The peculiar problems arising in Example 1 do not arise in economies with a large finite number of assets or in economies with a continuum of assets. In both cases, one can be explicit about a measure to use in making sense of ideas involving statements like *most* assets, a *typical* asset, and so on. In fact, in Section 5.2, it is shown that these problems do not appear in the continuum-economy analogue of Example 1. A useful analogy here is that of a large exchange economy comprised of agents of one of two possible types. If we know all the features of the economy, e.g., endowments, production possibilities, the preferences of each type, etc., but not what the ratio of each type in the population is. This would be an incomplete model about which little of interest can be said about things like equilibrium prices and allocations because these concepts depend on the relative weight, or measure, of agent types in the economy.

# 2. THE MODEL

### *2.1. Assets and Returns*

There is a continuum of assets represented by the measure space $(T, \mathcal{T})$ where $T = [0, 1]$ and $\mathcal{T}$ is the set of Lebesgue measurable subsets of $T$. The supply of assets is represented by a probability distribution $\tau$ on $(T, \mathcal{T})$ which assigns to each subset $A \subset T$ its weight $\tau(A)$ relative to the entire asset market.

To formalize the intuition of a market consisting of a large number of negligible assets, it is natural to assume $\tau$ to be non-atomic. That is, each asset $t$ has a negligible weight $\tau(t) = 0$ in the economy. For simplicity, assume that $\tau$ is the Lebesgue measure on $[0,1]$. All the results go through if we use any other distribution on the space of assets provided it is absolutely continuous with respect to the Lebesgue measure.

Assets have uncertain returns. Formally, there is a probability space $(\Omega, \Sigma, P)$ such that asset $t$ pays a rate of return $\tilde{r}_t(\omega)$ in state $\omega \in \Omega$.[4] Using $L_2$ to denote the space of random variables with finite mean and variance, an *asset return process* is a function $\mathbf{r} : [0, 1] \to L_2$ assigning a random return $\tilde{r}_t \in L_2$ to asset $t$.[5]

Define the inner product $(\tilde{x} \mid \tilde{y}) = \int_\Omega \tilde{x}\,\tilde{y}\,dP$ and the $L_2$-norm of $\tilde{x}$ is $\|\tilde{x}\| = (\tilde{x} \mid \tilde{x})^{1/2}$. Using $E$, $var$, $cov$ to denote expectation, variance and covariance, respectively, we have $(\tilde{x} \mid \tilde{y}) = cov(\tilde{x}, \tilde{y}) + E\tilde{x}\,E\tilde{y}$, and $\|\tilde{x}\|^2 = var(\tilde{x}) + (E\tilde{x})^2$. Bold face letters will always denote processes (*i.e.*, functions from $[0,1]$ into $L_2$); letters with a tilde '~' will be used to represent random variables; and letters with a bar '¯' will denote expected values. Thus, $\tilde{r}_t$ is a random variable whose expected value is $\bar{r}_t$. The symbol $\mathbf{r}$ then denotes the function $\mathbf{r} : T \to L_2$ defined by $\mathbf{r}_t = \tilde{r}_t$.

### 2.2. The Covariance Structure of Asset Returns

A process $\mathbf{r}$ determines an *expected return function* $E_{\mathbf{r}} : T \to I\!R$, where $E_{\mathbf{r}}(t) = r_t$ is the expected rate of return of asset $t$. In addition, we also have a *covariance function* $Cov : T \times T \to I\!R$, where $Cov_{\mathbf{r}}(t,s) = cov(\tilde{r}_t, \tilde{r}_s)$ is the covariance between the rates of return on assets $t$ and $s$. The diagonal $t \mapsto Cov_{\mathbf{r}}(t,t) = Var_{\mathbf{r}}(t)$ defines the *variance function*. The term *covariance structure* will refer to the functions $E_{\mathbf{r}}$ and $Cov_{\mathbf{r}}$ (the subscript will be omitted when $\mathbf{r}$ is clear from the context). A process $\mathbf{r}$ has a *measurable covariance structure* if $E_{\mathbf{r}}$ and $Cov_{\mathbf{r}}$ are Lebesgue measurable.[6] I also maintain the mild assumptions that $E_{\mathbf{r}} : T \to I\!R$ and $Var_{\mathbf{r}} : T \to I\!R$ are bounded.

The function $Cov$ may be thought of as representing the entries $Cov(t,s)$ of a 'matrix' with a continuum of rows and columns. $Cov$ is therefore a generator of the covariance matrix of every possible finite subset of assets $\{t_1, \cdots, t_n\}$ drawn from $T$.

### 2.3. Idiosyncratic Processes

A process $\mathbf{h}$ is *idiosyncratic* if $Cov_{\mathbf{h}}(t,s) = 0$ almost everywhere on $T \times T$. This definition formalizes the intuition that correlations in an idiosyncratic process must be 'sparse', so the corresponding risks are negligible in the aggregate. Proposition A.1 in the Appendix shows that this definition is in fact equivalent to any one of two seemingly stronger technical conditions on $\mathbf{h}$.

Our definition of idiosyncratic residuals is a natural extension of the standard definition of idiosyncratic residuals for finite sets of random variables: An alternative (in fact equivalent) way to interpret the definition is to say that if we draw a subset of $N$ assets from the underlying economy at random, then they will be uncorrelated with probability 1. In particular, the $N \times N$ covariance matrix corresponding to this finite subset of assets will be diagonal with probability 1 (see Sections 4 and 5.6 for more detailed discussion).

### 2.4. Factor Spaces, Projections and Factor Rotations

A *factor space* is any linear subspace $F$ spanned by a finite subset of zero-mean random variables in $L_2$. The orthogonal projection $\text{Proj}_F \tilde{x}$ of a random variable $\tilde{x} \in L_2$ on $F$ represents the *F-factor risk*; *i.e.*, the part of total risk that can be explained by $F$. The difference $\tilde{x} - x - \text{Proj}_F \tilde{x}$ is a random variable orthogonal to the subspace $F$, and represents *residual risk* which cannot be explained by $F$.

It is often convenient to work with factors rather than factor spaces. A *set of factors* $\Delta = \{\tilde{\delta}_1, \ldots, \tilde{\delta}_K\}$ for $F$ is any orthonormal basis for $F$ (*i.e.*, $F = span\,\Delta$, $\delta_k = 0$, $var(\tilde{\delta}_k) = 1$, and $cov(\tilde{\delta}_k, \tilde{\delta}_s) = 0$ for all $k \neq s$). A set of factors $\Delta'$ is a *rotation* of $\Delta$ if they span the same factor space.

Orthogonal projections have a simple representation relative to a set of factors $\Delta$:

$$\text{Proj}_F \tilde{x} = \beta_1 \tilde{\delta}_1 + \cdots + \beta_K \tilde{\delta}_K,$$

where the $\beta_k$'s are real numbers called the *factor loadings*, or *betas*, of $\tilde{x}$ relative to $\Delta$, and represent the sensitivity or exposure of $\tilde{x}$ to the corresponding factor risks. Geometrically, $\text{Proj}_F \tilde{x}$ is a coordinate-free description of the orthogonal projection of $\tilde{x}$ on $F$, while $\sum_k \beta_k \tilde{\delta}_k$ is its representation relative to the basis $\{\tilde{\delta}_1, \ldots, \tilde{\delta}_K\}$.

### 2.5. Examples

<u>*I.i.d. Processes:*</u> A process $\mathbf{h}$ is an i.i.d. process if, for every finite set of indices $\{t_1, \ldots, t_N\}$, the random variables $\{\tilde{h}_{t_1}, \ldots, \tilde{h}_{t_N}\}$ are independently and identically distributed with mean $\mu$ and variance $\sigma$. Clearly, $E_{\mathbf{h}}$ is a constant function which assumes the value $\mu$. The covariance function $Cov_{\mathbf{h}}$ is zero on $T^2$ except on the diagonal where it is equal to $\sigma$. Obviously, $\mathbf{h}$ has a measurable covariance structure and is, in fact, idiosyncratic provided $\mu = 0$.

*Finitely Generated Processes:* A process $\mathbf{g}$ is finitely generated if there is a set of factors $\tilde{\delta}_1, \ldots, \tilde{\delta}_K$ such that

$$\tilde{g}_t = r_t + \sum_{k=1}^{K} \beta_{kt}\tilde{\delta}_k, \qquad \tau - \text{a.e.}$$

where $\alpha_t, \beta_{1t}, \ldots, \beta_{Kt}$ are bounded, measurable real-valued functions. Since $\mathbf{g}$ is bounded, $E_{\mathbf{g}}(t) = r_t$ and $Cov_{\mathbf{g}}(t,s) = \sum_{k=1}^{K} \beta_{kt}\beta_{ks}$, the process $\mathbf{g}$ has a measurable covariance structure.

## 2.6. *Strict K-Factor Structures*

A process $\mathbf{r}$ has a strict $K$-factor structure if there is a $K$-factor space $F$ such that

$$\tilde{r}_t = r_t + \text{Proj}_F \tilde{r}_t + \tilde{h}_t \tag{2.1}$$

such that (1) $\mathbf{h}$ is idiosyncratic; (2) $\text{Proj}_F \tilde{h}_t = 0$ for $\tau$ - a.e. $t$; and (3) $F$ is *minimal* in the sense that there is no proper subspace $F' \subset F$ with these properties.

This definition closely parallels the traditional definition of strict factor structure: The risk $\tilde{r}_t - r_t$ can be decomposed into $F$-factor risk $\text{Proj}_F \tilde{r}_t$ and an idiosyncratic risk $\tilde{h}_t$ which cannot be explained by $F$. The minimality condition ensures that $F$ does not contain superfluous factors which do not significantly contribute to $F$'s ability to explain asset returns.

If $\Delta = \{\tilde{\delta}_1, \ldots, \tilde{\delta}_K\}$ is any set of factors for $F$, then a process with a strict $K$-factor has the familiar representation:

$$\tilde{r}_t = r_t + \beta_{1t}\tilde{\delta}_1 + \cdots + \beta_{Kt}\tilde{\delta}_K + \tilde{h}_t \tag{2.2}$$

A rotation $\Delta'$ of the factors will change the representation (2.2) by changing the factor loadings, but will have no effect on the decomposition of risk into systematic and idiosyncratic parts.

## 3. ASSET PRICING UNDER STRICT FACTOR STRUCTURE

In this section I derive exact factor-pricing under the assumption that returns have a strict factor structure. While this assumption is strong, its use simplifies the comparison between the framework of this paper and much of the work on the APT where this assumption is common. Sections 5 and 6 will be concerned with the implications of dropping this assumption.

### *3.1. Portfolios*

A *portfolio* $\mathbf{w}$ is characterized by its *support* $\{t_1, \ldots, t_n\} \subset T$ and a corresponding set of *portfolio weights* $\{\alpha_1, \ldots, \alpha_n\} \subset I\!R$. The cost of a portfolio $\mathbf{w}$ is $C(\mathbf{w}) = \sum_{i=1}^{n} \alpha_i$. I follow the literature by assuming that $C(\mathbf{w}) = 1$, so $\alpha_i$ represents the percentage of the portfolio invested in asset $t_i$. A negative $\alpha_i$ reflects the possibility that asset $t_i$ is sold short.[7] A portfolio $\mathbf{w}$ defines a random return $\tilde{w} = \sum_{i=1}^{n} \alpha_i \tilde{r}_{t_i}$ whose expected return $w$ and variance $var(\tilde{w})$ are defined in the usual way.

An *arbitrage portfolio* is the difference between two portfolios. That is, $\mathbf{w}$ is an arbitrage portfolio if it has the form $\mathbf{w} = \mathbf{w}_+ - \mathbf{w}_-$, in which case $\mathbf{w}_+$ and $\mathbf{w}_-$ will be referred to as the positive and the negative parts of $\mathbf{w}$. An arbitrage portfolio costs nothing because it finances the purchase of one portfolio $\mathbf{w}_+$ by short selling another portfolio $\mathbf{w}_-$. The 'rate' of return of an arbitrage portfolio $\mathbf{w}$ is the random variable $\tilde{w} = \tilde{w}_+ - \tilde{w}_-$, so its expected return and variance are $w_+ - w_-$ and $var(\tilde{w}_+ - \tilde{w}_-)$.

### *3.2. Risk Premia and Arbitrage Opportunities*

For simplicity, I will assume throughout that there is an asset or a portfolio paying a riskless rate $\gamma_0$. Given this, it is useful to think of $\tilde{r}_t$ as a sum

$$\tilde{r}_t = \gamma_0 + (\tilde{r}_t - r_t) + (r_t - \gamma_0)$$

of:

i) A *riskless rate of return* $\gamma_0$;

ii) A *pure risk* $\tilde{r}_t - r_t$ representing fluctuations around the expected return $r_t$; and

iii) An *excess return* or a *risk premium* $r_t - \gamma_0$ paid to compensate for these fluctuations.

More generally, call a random variable a *pure risk* if it has zero mean. Our goal is to define a function $\upsilon$ which determines the risk premium paid for holding any pure risk generated by either an asset or by a portfolio of assets. In addition, we also want $\upsilon$ to be continuous in the sense that if two pure risks are close then so are their corresponding premia. Formally, let $L^* = \{\tilde{r}_t - r_t : t \in T\} \subset L_2$ denote the set of pure risks associated with the process $\mathbf{r}$, and let $span(L^*)$ be the closed linear space spanned by $L^*$. Then we seek a continuous linear function $\upsilon : span(L^*) \to I\!R$ which is consistent with $\mathbf{r}$ in the sense that $\upsilon(\tilde{r}_t - r_t) = r_t - \gamma_0$ for all $t$.

It is easy to find examples of return processes that are inconsistent with any risk premium function. For example, it suffices that there are two assets $t$ and $s$ which have identical pure risks (*i.e.*, $\tilde{r}_t - r_t = \tilde{r}_s - r_s$) but have different expected returns $r_t \neq r_s$.

The next result characterizes asset return processes which are consistent with a (norm) continuous linear risk premium function. We will say that an asset return process $\mathbf{r}$ *admits no arbitrage opportunities* if for every sequence of arbitrage portfolios $\{\mathbf{w}^k\}$, $var(\tilde{w}^k) \to 0$ implies $w^k \to 0$. In other words, an arbitrage opportunity would exist if one can, at no cost, make an essentially riskless investment that earns a return bounded away from zero.

**PROPOSITION 1:** *There is a continuous linear risk premium function $\upsilon : span(L^*) \to I\!R$ consistent with $\mathbf{r}$ if and only if $\mathbf{r}$ admits no arbitrage opportunities. If such a function exists, then it is unique.*

This result confirms the relationship between the absence of arbitrage opportunities and the continuity of asset prices, a result which is well-known for the traditional sequence model (*e.g.*, Chamberlain and Rothchild [5]). While additional restrictions on asset prices, such as positivity, may be natural, only continuity plays any role in the pure form of the APT.

Since $\mathbf{r}$ has a strict factor structure, the pure risk of an asset $t$ is the sum

$$\tilde{r}_t - r_t = \mathrm{Proj}_F \tilde{r}_t + \tilde{h}_t.$$

If $\upsilon$ is a linear (not necessarily continuous) risk premium function consistent with $\mathbf{r}$, then the risk premium on asset $t$ is the sum of a premium paid for holding the asset's factor risk and a premium for holding its idiosyncratic risk:

$$\begin{aligned} r_t - \gamma_0 &= \upsilon(\tilde{r}_t - r_t) \\ &= \upsilon(\mathrm{Proj}_F \tilde{r}_t + \tilde{h}_t) \\ &= \upsilon(\mathrm{Proj}_F \tilde{r}_t) + \upsilon(\tilde{h}_t). \end{aligned}$$

The ability to eliminate idiosyncratic risk through diversification in a large economy and the absence of arbitrage opportunities suggest that no premium will be paid for bearing idiosyncratic risks. Exact factor pricing means that $\upsilon(\tilde{h}_t) = 0$, so any excess return paid for an asset must be due entirely to that asset's factor risk. The next result provides an arbitrage pricing result for our model:

**PROPOSITION 2:** *If* $\mathbf{r}$ *admits no arbitrage opportunities, then for almost every asset* $t$,

$$r_t - \gamma_0 = \upsilon(\mathrm{Proj}_F \tilde{r}_t). \tag{3.1}$$

*In particular, if* $\Delta = \{\tilde{\delta}_1, \ldots, \tilde{\delta}_K\}$ *is any set of factors generating* $F$, *then there exist constants* $\gamma_1, \ldots, \gamma_K$, *representing excess returns per unit of factor risks, such that the expected return on almost every asset* $t$ *satisfies:*

$$r_t = \gamma_0 + \beta_{1t}\gamma_1 + \cdots + \beta_{Kt}\gamma_K. \tag{3.2}$$

The continuity of $\upsilon$ implies that there is a pricing vector $p \in span(L^*)$ such that the premium $\upsilon(\tilde{r}_t - r_t)$ is equal to the (inner) product of $\tilde{r}_t - r_t$ with $p$. It is easy to see that $p = \gamma_1\delta_1 + \cdots + \gamma_K\delta_K$. Since the only risk within $F$ that earns a premium is the one spanned by $p$, this might lead to the conclusion that the APT is a one-factor model (the one factor being $p$). This, however, misses a basic point: The APT's assumptions have little to say about the factor risk premia $\gamma_1, \ldots, \gamma_K$ (hence the position of $p$ within the factor space $F$). These premia would depend on such things as consumer preferences, endowments and market equilibrium, on which no restrictions are imposed by the APT's assumptions. Rather, the 'bite' of the pricing result is in restricting $p$ to lie in the factor space $F$ spanned by $\{\delta_1, \ldots, \delta_K\}$ (instead of being some arbitrary vector in $span(L^*)$).

### *3.4. Comparison with the Traditional APT*

Define the *pricing error* of asset $t$ by:

$$\begin{aligned} a_t &= r_t - \gamma_0 - \beta_{1t}\gamma_1 - \cdots - \beta_{Kt}\gamma_K \\ &=: v(\tilde{h}_t). \end{aligned}$$

That is, $a_t$ represents expected excess returns which cannot be explained by factor-pricing, and, therefore, represents an asset-specific premium. The traditional APT assumes an infinite sequence $\{\tilde{r}_1, \ldots, \tilde{r}_n, \ldots\}$ with a corresponding sequence of pricing errors $\{a_n\}$ and reaches the familiar conclusion:

$$\sum_{n=1}^{\infty} a_n^2 < \infty. \tag{3.3}$$

This conclusion is often stated as a double approximation: The pricing error *approximately* equals zero for *most* assets.[8] On the other hand, Proposition 2 implies:

$$\int_T a_t^2 \, d\tau = 0. \tag{3.4}$$

Conclusion (3.3) is, in principle, consistent with a situation in which $a_n \neq 0$ for every $n$, so not a single asset is correctly priced, while (3.4) implies that the set of assets which are exactly correctly priced has measure one.

This difference is significant for practical as well as theoretical reasons. In practice, the pricing error is assumed to be zero, usually because "currently available statistical tests are not amenable to testing approximate relationships" (Connor and Korajczyk [7]). Since the traditional APT (equation (3.3)) cannot guarantee zero pricing errors for any asset, additional assumptions on market equilibrium, investor' preferences and distributional properties of asset returns are often introduced. Shanken [16] pointed out that adding these assumptions means that empirical tests will, in fact, be tests of the joint hypothesis of the APT plus these additional assumptions, casting doubt on whether the APT is itself testable.

The difference in the conclusions (3.1) and (3.3) can be explained as follows. At a technical level, the source is the difference between the measurable structures of the sequence and continuum models. Let $A_n$ denote the subset of assets whose squared pricing errors exceed $\frac{1}{n}$. One can show that the absence of arbitrage opportunities implies that each $A_n$ is finite, and therefore 'negligible,' in a market containing an infinite number of assets. The problem in the sequence model is that, while each $A_n$ is finite, hence negligible relative to the entire economy, the limit $\cup_{n=1}^{\infty} A_n$ may be the entire space of assets, which is obviously non-negligible. This breakdown in the continuity of the notion of negligibility cannot occur in the continuum model because the countable union of negligible sets must also be negligible.

The economic interpretation of the technical observations made in the last paragraph hinges on the ratio between the number of assets needed for a given level of diversification and the total number of assets available in the economy. This ratio is not well-defined for the sequence economy. The continuum model, on the other hand, captures the basic intuition underlying the APT, namely that the economy is very large relative to the extent of diversification needed to nearly completely eliminate idiosyncratic risk.

### *3.5. Alternative Approaches*

An alternative approach that also yields exact factor pricing is based on the representation of the space of assets as an infinite sequence with a finitely additive measure in which each asset has zero weight (see Werner [18]). This, however, is not a measure space in the usual sense, so many standard probabilistic and statistical tools may be inapplicable to it. For example, dominated convergence theorem fails and the integral of a strictly positive function may be zero.[9] Moreover, it is not clear how to define and derive properties of random sampling within this framework. Also, since there are examples of sequence economies in which no single asset is correctly priced, an analogue of Proposition 2--that the set of correctly priced assets has full measure (in particular, that it is non-empty)--will not hold for a sequence model, regardless of whether the underlying measure is countably additive or not. Finally, economies modeled as purely finitely additive measures have a built-in discontinuity in the limit which makes them difficult to interpret as models of large but finite asset economies. The reason is that such interpretation is essentially a statement of continuity between large finite models and the limiting infinite model.

Another approach that tries to circumvent the problems arising in the standard sequence model is based on the techniques of non-standard analysis. In a paper subsequent to this work, Khan and Sun (1995, [11]) use such techniques to model a large economy in which assets are indexed by an atomless measure space. They arrive at asset pricing and factor structure results that are virtually identical in substance and economic interpretation to those first reported in this paper. These include their result of exact factor pricing of almost all assets, optimal extraction of sets of factors based upon a criterion of explanatory power, the decomposition of risk into factor risk and idiosyncratic risk, and the use of an infinite-dimensional analogue of the variance-covariance matrix to derive such decomposition. Khan and Sun also show that asset prices are determined by their exposures to a benchmark portfolio, in a CAPM-like relationship. They refer to this result as the "unification" of the APT and the CAPM. As noted by the Associate Editor, the existence of such portfolio is a consequence of the continuity of the pricing function (as in [5] or in Proposition 1 above). However, the fact that asset prices have a CAPM-like relationship to the pricing vector does not make the APT equivalent to the CAPM. It is well-understood that, unlike the APT, the CAPM's conclusions require restrictions on individual behavior and market equilibrium, making the two theories profoundly different in assumptions and implications (see, for example, [10, p. 178]).

## 4. IMPLICATIONS FOR PRICES IN FINITE RANDOM SAMPLES OF ASSETS

Imagine an outside observer who is interested in inferring properties of the asset pricing relationships in an unknown underlying economy using the limited information contained in a finite subset of assets. Such inference is clearly groundless without a framework within which the properties of the subset can be related to those of the underlying market. This is analogous to the case of a statistician who wants to use a given, finite set of observations to draw general conclusions: it is necessary to view the sample as a random draw from an underlying population according to some probability law.

To rationalize this inference, consider the question as a problem of statistical inference in which finite subsets of assets are interpreted as random draws from the underlying asset economy. Thinking of a given small subset of assets as a particular realization of a sampling procedure provides the necessary statistical linkage between the observed properties of the sample and the corresponding properties of the market on which inference is based. This statistical interpretation is used (here and in Section 6.2) to derive rather strong implications. I begin with a formal description of the sampling model.

### *4.1. The Sampling Model*

Let $(T^\infty, \mathcal{T}^\infty)$ be the set of infinite sequences of assets with the product $\sigma$-algebra, and define $T^n$ to be the $n$-fold product of $T$, which we view as a subset of $T^\infty$ in the usual way. Points $(t_1, \ldots, t_n) \in T^n$ will be interpreted as randomly drawn subsets of $n$ assets, while $(t_1, t_2, \ldots) \in T^\infty$ will be interpreted as a random draw of an entire infinite sequence economy. I will assume, for concreteness, that assets are drawn independently using the distribution $\tau$. Formally, sequences of assets $(t_1, t_n, \ldots) \in T^\infty$ are drawn according to the product measure $\tau^\infty$ on $T^\infty$. The $n$-fold product of $\tau$ will be denoted $\tau^n$ and will represent the probability law generating finite draws $(t_1, \ldots, t_n)$.

It is important to emphasize that the assumption that draws are made independently according to $\tau$ is made here as a simple way to illustrate a basic point: One can view the continuum model as a sample space representing an outside observer's abstract model of the economy from which finite samples of assets are drawn. One can, for example, extend this by introducing correlations and biases in the way assets are sampled or by making the likelihood of picking a particular asset depend on the price functions. All that is required is that the sampling law generates draws which are representative of the underlying population (*e.g.*, samples will not be concentrated in a subregion of $T$).

### *4.2. Pricing Result for Finite Samples*

Combining the sampling model of Section 4.1 with Proposition 1 yields the result that there is probability 1 that, in a randomly drawn finite sample, all assets are exactly factor priced:

**PROPOSITION 3:** *Suppose that* **r** *admits no arbitrage opportunities. Then, for any sample size* $n$,

$$\tau^n\{(t_1,\ldots,t_n) : a_{t_i} = 0, \quad \text{for } i = 1,\ldots,n\} = 1. \tag{4.1}$$

Proposition 3 is a statement about the properties of a typical or average subset of assets, rather than about pricing errors in a *particular sample*. The point is that, while pricing errors can be large in particular subsets, these subsets are unlikely to be drawn in the sense of the sampling model of Section 4.1.

Proposition 3 shows only the testability *in principle* of the asset pricing relationships. The reason for this qualification is that the expression $a_{t_i} = 0$ in (4.1) involves three quantities not directly observed in practice: (1) the expected return on asset $t_i$; (2) the factor loadings of each asset $\beta_{kt_i}$; and (3) the factor's risk premia $\gamma_k$. However, estimating these quantities is an issue distinct from whether the APT itself is testable. The focus of the debate about the theoretical feasibility of testing the APT was whether the APT had any implication at all for samples of fixed finite size, even assuming that the quantities 1-3 are perfectly known.The significance of Proposition 3 is that it gives a sharp affirmative answer to this question.[10]

### *4.3. Comparison with Testing the Traditional APT*

The sampling model of Section 4.1 provides an interesting perspective on the implications of the traditional APT for finite subsets of assets. In our context, the traditional APT may be viewed as a theory of pricing and returns for a fixed infinite draw of assets $(t_1^*, t_2^*, \ldots)$. The theory provides no sampling space $(T^\infty, \mathcal{T}^\infty)$ or a sampling procedure $\tau^\infty$ to explain how this draw was generated or what relationship might exist between its properties and the properties of the underlying asset market. In particular, the concepts of 'representative' versus 'exceptional' draws cannot be given formal meaning in the traditional model. Lacking an underlying sampling story, the traditional APT does its best to derive an asset pricing conclusion which is valid for any *arbitrary* sequence of assets. The asymptotic conclusion $\sum_k a_t^2 < \infty$ appears to be the strongest such statement.

By contrast, the framework of this paper focuses on statements about the space of *all* possible draws $(T^\infty, \mathcal{T}^\infty)$ and interprets the drawn sequence $(t_1^*, t_2^*, \ldots)$ as the outcome of a random sampling experiment. Interestingly, while for an arbitrary draw one can, at best, obtain a weak asymptotic result, a substantially sharper statement can be made about the probability-1 subset of representative draws.

The reader might wonder whether an analogue of Proposition 3 can be obtained for the sequence model by defining a sampling structure on $(t_1^*, t_2^*, \ldots)$ along the lines of Section 4.1. Any such attempt must confront the problem that a probability measure $\lambda$ on the infinite sequence will assign an arbitrarily large mass to the first $N$ assets for large enough $N$. Large samples drawn using $\lambda$ will give excessive weight to the first $N$ assets. Sampling according to $\lambda$ will not produce a sequence analogue to Proposition 3 because the APT is consistent with large pricing errors in the first $N$ assets.

# 5. APPROXIMATE FACTOR STRUCTURES: EXISTENCE, OPTIMALITY AND ROBUSTNESS

## *5.1. The Explanatory Power of Factor Spaces*

Recall that $span(L^*)$ is the closed linear space spanned by all of the form $\tilde{r}_t - r_t$, for some asset $t$. Let $\mathcal{F}_K$ denote the set of all factor spaces in $span(L^*)$ of dimension $K = 0, 1, 2, \ldots$ and let $\mathcal{F}$ be the set of all finite dimensional factor spaces (*i.e.*, $\mathcal{F} = \cup_{K=1}^{\infty} \mathcal{F}_K$). The *explanatory power* is the function $V : \mathcal{F} \to I\!R$ defined by:

$$V(F) := \frac{\int_T \operatorname{var}(\operatorname{Proj}_F \tilde{r}_t)\, d\tau}{\int_T \operatorname{var}(\tilde{r}_t)\, d\tau}.$$

If $\Delta = \{\tilde{\delta}_1, \ldots, \tilde{\delta}_K\}$ is a set of factors which span the factor space $F$, then we will write $V(\tilde{\delta}_1, \ldots, \tilde{\delta}_K)$ to denote $V(F)$. Division by $\int_T \operatorname{var}(\tilde{r}_t)\, d\tau$ normalizes $V$ so that $0 \le V(F) \le 1$ for every $F$, but otherwise plays no role in the analysis.[11] I assume throughout that $\mathbf{r}$ is a bounded asset return process with measurable covariance structure. Under these assumptions, Proposition A.2 in the Appendix shows that $V$ is well-defined.

To motivate the definition of $V$, note that $\operatorname{Proj}_F \tilde{r}_t$ represents the part of asset $t$'s return that can be explained by the factor space $F$. The numerator $\int_T \operatorname{var}(\operatorname{Proj}_F \tilde{r}_t)\, d\tau$ represents a measure of the average variation in returns that can be explained by $F$. Therefore, $V(F)$ is the average variation explained by $F$ as a percentage of the average total variation in asset returns. The reader might find it useful to note the similarity between the definition of $V(F)$ and the standard definition of $R^2$ in Statistics. Both concepts attempt to measure the average goodness of fit relative to the linear subspace spanned by a given set of regressors.

It is worth noting that the function $V$ can be alternatively defined in terms of the gross (rather than the rates of) return on assets. This amounts to modifying the measure $\tau$ to take account of the market values of the various assets. While this has no effect on the results at a technical level, the interpretation of what constitutes a 'good' set of factors will, of course, change. Basically, the measure $\tau$ is a criterion for ranking subsets of assets by their relative importance in the economy and what constitutes a good fit will reflect that. This is similar to the usual definition of $R^2$, reflecting the implicit criterion that all sample points are given the same weight.

### 5.2. *Optimal Factor Extraction*

One way to think of the problem of finding an approximate factor structure is to follow a sequential procedure: (1) Start with the factor $\tilde{\delta}_1^*$ that has the highest explanatory power; (2) "regress" $\mathbf{r}$ on $\tilde{\delta}_1^*$ to obtain a residual process $\mathbf{r}_2$ in which all systematic variations explained by $\tilde{\delta}_1^*$ have been removed; (3) Repeat these two steps with the return process $\mathbf{r}_2$ to extract a new factor $\tilde{\delta}_2^*$, and so on. The resulting *optimal sequence of factors* $\{\tilde{\delta}_1^*, \tilde{\delta}_2^*, \ldots\}$ generates a sequence of factor spaces $F_K = span\{\tilde{\delta}_1^*, \ldots, \tilde{\delta}_K^*\}$ with increasing explanatory powers. The following proposition formalizes this intuition and shows, in particular, that this sequential method of extracting factors is well-defined:

**PROPOSITION 4:** *For any process* $\mathbf{r}$,

i) *There is* $\tilde{\delta}^*$ *such that* $V(\tilde{\delta}^*) \geq V(\tilde{\delta})$ *for every factor* $\tilde{\delta}$.

ii) *An optimal sequence of factors exists. That is, there is a sequence* $\{\tilde{\delta}_1^*, \tilde{\delta}_2^*, \ldots\}$ *satisfying*

$$V(\tilde{\delta}_1^*) = \max_{\tilde{\delta}} V(\tilde{\delta})$$

$$V(\tilde{\delta}_K^*) = \max_{\tilde{\delta} \perp \{\tilde{\delta}_1^*, \ldots, \tilde{\delta}_{K-1}^*\}} V(\tilde{\delta}), \qquad K = 2, \ldots$$

iii) $\mathbf{r}$ *has strict K-factor structure if and only if there is an optimal sequence of factors* $\{\tilde{\delta}_1^*, \tilde{\delta}_2^*, \ldots\}$ *with* $V(\tilde{\delta}_K^*) > V(\tilde{\delta}_{K+1}^*) = 0$.

Recall that Example 1 in the Introduction described a sequence economy with no approximate factor structure. In that example, there was no obvious way to rank two factors $\tilde{\eta}_m$ and $\tilde{\eta}_{m'}$ according to their explanatory power. By contrast, consider the following example which gives a continuum-economy analogue to Example 1:

*Example 2:* Let $\{\tilde{\eta}_m\}$ be a sequence of i.i.d.random variable with unit variance and zero mean. Call a process $\mathbf{r}$ *countably simple* if for every $t$, $\mathbf{r}_t = \tilde{\eta}_m$ for some $m$. If we define $A_m = \{t : \mathbf{r}_t = \tilde{\eta}_m\}$, then the measurability of $\mathbf{r}$ implies that the $A_m$ form a countable partition of $T$ by measurable sets, and that $V(\tilde{\eta}_m) = \tau(A_m)$. Assume further that $\tau(A_m) > 0$ for all $m$, so there is an infinite number of non-trivial factors.

In Example 2, it is always possible to rank factors by their explanatory power. An approximate factor structure can then be found by looking for a set of factors with the largest explanatory power. By contrast, the problem in Example 1 was the lack of an obvious criterion to meaningfully compare the relative size of the sets of assets whose returns are given by $\tilde{\eta}_m$ and $\tilde{\eta}_{m'}$ respectively.

### *5.3. Optimal Approximate Factor Structures*

Approximate factor structures are useful because identify the most significant factors and discard factors which contribute little to explaining asset returns. There are two reasons why such approximation may be important. First, the underlying process $\mathbf{r}$ may be one with no strict factor structure at all (as in the case of Examples 1 and 2), so approximation is the only way to get a factor representation. Second, the definition of a strict factor structure can, in some cases, be "sufficiently stringent that it is unlikely that any large asset market has ... a usefully small number of factors". (Chamberlain and Rothchild [5], p. 1282). Thus, even if a strict $K$-factor structure existed, $K$ might be so large that a more useful model would be an approximate factor model with $L < K$ factors. To illustrate this, consider the following example:

*Example 3:* Let $\{\tilde{\eta}_m\}$ be as in Examples 1 and 2 and let $0 < b < a < 1$. Define the process $\mathbf{r}$ by setting $\tilde{r}_t = \tilde{\eta}_1$ on $(a, 1]$ and $\tilde{r}_t = \tilde{\eta}_2$ on $(b, a]$. Divide $[0, b]$ into $2^n$ equal subintervals (the end-points will not matter), and let $\tilde{r}_t = \tilde{\eta}_3$ on the subinterval with the highest endpoints, $\tilde{r}_t = \tilde{\eta}_4$ on the subinterval immediately to its left, and so on.

This process has strict factor structure with $2^n + 2$ factors. The number of factors needed to represent asset returns using a strict factor structure increases exponentially with $n$. On the other hand, if $b$ is close to 0, an approximate factor model with two factors $\{\tilde{\eta}_1, \tilde{\eta}_2\}$ performs well at explaining 'most of the variation in most of the assets.'

To formalize this, an *optimal L-factor structure* for a process $\mathbf{r}$ is a factor space $F_L$ with dimension at most $L$ such that it has the highest explanatory power among all other factor spaces of dimension at most $L$, and parsimonious in the sense of containing no superfluous factors that make no contribution to its explanatory power. Formally,

i) $\hat{F}_L$ solves $\max_{\dim F \leq L} V(F)$;

ii) $F_L$ is *minimal*: $\dim F < L$ implies $V(F) < V(F_L)$.

**PROPOSITION 5:** *If* $\mathbf{r}$ *has a* $K$*-strict factor structure, then an optimal approximate* $L$*-factor structure* $F_K$ *exists for each* $L$.[12]

It is worth emphasizing that this result has no counterpart in the sequence approach: If a sequence economy has a strict factor structure with $K$ non-trivial factors,[13] then it cannot have an approximate $L$-factor structure for any $L < K$.

### *5.4. Asymptotic Properties*

One often finds in the literature the intuition that if sufficiently many relevant factors are included in the model, one will eventually be able to capture nearly all systematic variations in asset returns, leaving residuals which are approximately idiosyncratic. Example 1 demonstrates that this intuition is difficult to formalize (let alone prove) within the sequence model.

The next proposition confirms this intuition in the model with a continuum of assets by investigating the asymptotic properties of factor spaces as the number of factors increases. Before stating the theorem, we need the following two definitions:

$$V_K^{\max} = \sup_{F \in \mathcal{F}_K} V(F)$$

$$V^{\max} = \sup_{F \in \mathcal{F}} V(F).$$

**PROPOSITION 6:**

i) $V_K^{\max} \uparrow V^{\max}$ *as* $K \to \infty$;

ii) $V(\{\tilde{\delta}_1^*, \ldots, \tilde{\delta}_K^*\}) \to V^{\max}$ *as* $K \to \infty$ *for any optimal sequence of factors* $\{\tilde{\delta}_1^*, \tilde{\delta}_2^*, \ldots\}$;

iii) *There exists a unique minimal factor space* $F_\infty$ *such that* $V^{\max} = V(F_\infty)$;

iv) *The residual process* $\tilde{h}_t = \tilde{r}_t - r_t - \mathrm{Proj}_{F_\infty} \tilde{r}_t$ *is idiosyncratic.*

Propositions 4-6 refine and sharpen a related result on the decomposition of risk in abstract settings in Al-Najjar [1]. The key improvement here is that the present framework gives an efficient and parsimonious way to extract the factors. This difference is crucial in applications: for example, if **r** has a strict 1-factor structure, then, by Proposition 4 (i), the true optimal factor can be found. Al-Najjar [1] showed only that there is a countable set of factors spanning the range of the aggregate part of **r**.

The proofs offered here are also new, and in fact independent of the ones found in Al-Najjar [1]. The main innovation is the introduction of the function $V$ which, in addition to providing a better intuition, also makes it possible to develop an elementary proof of decomposition (Proposition A.4).[14]

### *5.5. Reference Variables*

In practice, the true factor space will not be known a priori. Suppose, for example, that $\{\tilde{\delta}_1, \ldots, \tilde{\delta}_K\}$ is a strict factor structure for a given asset economy **r**. An empirical investigation will typically have to rely on a set of *proxies* or *reference variables* to approximate these factors. Such proxies will generally not perform as well as the true factors in explaining asset returns, but they might be expected to perform reasonably well if they happen to be highly correlated with the true factors.

To formalize this, consider two sets of factors $\{\tilde{\delta}_1, \ldots, \tilde{\delta}_K\}$ and $\{\tilde{\delta}'_1, \ldots, \tilde{\delta}'_K\}$. Since factors are scaled to have norm one, $(\tilde{\delta}_k \mid \tilde{\delta}'_k)$ is the correlation coefficient between $\tilde{\delta}_k$ and $\tilde{\delta}'_k$. A sequence of sets of factors $\{\tilde{\delta}_1^n, \ldots, \tilde{\delta}_K^n\}_{n=1}^\infty$ converges to $\{\tilde{\delta}_1, \ldots, \tilde{\delta}_K\}$ if $\min_k (\tilde{\delta}_k \mid \tilde{\delta}_k^n) \to 1$ as $n \to \infty$. In words, two sets of factors are close if each factor in the first set is highly correlated with the corresponding factor in the second set.[15]

**PROPOSITION 7:** *Suppose that, for each n, $\{\tilde{\delta}_1^n, \ldots, \tilde{\delta}_K^n\}$ is a set of reference variables for $\{\tilde{\delta}_1, \ldots, \tilde{\delta}_K\}$ with corresponding factor spaces $F^n$ and $F$. Then $\min_k(\tilde{\delta}_k, \tilde{\delta}_k^n) \rightarrow 1$ implies that $V(F^n) \rightarrow V(F)$.*

In the context of the sequence model, Reisman [13] pointed out that a set of reference variables obtained through a slight perturbation of the factors might not constitute an approximate factor structure. This is problematic because estimates of the true factors will typically be based on the limited - and noisy - information available from observations of asset returns; so, they should not be expected to coincide with the true factors in most situations of interest.

Proposition 7 shows that small errors in estimating a set of factors produce only small differences in the resulting explanatory power. In particular, if **r** has a strict factor structure with factor space $F$, and if $F'$ is a set of reference variables sufficiently close to $F$, then **r** has an approximate factor structure relative to $F'$. The proposition therefore suggests that the lack of robustness in the sequence model is due to the difficulty in assigning relative weights to subsets of assets in a meaningful way. To see this, note that while there is no difficulty in defining a function analogous to $V$ in the sequence context, such definition requires an explicit measure on the space of assets. But any such measure will necessarily put a mass of almost 1 on the first $N$ assets, for $N$ large enough, thus ignoring assets in the tail of the sequence.

### *5.6. Relationship to Chamberlain and Rothchild's Definition*

Fix the sequence economy $\{\tilde{r}_{t_1}, \tilde{r}_{t_2}, \ldots\}$ and let $\Sigma_N$ be the covariance matrix of the first $N$ assets. Chamberlain and Rothchild [5] say that this sequence has "an approximate $K$-factor structure if and only if exactly $K$ of the eigenvalues of the covariance matrices $\Sigma_N$ increase without bound and all other eigenvalues are bounded." (p. 1284).[16]

The definition of an approximate factor structure given earlier differs from Chamberlain and Rothchild's in a number of important respects. First, in our definition it is possible to evaluate the relative performance of alternative candidate factor spaces, while, in Chamberlain and Rothchild's definition, a sequence economy either has an approximate factor structure of some order $K$, or it has no approximate factor structure at all. Second, our definition allows for a greater range of asset return processes to have approximate factor structures than suggested in Chamberlain and Rothchild's definition. Consider the return process in the following example:

*Example 4:* Let $\{\tilde{\eta}_m\}$ be as in Examples 1 and 2, and define the process $\mathbf{r} = \sum_k \beta_{kt}\tilde{\eta}_k$ by letting $\beta_{kt}$ be equal to 1 on the half open interval $\left(\frac{1}{2^{-(k+1)}}, \frac{1}{2^{-k}}\right]$ and zero otherwise. Thus, $\beta_{1t}$ is the characteristic function of $\left(\frac{1}{2}, 1\right]$, $\beta_{2t}$ is the characteristic function of $\left(\frac{1}{4}, \frac{1}{2}\right]$, and so on.

With $\tau^\infty$-probability 1, any sequence economy $\{\tilde{r}_1, \tilde{r}_2, \ldots\}$ drawn randomly from $T$ will fail to have a factor structure in the sense of Chamberlain and Rothchild. The reason, roughly, is that a typical sequence will contain infinitely many points in each interval $\left(\frac{1}{2^{-(k+1)}}, \frac{1}{2^{-k}}\right]$, so each random variable $\tilde{\eta}_k$ must be included as a factor. By contrast, it is intuitively clear (and Proposition 4 formally proved) that $\mathbf{r}$ has an approximate factor structure because, for moderately large $k$, the set of factors $\{\tilde{\eta}_1, \ldots, \tilde{\eta}_k\}$ will be enough to explain the variation in returns on most assets. Chamberlain and Rothchild's criterion of the number of exploding eigenvalues ignores the rate at which different eigenvalues explode. In Example 4, the eigenvalue corresponding to $\tilde{\eta}_k$ for large $k$ explodes at a slower rate than, say, the one corresponding to $\tilde{\eta}_1$.

## 6. APPROXIMATE FACTOR-PRICING

### *6.1. The Approximate Pricing Theorem*

**PROPOSITION 8:** *Suppose that* **r** *admits no arbitrage opportunities. Then, for every $\epsilon > 0$, there is a $K$-factor space $F_K$ and a subset of assets $A \subset T$ with $\tau(A) > 1 - \epsilon$ such that the pricing error $a_t$ relative to $F_K$ for every $t \in A$ satisfies:*

$$|a_t| = |\bar{r}_t - \gamma_0 - \psi\left(\text{Proj}_{F_K}\tilde{r}_t\right)| < \epsilon.$$

The double approximation in Proposition 8 might initially suggest that the pricing result in this proposition is similar to the approximate factor-pricing theorems in the sequence approach. A number of important differences should be emphasized. First, Proposition 8 does not assume **r** to have either a strict or approximate $K$-factor structure. Second, the approximation in Proposition 8 improves as new factors are introduced. This contrasts with the approximate arbitrage result of Chamberlain and Rothchild [5] where the quality of the APT approximation is not defined, and so increasing the number of factors had no clear effect. Finally, the pricing formula in Proposition 8 has different implications for asset returns and prices in finite subsets of assets.

The intuition for Proposition 8 may be explained as follows. Given an approximate factor space $F_K$, Proposition 6 can be used to show that there is a unique linear subspace $H$ such that the rate of return on any asset $t$ can be written (uniquely) in the form:

$$\tilde{r}_t = \bar{r}_t + \text{Proj}_{F_K}\tilde{r}_t + \text{Proj}_H\tilde{r}_t + \tilde{h}_t,$$

where $\tilde{h}_t$ is idiosyncratic. If there are no arbitrage opportunities, the risk premium function $\psi$ will have the properties asserted in Proposition 1, so we may write:

$$\bar{r}_t - \gamma_0 = \psi\left(\text{Proj}_{F_K}\tilde{r}_t\right) + \psi\left(\text{Proj}_H\tilde{r}_t\right) + \psi(\tilde{h}_t).$$

The only difference between this equation and the corresponding one in Section 3.3 (where a strict factor structure is assumed) is the appearance of the term $\psi\left(\text{Proj}_H\tilde{r}_t\right)$. The term $\text{Proj}_H\tilde{r}_t$ represents additional systematic risks not captured by the factor space $F_K$ (which I will refer to as *H-factor risk*). While an asset's exposure to $H$-factor risk will contribute to that asset's excess return, a rich enough $F_K$ will ensure that most assets' exposure to that risk is small, implying that $\text{Proj}_H \bar{r}_t$ is close to zero for most assets.

### 6.2. Pricing Result for Finite Subsets of Assets

For finite subsets of assets drawn randomly from the underlying asset economy, Proposition 9 roughly states that there is a high probability of drawing a sample in which a large percentage of assets are approximately correctly priced.

**PROPOSITION 9:** *Suppose that* $\mathbf{r}$ *admits no arbitrage opportunities. For every* $\epsilon > 0$ *there is a sample size* $n$ *and a factor space* $F \in \mathcal{F}$ *such that*

$$\tau^n \left\{ (t_1, \ldots, t_n) : \frac{\text{number of assets } t_i \text{ with } |a_{t_i}| < \epsilon}{n} > 1 - \epsilon \right\} > 1 - \epsilon.$$

One difference between this result and Proposition 4 for asset economies with strict factor structures is the role played by the sample size $n$. In Proposition 4, $n$ played no role; in particular, a larger sample size presented no advantage as far as testing the APT was concerned. In Proposition 9, the approximate factor space $F$ does not necessarily capture all systematic risk. Thus, there may well be a subset of assets in the economy with too high an exposure to $H$-risks to have their excess return adequately explained by $F$. In this case, a larger sample size is important because it reduces the chance of drawing a subset in which assets with high exposure to $H$-risk are over-represented.

### 6.3. Pricing with Reference Variables

Consider a sequence economy with a strict factor space $F$. Reisman argued that the traditional APT's main pricing result $\sum_n a_n^2 < \infty$ is valid with respect to almost any factor space $F'$. This result is disturbing because $F'$ "might account for only a trivial fraction of the common variation in security returns and still an APT approximation [...] must hold" (Shanken [17], p. 1570).

I will now show that this situation does not arise when the underlying space of assets is represented by $T = [0, 1]$ and point out that the disturbing implications of replacing the factor space by a set of reference variables is an artifact of the sequence approach to modeling large asset markets.

For simplicity, assume that $\mathbf{r}$ is a one-factor asset return process $\tilde{r}_t = \beta_t \tilde{\delta} + \tilde{h}_t$ with no arbitrage opportunities and let $\psi$ denote its unique continuous pricing function. By Proposition 2, we have $\psi(\tilde{h}_t) = 0$ except on a set of assets, $A$, of measure zero. If $\tilde{\delta}' \in span(L^*)$, $\tilde{\delta}' \neq \tilde{\delta}$, is any reference variable orthogonal to the idiosyncratic components $\tilde{h}_t$ for all $t \subset A$.[17] then we must have $\psi(\tilde{\delta}) > \psi(\tilde{\delta}') > 0$. So

$$\psi(\mathrm{Proj}_{\tilde{\delta}'} \tilde{r}_t) \;=\; \beta_t(\tilde{\delta}' \,|\, \tilde{\delta})\psi(\tilde{\delta}') \neq \; \beta_t \psi(\tilde{\delta}) \;=\; \psi(\mathrm{Proj}_{\tilde{\delta}} \tilde{r}_t)$$

for almost every asset $t$. On the other hand, Proposition 2 implies that $a_t^2 = (\bar{r}_t - \gamma_0 - \psi(\mathrm{Proj}_{\tilde{\delta}} \tilde{r}_t))^2 = 0$ for almost every asset. Using $a_t'$ to denote the pricing error obtained when the factor model is misspecified as $\tilde{r}_t = \beta_t \tilde{\delta}' + \tilde{h}_t$, this implies that $a_t' \neq 0$ for almost all $t$, hence

$$\int_T (a_t')^2 \, d\tau > \int_T a_t^2 \, d\tau = 0.$$

That is, the quality of the factor-pricing result (measured by the average pricing errors) deteriorates as the true factor is replaced by a reference variable. This shows that the concerns raised in the literature on the use of proxies are not serious in a large economy with a continuum of assets. Note further that since $(\tilde{\delta}' \; \tilde{\delta}) \rightarrow 1$, $\tilde{\delta}'$ is an increasingly accurate estimate of the true factor $\tilde{\delta}$. So, the quality of the approximation improves in the sense that $\int_T (a_t')^2 \, d\tau \rightarrow 0$.[18]

## 7. CONCLUDING REMARKS

The underlying theme of this paper is that a complete description of an economy requires an explicit description of the relative weight, or measure, of subsets of assets. In a sense, the sequence model is an incomplete description of an asset economy because it does not allow for measures that appropriately reflect basic concepts that are central to the APT and factor analysis. By contrast, the model of this paper allows for measures that have a simple and natural representation, making it possible to give a new perspective on such issues as pricing, factor extraction, and sampling.

Within this framework, results for factor structures and asset pricing are derived. Some of these results represent cleaner and sharper statements of known results or widely shared intuitions, thus providing a plausibility check on the model. Other results are new with no counterpart in the sequence model, illustrating the incremental contribution of the framework with a continuum of assets. It is worth noting that while factor analysis is cast in the context of asset pricing and the APT, the concepts and results are valid in other contexts in which there is a need for a parsimonious and tractable representation of individual risks in terms of common, economy-wide risks.

The factor-pricing results reported here suggest that some of the critiques of the APT are brought about by the particular formalism of an infinite sequence of assets. This will hopefully focus the debate on more substantive conceptual issues concerning the APT's basic assumptions of absence of arbitrage opportunities, symmetric information about assets' stochastic returns, strict factor structure, ... etc. Finally, while the paper vindicates the APT's basic claim that no-arbitrage assumptions impose strong pricing restrictions, empirical evidence against the APT is also more damaging within the present framework, compared to the traditional sequence APT which makes no definite prediction about the likelihood of pricing errors in finite samples of assets.

## APPENDIX

## PROOFS

**Proof of Proposition 1:** Suppose that $\psi$ is continuous and let $\{\mathbf{w}^k\}$ be a sequence of arbitrage portfolios with corresponding positive and negative parts $\mathbf{w}^k_+$ and $\mathbf{w}^k_-$, respectively. If $var(\tilde{w}^k) \to 0$, then $\| (\tilde{w}^k_+ - \tilde{w}^k_-) - (\tilde{w}^k_+ - w^k_+) \| \to 0$. Norm continuity and the linearity $\psi$ imply that $|\psi(\tilde{w}^k_+ - w^k_+) - \psi(\tilde{w}^k - w^k_-)| \to 0$. By the definition of $\psi$, this means $|(w^k_+ - \gamma_0) - (w^k_- - \gamma_0)| = |\tilde{w}^k_+ - w^k_-| \to 0$, as required.

Conversely, Let $span_f(L^*)$ be the linear space of all *finite* linear combinations spanned by $L^*$. If $\tilde{w} \in span_f(L^*)$ is the random rate of return on a portfolio $\mathbf{w}$ with support $\{t_1, \ldots, t_n\}$ and weights $\alpha_i$, define $\psi(\tilde{w} - \tilde{w}) = \sum_i \alpha_i r_{t_i} - \gamma_0$. This definition makes sense because if $\tilde{w}$ is the rate of return on two different portfolios $\mathbf{w}$ and $\mathbf{w}'$ with supports $t_i$, $t_j$ and weights $\alpha_i$ and $\alpha_j$, but, say, $\sum_i \alpha_i r_{t_i} > \sum_j \alpha_j \bar{r}_{t_j}$, then $\mathbf{w} - \mathbf{w}'$ is an arbitrage portfolio such that $var(\tilde{w} - \tilde{w}') = 0$ yet $w - w' \neq 0$, a contradiction with the assumption that $\mathbf{r}$ admits no arbitrage opportunities. Since there is a riskless portfolio, it is also clear that $\psi(0) = 0$.

This shows that $\psi$ can be extended linearly to all of $span_f(L^*)$. Note that $span_f(L^*)$ is a norm dense linear subspace of $span(L^*)$. Since $\psi$ is continuous (hence uniformly continuous) on $span_f(L^*)$, $\psi$ has a unique continuous extension to $span(L^*)$.

*Q.E.D.*

To prove Proposition 2, I begin with simple characterizations which further clarify the structure of idiosyncratic processes. Part (i), in particular, shows that the definition of an idiosyncratic process given here is in fact equivalent to the seemingly stronger and more abstract definition in Al-Najjar [1] for general processes. Let $\mathcal{F}_K$ denote the set of all factor spaces in $span(L^*)$ of dimension $K = 0, 1, 2, \ldots$ and let $\mathcal{F}$ be the set of all finite dimensional factor spaces (*i.e.*, $\mathcal{F} = \cup_{K=1}^{\infty} \mathcal{F}_K$). It is also convenient to define the set $\mathcal{F}_\infty$ of all countably infinite dimensional closed subspaces of $span(L^*)$.

**PROPOSITION A.1:**

i) **h** *is idiosyncratic if and only if for every random variable* $\tilde{x}$, $cov(\tilde{x}, \tilde{h}_t) = 0$ *for almost every* $t$;

ii) **h** *is idiosyncratic if and only if for every* $H \in \mathcal{F} \cup \mathcal{F}_\infty$ *we have* $\tilde{h}_t \perp H$, $\tau$ – a.e.

**Proof:**

i) Define $H$ to be the closed linear spaced spanned by $\{\tilde{h}_t : t \in [0, 1]\}$. If **h** is idiosyncratic then for every $t$, $cov(\tilde{h}_t, \tilde{h}_s) = 0$ for almost every $s$. The linearity of the covariance implies that this claim is also true for any $\tilde{x} \in H$ which is a finite linear combination of elements in $\{\tilde{h}_t : t \in [0, 1]\}$. Finally, the claim holds for any $\tilde{x}$ in $H$ by continuity of the covariance function. Finally, writing the direct sum $L_2 = H \oplus H^\perp$, and noting that for any $\tilde{y} \in H^\perp$ we have $cov(\tilde{y}, \tilde{h}_s) = 0$ for every $s \in T$, we conclude that for any $\tilde{x} \in L_2$, $cov(\tilde{x}, \tilde{h}_s) = 0$ for almost every $s$.

In the other direction, suppose that for every $\tilde{x}$, $cov(\tilde{x}, \tilde{h}_t) = 0$ for almost every $t$. Then for every $t$, $cov(\tilde{h}_t, \tilde{h}_s) = Cov(t, s) = 0$, $\tau(s)$ – a.e., so $\int_T |Cov(s, t)| \, d\tau(t) = 0$. By Fubini's Theorem,

$$\int_{T \times T} |Cov(s, t)| \; d\tau^2 = \int_T \left[ \int_T |Cov(s, t)| \, d\tau \right] d\tau = 0,$$

implying that $Cov(t, s) = 0$, $\tau^2$ – a.e., so **h** is idiosyncratic.

ii) One direction follows immediately from the definition. In the other direction, suppose that **h** is idiosyncratic. If $H \in \mathcal{F}_\infty$, then the definition of $\mathcal{F}_\infty$ implies that $H$ has a countable orthonormal basis $\{\gamma_1, \gamma_2 \ldots\}$.[19] From part (i), the fact that **h** is idiosyncratic implies that for any $l$, $\tilde{h}_t \perp \gamma_l$ except for $t$'s in a subset of assets $S_l \subset T$ with $\tau(S_l) = 0$. Define $S = \cup_{l=1}^{\infty} S_l$ and note that $\tau(S) \leq \sum_{l=1}^{\infty} \tau(S_l) = 0$. For every $t \notin S$, we have $\tilde{h}_t \perp \gamma_l$ for all $l = 1, 2, \ldots$. Since $\{\gamma_1, \gamma_2 \ldots\}$ is a spanning set for $H$, we conclude that $\tilde{h}_t \perp H$ for all assets $t \notin S$. That is, for all assets outside a set of measure zero $S$ are orthogonal to $H$ as required. This proves the result in the case $H \in \mathcal{F}_\infty$. In the remaining case $H \in \mathcal{F}$, the spanning set $G$ is finite and the same argument applies with only minor modifications.

*Q.E.D.*

**Proof of Proposition 2:** Let **r** be an asset return process with strict factor space $F \in \mathcal{F} \cup \mathcal{F}_\infty$ and continuous pricing function $v$. Let $\{\tilde{z}_\alpha : \alpha \in \mathcal{A}\}$ be an orthonormal basis for $span(L^*)$ where $\mathcal{A}$ is an arbitrary index set. Since $v$ is a continuous linear functional on $span(L^*)$, there is a vector $\theta \in span(L^*)$ such that $v(\tilde{x}) = (\theta|\tilde{x})$ for every $\tilde{x} \in span(L^*)$. By Theorem IV.4.10 of Dunford and Schwartz [8], there is a countable subset $A \subset \mathcal{A}$ such that $\theta \perp \tilde{z}_\alpha$ for every $\alpha \notin A$. Define $H = span\{\tilde{z}_a : a \in A\}$, so by Proposition A.1 (ii), $\bar{h}_t \in H^\perp$, $\tau - a.e.$ Since $v(\tilde{x}) = 0$ for every $\tilde{x} \in H^\perp$ by construction, we conclude that $v(\bar{h}_t) = 0$, $\tau - a.e.$

*Q.E.D.*

A more direct proof of Proposition 2 using limits of arbitrage portfolios is also possible. The advantage of the present proof (aside from being shorter) is that it better highlights the roles played by the continuity of $v$ and the structure of Hilbert spaces. The economic reasoning enters in a subtle way in the step that the union of negligible sets is negligible in Proposition A.1 (ii).

**Proof of Proposition 3:** By Proposition 2, the set of correctly priced assets $A$ satisfies $\tau(A) = 1$. Thus,

$$
\begin{aligned}
\tau^n\{(t_1, \dots, t_n) : a_{t_i} = 0, \quad \text{for } = 1, \dots, n\} &= \tau^n(A \times \cdots \times A) \\
&= \underbrace{\tau(A) \cdots \tau(A)}_{n \text{times}} \\
&= 1,
\end{aligned}
$$

where the second equality follows from the fact that $\tau^n$ is a product measure.

*Q.E.D.*

To prove Proposition 4, I begin with three preliminary results which may be of independent interest. It is important to note that the measurability of $t \mapsto var(\tilde{r}_t)$ is made only for expository convenience. All the analysis would go through if we define $V$ to be $\int_T var(\text{Proj}_F \tilde{r}_t) d\tau$.

**PROPOSITION A.2:** *Assume that* **r** *is bounded and has a measurable covariance structure. Then for any $F \in \mathcal{F} \cup \mathcal{F}_\infty$, the function $t \mapsto var(\text{Proj}_F \tilde{r}_t)$ is bounded and measurable. In particular, if $t \mapsto var(\tilde{r}_t)$ is measurable, then $V(F)$ is well defined.*

**Proof:** First, since $\mathbf{r}$ is norm bounded by a constant $M$, $\| \text{Proj}_F \tilde{r}_t \| \leq \| \tilde{r}_t \| \leq M$ for all $t$. Second, if $\{\tilde{\epsilon}_k\}_{k=1}^{\infty}$ is any orthonormal basis for $F$, then the measurability of the covariance structure of $\mathbf{r}$ (which implies weak measurability by Proposition A.1) means that $t \mapsto cov(\tilde{\epsilon}_k, \tilde{r}_t)^2 = var(\text{Proj}_{\tilde{\epsilon}_k} \tilde{r}_t)$ is measurable for every $k$. Since the $\tilde{\epsilon}_k$'s are orthogonal, $var(\text{Proj}_F \tilde{r}_t) = \sum_{k=1}^{\infty} var(\text{Proj}_{\tilde{\epsilon}_k} \tilde{r}_t)$. The function $t \mapsto var(\text{Proj}_F \tilde{r}_t)$ is measurable since it is the pointwise limit of the sequence of measurable functions $t \mapsto \sum_{k=1}^{K} var(\text{Proj}_{\tilde{\epsilon}_k} \tilde{r}_t)$. Thus, the function $t \mapsto var(\text{Proj}_F \tilde{r}_t)$ is a bounded measurable function, so $V$ is well-defined.

*Q.E.D.*

**PROPOSITION A.3:** *Suppose that $F, F' \in \mathcal{F} \cup \mathcal{F}_\infty$ are orthogonal and let $F_\cup = span(F \cup F')$. Then $V(F_\cup) = V(F) + V(F')$.*

**Proof:** Since $F \oplus F' = F_\cup$, we have $\text{Proj}_{F_\cup} \tilde{r} = \text{Proj}_F \tilde{r} + \text{Proj}_{F'} \tilde{r}$ and

$$var(\text{Proj}_{F_\cup} \tilde{r}) = var(\text{Proj}_F \tilde{r}) + var(\text{Proj}_{F'} \tilde{r}).$$

The additivity of the integral implies:

$$\int_T var(\text{Proj}_{F_\cup} \tilde{r})\, d\tau = \int_T var(\text{Proj}_F \tilde{r})\, d\tau + \int_T var(\text{Proj}_{F'} \tilde{r})\, d\tau.$$

The result now follows by substituting in the definition of $V(F_\cup)$.

*Q.E.D.*

**PROPOSITION A.4:** *Let $\{\tilde{\epsilon}_\alpha : \alpha \in \mathcal{A}\}$ be an orthonormal basis for $span(L^*)$, where the index set $\mathcal{A}$ may be uncountable. Then there is a countable set $A \subset \mathcal{A}$ such that $V(\tilde{\epsilon}_\alpha) > 0$ if and only if $\alpha \in A$.*

**Proof:** For each $n = 1, 2, \ldots$, define $A_n = \{\alpha : V(\tilde{\epsilon}_\alpha) \geq \frac{1}{n}\}$. If $A_n$ contained infinitely many indices for some $n$, then for any $m$ we can find $m$ distinct indices $\{\alpha^1, \ldots, \alpha^m\} \subset A_n$. Using Proposition A.3, we have $V(span\{\tilde{\epsilon}_{\alpha^1}, \ldots, \tilde{\epsilon}_{\alpha^m}\}) = \sum_m V(\tilde{\epsilon}_{\alpha^m}) \geq \frac{m}{n}$. This is impossible since $V(F) \leq 1$ for all $F \in \mathcal{F}$. We conclude that $A_n$ must be finite for each $n$, hence $A = \cup_n A_n = \{\alpha : V(\tilde{\epsilon}_\alpha) > 0\}$ is countable.

*Q.E.D.*

Since the choice of the basis $\{\tilde{\delta}_\alpha : \alpha \in \mathcal{A}\}$ was arbitrary, the countable set of indices $A$ may be highly 'inefficient'. For example, even if $\mathbf{r}$ has a strict one-factor structure, the set $A$ whose existence is asserted in Proposition A.4 may be infinite. On the other hand, this proposition is useful because it reduces the problem of searching for an optimal set of factors to a countable dimensional subspace, namely the subspace spanned by $\{\tilde{\delta}_\alpha : \alpha \in A\}$.

**Proof of Proposition 4:** (i) Recall the definition $V_1^{\max} = \sup_{F \in \mathcal{F}_1} V(F) < \infty$. Let $\{\tilde{\delta}^n\}_{n=1}^{\infty}$ be a sequence of factors such that $V(\tilde{\delta}^n) \uparrow V_1^{\max}$. From Proposition A.4, we may assume, without loss of generality, that the sequence $\{\tilde{\delta}^n\}$ was chosen so that it lies in a countable dimensional subspace $L'$ of $L^*$. Since each $\tilde{\delta}^n$ has norm equal to one, these vectors belong to the unit sphere $B'$ of $L'$. By the Alaoglu Theorem [8, V.4.2, p. 424], $B'$ is compact in the weak topology on $L_2$ (which coincides with the weak* topology in $L_2$). Since $B'$ is a subset of the countable dimensional (hence separable) subspace $L'$, the weak topology on $B'$ is metrizable. With some abuse of notation, we may therefore assume that there is a random variable $\tilde{\delta}^*$, with $\| \tilde{\delta}^* \| \leq 1$ such that $\tilde{\delta}^n \to \tilde{\delta}^*$ weakly.

Defining $d^n = \tilde{\delta}^n - \tilde{\delta}^*$, we have

$$\begin{aligned}
\int_T \| \operatorname{Proj}_{\tilde{\delta}^n} \tilde{r}_t \|^2 \, dt &= \int_T (\tilde{\delta}^n \mid \tilde{r}_t)^2 \, dt \\
&= \int_T \Big[ (\tilde{\delta}^* \mid \tilde{r}_t) + (d^n \mid \tilde{r}_t) \Big]^2 \, dt \\
&= \int_T (\tilde{\delta}^* \mid \tilde{r}_t)^2 \, dt + \int_T (d^n \mid \tilde{r}_t)^2 \, dt + \int_T 2(\tilde{\delta}^* \mid \tilde{r}_t)(d^n \mid \tilde{r}_t) \, dt
\end{aligned}$$

The fact that $d^n \to 0$ weakly means that the sequence of functions $t \to (d^n \mid \tilde{r}_t)$ converges to 0 almost everywhere. This implies that the second and the third integrals converge to zero as $n$ goes to infinity. This and the assumption that $\int_T \| \operatorname{Proj}_{\tilde{\delta}^n} \tilde{r}_t \|^2 \, dt \to V_1^{\max}$ imply that $\int_T (\tilde{\delta}^* \mid \tilde{r}_t)^2 \, dt = V_1^{\max}$. On the other hand, since $V\left(\frac{1}{\| \tilde{\delta}^* \|} \tilde{\delta}^*\right) = \frac{1}{\| \tilde{\delta}^* \|^2} \int_T (\tilde{\delta}^* \mid \tilde{r}_t)^2 \, dt \geq \int_T (\tilde{\delta}^* \mid \tilde{r}_t)^2 \, dt$, it must also be the case that $\| \tilde{\delta}^* \| = 1$, hence $V(\tilde{\delta}^*) = V_1^{\max}$.

(ii) Apply part (i) to the process $\mathbf{r}$ to extract an optimal 1-factor space $\tilde{\delta}_1^*$ with corresponding risk exposure function $\beta_{1t}$. Write $\mathbf{r}_2 = \mathbf{r} - \beta_{1t}\tilde{\delta}_1^*$. The process $\mathbf{r}_2$ is clearly bounded and has a measurable covariance structure. We can therefore again apply part (i) to extract an optimal factor $\tilde{\delta}_2^*$ for $\mathbf{r}_2$.

Note that we must have $V(\tilde{\delta}_2^*) \leq V(\tilde{\delta}_1^*)$. Write $\mathbf{r}_3 = \mathbf{r}_2 - \beta_{2t}\tilde{\delta}_2^*$. Repeating the process produces the desired ordered sequence of factors $\{\tilde{\delta}_1^*, \tilde{\delta}_2^*, \ldots\}$.

(iii) is trivial.

*Q.E.D.*

The complication in the proof of part (i) arises because the Alaoglu Theorem ensures only the existence of weak limits for the sequence $\{\tilde{\delta}_1^n, \ldots, \tilde{\delta}_K^n\}$. While $V$ is continuous in the strong (norm) topology on $L^*$, it is not continuous in the weak topology so we cannot pass to the limit and conclude that $\tilde{\delta}^*$ is an optimal factor. The proof takes as candidate the weak limit $\tilde{\delta}^*$ then show 'by hand' that it is indeed optimal. This weakness of weak convergence in $L_2$ also explains the need for the restriction that $\mathbf{r}$ has $K$-strict factor structure in Proposition 5. The early part of the proof of part (i) can be extended to the sequence of $L$-factor spaces used in Proposition 5. However, the (weak) limiting factors not only might fail to have norm one, but may even be correlated.

**Proof of Proposition 5:** Let $F_K$ be the strict factor space for $\mathbf{r}$. Let $\{F^n\}_{n=1}^{\infty}$ be a sequence of $L$-factor spaces such that $V(F^n) \uparrow V_L^{\max}$. We may assume, without loss of generality, that $F^n \subset F_K$ for all $n$. For each $n$, write $F^n = span\{\tilde{\delta}_1^n, \ldots, \tilde{\delta}_L^n\}$ and note that, by proposition A.3, $V(F^n) = \sum_l V(\tilde{\delta}_l^n)$. Since each sequence $\{\tilde{\delta}_l^n\}$ is bounded and lies in the finite dimensional subspace $F_K$, there must be a $\tilde{\delta}_l$ such that $\tilde{\delta}_l^n \to \tilde{\delta}_l$ in norm. Since the inner product is jointly continuous in norm, $\{\tilde{\delta}_1, \ldots, \tilde{\delta}_L\}$ is a set of factors. Since $V$ is norm continuous (for an argument, see the proof of Proposition 7), $V_L^{\max} = V(\tilde{\delta}_1, \ldots, \tilde{\delta}_L)$.

*Q.E.D.*

To prove Proposition 6, I first show the following intermediate result which further refines the construction of Proposition A.4.

**PROPOSITION A.5:** *There is a unique minimal linear space $F_\infty \in \mathcal{F} \cup \mathcal{F}_\infty$ with $V(F) = V^{\max}$.*

**Proof:** Let $A$ be a countable set of indices as in Proposition A.4, and define $L' = span\{\tilde{\delta}_\alpha : \alpha \in A\}$. It is easy to see that $V(L') = V^{\max}$ and that for any $L \in \mathcal{F}_\infty$ with $V(L) = V^{\max}$, we also have

$V(L \cap L') = V^{max}$. Thus, without loss of generality, we may assume that any $L \in \mathcal{F}_\infty$ with $V(L) = V^{max}$ is a subspace of $L'$.

Define $H = \{\tilde{\eta} \in L' : V(\tilde{\eta}) = 0\}$. I first show that $H$ is a closed linear subspace. Suppose that $\tilde{\delta} = \sum_{n=1}^{N} \tilde{\eta}_n$, where $V(\tilde{\eta}_n) = 0$ for all $n$. This means that $(\tilde{r}_t \mid \tilde{\eta}_n) = 0$, except for $t$'s in a set $B_n \subset T$ with $\tau(B_n) = 0$. Thus, every $t$ in the set of measure zero $B = \cup B_n$ is orthogonal to each $\tilde{\eta}_n$, hence orthogonal to the subspace they span. This implies that $V(\tilde{\delta}) = 0$. That $H$ is closed follows from an argument analogous to the one used to prove Proposition 7.

To complete the proof, the equation $L' = \bar{F}_\infty \oplus H$ defines $\bar{F}_\infty$ uniquely. Since $V(H) = 0$, we must have $V(\bar{F}_\infty) = V^{max}$. It is easy to see that $\bar{F}_\infty$ must be minimal.

*Q.E.D.*

**Proof of Proposition 6:** Part (i) is immediate and part (iii) follows from Proposition A.5. To prove (ii), it is enough to show that $F_\infty = span\{\tilde{\delta}_1, \tilde{\delta}_2, \ldots\}$. Clearly, $\tilde{\delta} \in F_\infty$. Define $F_\infty^2$ by $\bar{F}_\infty = \tilde{\delta}_1 \oplus F_\infty^2$. Since $\tilde{\delta}_2 \perp \tilde{\delta}_1$, it is clear that $\tilde{\delta}_2$ must belong to $F_\infty^2$. Repeating this process establishes that $\tilde{\delta}_k \in F_\infty$ for all $k$, hence $span\{\tilde{\delta}_1, \tilde{\delta}_2, \ldots\} \subset F_\infty$. If the inclusion were proper, then, by the minimality of $F_\infty$, there is $\tilde{\eta} \in \bar{F}_\infty$ with $\tilde{\eta} \perp \tilde{\delta}_k$ for all $k$ such that $V(\tilde{\eta}) > 0$. But this would imply that $V(\tilde{\eta}) > V(\tilde{\delta}_k)$ for at least one $k$ (in fact infinitely many $k$'s), contradicting the assumption that each $\tilde{\delta}_k$ was selected optimally.

To prove part (v), recall that $V(\tilde{\eta}) = 0$ for any $\tilde{\eta} \in F_\infty^\perp$. Thus, $\int \| \text{Proj}_\eta \tilde{r}_t \|^2 \, d\tau = 0$, implying that $\text{Proj}_\eta \tilde{r}_t = 0$, $\tau - a.e.\ t$. Since $\tilde{\eta} \perp \{\tilde{\delta}_1, \tilde{\delta}_2, \ldots\}$, $\text{Proj}_\eta \tilde{r}_t = \text{Proj}_\eta [\text{Proj}_{F_\infty} \tilde{r}_t + \text{Proj}_{F_\infty^\perp} \tilde{r}_t] = \text{Proj}_\eta \text{Proj}_{F_\infty^\perp} \tilde{r}_t = \text{Proj}_\eta \tilde{h}_t$.

*Q.E.D.*

**Proof of Proposition 7:** By the additivity of $V$, we have $V(F^n) = \sum_k V(\tilde{\delta}_k^n)$ and a similar expression for $V(F)$. It is therefore sufficient to prove that $V(\tilde{\delta}_k^n) \to V(\tilde{\delta}_k)$ for each $k$.

Since $\tilde{\delta}_k^n$ converges to $\tilde{\delta}_k$ in norm, we have, $var(\text{Proj}_{\tilde{\delta}_k^n} \tilde{r}_t) = (\tilde{r}_t \mid \tilde{\delta}_k^n)^2 \to (\tilde{r}_t \mid \tilde{\delta}_k)^2 = var(\text{Proj}_{\tilde{\delta}_k} \tilde{r}_t)$, for each asset $t$. Since $\mathbf{r}$ is bounded, The Dominated Convergence Theorem implies that

$$\int_T var(\text{Proj}_{\tilde{\delta}_k^n})\tilde{r}_t \, dt \to \int_T var(\text{Proj}_{\tilde{\delta}_k})\tilde{r}_t \, dt,$$

as required.

*Q.E.D.*

**Proof of Proposition 8:** The proof of Proposition 2 already established that:

$$\bar{r}_t - \gamma_0 = \psi(\text{Proj}_{F^\perp} \tilde{r}_t).$$

The linearity of $\psi$ implies

$$\bar{r}_t - \gamma_0 = \psi(\text{Proj}_{F_K} \tilde{r}_t) + \psi(\text{Proj}_{F_K^\perp} \tilde{r}_t)$$

for any optimal $K$-factor space $\hat{F}_K$. Since $\psi$ is continuous, hence uniformly continuous, for any $\epsilon > 0$ there is $\alpha > 0$ such that $var(\tilde{x}) < \alpha$ implies $|\psi(\tilde{x})| < \epsilon$. Proposition 6 (i) implies that for any $\alpha > 0$, there is $K$ such that $\tau\{t : var(\text{Proj}_{F_K^\perp} \tilde{r}_t) > \alpha\} < \alpha$. The conclusion of the proposition follows by choosing $\alpha$ and $K$ appropriately.

*Q.E.D.*

**Proof of Proposition 9:** Fix $\epsilon > 0$ and apply Proposition 8 to obtain a $K$-factor space $F$ ensuring $|a_t| < \epsilon$ for all $t$ in a subset $A \subset T$ with $\tau(A) > 1 - \epsilon$. Since the draws are i.i.d., the proposition follows by applying the law of large number.

*Q.E.D.*

## ENDNOTES

1- In this paper I focus on the 'arbitrage-APT' model (*e.g.*, Ross [14], Chamberlain and Rothchild [5]), rather than the 'equilibrium-APT' (*e.g.*, Connor [6] and Milne [12]).

2- Pointed out by an Associate Editor.

3- Another issue, addressed in a companion paper (Al-Najjar [2]), concerns the robustness of factor structures to seemingly irrelevant repackagings of assets. A number of authors, beginning with Shanken [15, 16] and later followed by others (*e.g.*, Gilles and LeRoy [9]), argued that the factor structure in a sequence economy can be arbitrarily changed as a result of 'repackaging' assets. Using a model similar to the one presented here, Al-Najjar [2] shows that when repackaging is appropriately defined, factor structures in a continuum economy are robust in the sense that repackaging can never create new factors. See Al-Najjar [2] for a more detailed discussion of the literature on this problem.

4- This rate is obtained by dividing the gross return on that asset, which is a random variable denoted $\tilde{f}_t$, by the price of the asset (assuming this price is not zero). Thus, a theory of rates of returns is also implicitly a theory of asset prices.

5- The existence of a non-trivial idiosyncratic component imposes certain restrictions on the underlying probability space $(\Omega, \Sigma, P)$. For example, this space cannot be generated by the $\sigma$-algebra of a complete separable metric space. It is worth noting that the existence of such large probability spaces is guaranteed by standard constructions using the Kolmogorov Extension Theorem which applies to arbitrary index sets (see, *e.g.*, , Ash [4], Theorem 4.4.3, p. 191). For example, it is straightforward to construct a probability space on which a continuum of i.i.d. random variables are defined. See Al-Najjar (1995, [1], p. 1199) for further discussion.

6- The condition that $Cov$ is Lebesgue measurable means that it is enough that $Cov$ be measurable relative to the $\tau^2$-completion of the measurable space $(T^2, \mathcal{T}^2)$. This is a weaker condition than the condition of being measurable relative to the product space $(T^2, \mathcal{T}^2)$. Recall that the $\tau^2$-completion of $(T^2, \mathcal{T}^2)$ is obtained by adding all subsets of sets of $\tau^2$-measure zero.

7- A portfolio may be thought of as a signed measure with finite support. General signed measures can be introduced as idealized portfolios, as suggested in Al-Najjar [1]. Such a generalization would have little impact on the results of this paper.

8- Formally, given any $\epsilon > 0$, there can be at most a finite number of assets for which $|a_n^2| > \epsilon$.

9- I thank Max Stinchcombe for pointing out these facts.

10- Note that while we maintain the strong assumptions that the observer knows the primitives of the economies (*e.g.*, the betas and the factor risk premia) to better illustrate the main point about testing, the qualitative result that all assets in a random sample are correctly priced holds for *any* specification of the primitives (provided the measurability and absence of arbitrage assumptions are met).

11- The measurability of the covariance structure does not require the variance function $Var(t) = Cov(t, t)$ to be measurable because the diagonal in $T \times T$ has measure zero, so any of its subsets is measurable by the completeness of the Lebesgue measure. For example, let $\mathbf{h}$ be an i.i.d. process with unit variance and define the process $\tilde{r}_t = \tilde{h}_t$ for $t \in A$ and 0 off $A$. Then $Cov(t, s)$ is identically zero except on the set $A' = \{(t, t) : t \in A\}$. However, $A'$ is measurable, being a subset of the diagonal which has measure zero. This is so even when $A$ is non-measurable, in which case $Var(t)$, being the indicator function of $A$, will not be a measurable function.

12- It seems reasonable to expect that a stronger version of this result holds in which the assumption that $\mathbf{r}$ has a $K$-strict factor structure is eliminated. I haven't been able to prove this stronger version for the reasons explained in the remark following the proof of Proposition 4.

13- For example, $K$ factors with betas uniformly bounded away from zero.

14- The decomposition in the present paper and in [1] are linear, in the sense that: (1) risk is written as the *sum* of idiosyncratic and factor risks; and (2) the residuals are mutually orthogonal (uncorrelated). Since the absence of correlation does not imply independence, the residuals in a linear decomposition of an asset's return may still contain infromation that can help predicting the returns of other assets. A stronger form of decomposition is provided in Al-Najjar (1996, [3]).

There, random *aggregate states* are extracted with the property that, conditional on knowledge of the realized aggregate state, individual shocks are indpendent.

15- One could have equivalently required $(\tilde{\delta}_k | \tilde{\delta}'_k) \leq -1 + \epsilon$ since it is the space spanned by the sets of factors which matter in the analysis. The present definition simplifies the exposition.

16- More formally, let $\{\lambda^n_K\}$ and $\{\lambda^n_{K+1}\}$ be the sequences of the $K$th and $(K+1)$st largest (in absolute value) eigenvalues of $\Sigma_N$. Then, the sequence economy has an approximate $K$-factor structure if and only if $\limsup |\lambda^n_{K+1}| = \infty$ while $\limsup |\lambda^n_{K+1}| < \infty$.

17- This assumption simplifies the exposition, but can be dispensed with easily.

18- This follows from the Lebesgue Dominated Convergence Theorem, the continuity of $v$, and the continuity of Proj, which implies that $a'_t \to a_t$, $\tau$ − a.e.

19- An orthonormal set $G = \{\gamma_1, \gamma_2 \ldots\} \subset H$ is a basis for a Hilbert space $H$ if $H$ is the norm-closure of the linear space generated by $G$. That is, every $h \in H$ is either a linear combination of elements of $G$, or the norm-limit of a sequence of such linear combinations. The dimension of $H$ is the cardinality of any orthonormal basis for $H$. Theorem IV.4.14 in Dunford and Schwartz (1958, p. 253) guarantees that this notion of dimension is well-defined.

## REFERENCES


1. Al-Najjar, N. I., Decomposition and Characterization of Risk with a Continuum of Random Variables, *Econometrica* **63** (1995), 1195-1224.

2. Al-Najjar, N. I., On The Robustness of Factor Structures to Asset Repackaging, (1995). Forthcoming in: *Journal of Mathematical Economics.*

3. Al-Najjar, N.I., "Aggregation and the Law of Large Numbers in Economies with a Continuum of Agents," MEDS Department, Kellogg GSM, CMSEMS working paper no. 1160, Northwestern University, March 1996.

4. Ash, R. B., "Real Analysis and Probability," Academic Press, New York, 1972.

5. Chamberlain, G. and M. Rothschild, Arbitrage, Factor Structure, and Mean-Variance Analysis on Large Asset Markets, *Econometrica* **51** (1983), 1281-1304.

6. Connor, G., A Unified Beta Pricing Theory, *J. Econ. Theory* **34** (1984), 13-31.

7. Connor, G. and R. A. Korajczyk (1992): "The Arbitrage Pricing Theory and Multifactor Models of Asset Returns". Forthcoming in: *Finance Handbook*, ed. by R. Jarraw, V. Maksimovic and W. Ziemba.

8. Dunford, N. and J.T. Schwartz, "Linear Operators, Part I," Interscience, New York, 1958.

9. Gilles, C. and S. F. LeRoy, On the Arbitrage Pricing Theory, *Econ. Theory* **1** (1991), 213-29.

10. Ingersoll Jr., J. E., "Theory of Financial Decision Making," Rowman & Littlefield, New Jersey, 1987.

11. Khan, M. A. and Y. Sun, "Hyperfinite Asset Pricing Theory," Johns Hopkins University, July 1995.

12. Milne, F., Arbitrage and Diversification in a General Equilibrium Asset Economy, *Econometrica* **56** (1988), 815-40.

13. Reisman, H., Reference Variables, Factor Structure, and the Approximate Multibeta Representations, *J. of Finance* **47** (1992), 1303-14.

14. Ross, S., The Arbitrage Theory of Capital Asset Pricing, *J. Econ. Theory* **13** (1976), 341-60.

15. Shanken, J., The Arbitrage Pricing Theory: Is it Testable?, *J. of Finance* **37** (1982), 1129-40.

16. Shanken, J., Multi-beta CAPM or Equilibrium-APT, *J. of Finance* **40** (1985), 1189-96.

17. Shanken, J., The Current State of the Arbitrage Pricing Theory, *J. Finance* **47** (1992), 1569-74.

18. Werner, J., Diversification and Equilibrium in Securities Markets, *Journal of Economic Theory* **75** (1997), 89-103.