Yes, The APT is Testable
Author(s): Philip H. Dybvig and Stephen A. Ross
Source: *The Journal of Finance*, Vol. 40, No. 4 (Sep., 1985), pp. 1173-1188
Published by: Blackwell Publishing for the American Finance Association
Stable URL: http://www.jstor.org/stable/2328401
Accessed: 26/11/2009 08:46

Yes, The APT Is Testable

PHILIP H. DYBVIG and STEPHEN A. ROSS*

ABSTRACT

The Arbitrage Pricing Theory (APT) has been proposed as an alternative to the mean-variance Capital Asset Pricing Model (CAPM). This paper considers the testability of the APT and points out the irrelevance for testing of the approximation error. We refute Shanken's objections, including his assertion that Roll's critique of the CAPM is applicable to the APT. We also explain the testability of the APT on subsets, and we explore the relationship between the APT and the CAPM.

THE ARBITRAGE PRICING THEORY (APT) has been the subject of extensive research efforts since it was introduced by Ross [36, 37]. The research has included theoretical papers[1] and a growing empirical literature.[2] Nonetheless, there seems to be some confusion in the literature about the relation between the theoretical underpinnings of the APT and its empirical tests. This paper is an attempt to alleviate the confusion. To illustrate the issues and to provide a focus, Section I of this paper is devoted to examining and refuting some points raised in Shanken [40] which considers the testability of the APT.

The underlying premise of the APT is that asset returns, $\tilde{R}_i$, are generated by a factor model of the form

$$\tilde{R}_i = E_i + \textstyle\sum_{j=1}^{k} b_{ij}\tilde{\delta}_j + \tilde{\varepsilon}_i, \tag{1}$$

where $\tilde{\delta}$ is a vector of systematic factors which commonly influence assets returns,

* Both authors are from the Yale School of Management, Yale University. The authors are grateful for valuable comments from Steve Brown, Jon Ingersoll, Dick Roll, Jay Shanken, and Mark Weinstein. We would like to recommend related papers by Pfleiderer [28] and Pfleiderer and Reiss [29]. Dybvig is grateful for support under the Batterymarch fellowship program, and Ross is grateful for support from the National Science Foundation.

[1] On the theoretical front, several separate but related strands can be identified. Chamberlain and Rothschild [5] Huberman [22], Ingersoll [24], and Stambaugh [42] have derived the APT using a definition of no arbitrage in an economy with infinitely many assets. Connor [11], Dybvig [17], and Grinblatt and Titman [21] derive the APT in an equilibrium setting and have refined the degree of approximation. Chen and Ingersoll [8], Cragg and Malkiel [14], and Dybvig [16] have derived the APT by assuming that some agent holds a well-diversified portfolio. Brock [1, 2], Constantinides [12], Cox, Ingersoll, and Ross [13], Garman and Ohlsen [18], Prescott [30], Roll and Ross [35], and Schmalensee [39] have examined the intertemporal consistency of the theory in dynamic equilibrium models. Shanken [41] futher examines equilibrium derivations of the APT.

[2] The APT has been examined by Brown and Weinstein [3], Chen [7], Chan, Chen, and Hsieh [6], Chen, Roll, and Ross [9], Dhrymes, Friend, and Gultenkin [15], Cho, Elton, and Gruber [10], Gehr [19], Gibbons [20], Hughes [23], Jobson [25], Oldfield and Rogalski [27], Reinganum [31], and Roll and Ross [35]. (Without judging its other merits, e.g., relative power, the Brown and Weinstein test conclusively demonstrates that the APT is testable, since their test is a maximum likelihood test.)

β_i is the vector of factor loadings on asset i, E_i is the expected return on asset i, and ε_i is the idiosyncratic noise term associated with asset i. Although weaker assumptions suffice, for concreteness we can assume that $\tilde{\varepsilon}_i$ is independent across assets, is independent of the factors, and has a mean of zero. We will also assume that the factors, $\tilde{\delta}_j$, have mean zero and that variances exist.[3] The APT says simply that the new assets' expected returns are linear in the factor loadings:

$$E_i = r + \textstyle\sum_{j=1}^{k} \lambda_j \beta_{ij}, \tag{2}$$

where r is a constant (which equals the riskless rate of return if there is a riskless asset), and λ is a vector of risk premia. The "Arbitrage" in the "Arbitrage Pricing Theory" comes from the simple proof that (1) implies (2), using absence of arbitrage alone, for the case in which all the $\tilde{\varepsilon}_i$'s are identically zero.[4]

More generally, deriving the APT involves proving that the $\tilde{\varepsilon}_i$'s can be neglected when pricing assets. Of the many existing derivations of the APT, most are based on the intuition that the idiosyncratic error terms represent diversifiable risk and therefore should not affect pricing. Because diversification is a notion that holds only approximately in a finite economy, theoretical motivations of the APT have generally dealt in approximations. The theoretical derivations of the APT differ both in the sense of approximation and in the additional assumptions that are made.[5] Some theoretical models conclude that the APT is a good approximation in a sequence economy, when there are "sufficiently many" assets (Chamberlain [4], Chamberlain and Rothschild [5], Huberman [22], Ingersoll [24], and Ross [37]) or in a finer sense (Connor [11]).[6] Other models show that the APT should be a good approximation in a finite economy (Cragg and Malkiel [14], Dybvig [17], Grinblatt and Titman [21], and in continuous time, Ross [38]), with

[3] Chamberlain [4], Chamberlain and Rothschild [5], Connor [11], Ingersoll [24], and Stambaugh [42], have significantly weakened the requirement of independent idiosyncratic error terms. One message that emerges from these papers is that a factor model is essentially one in which the covariance matrix has k large eigenvalues. Another message is that what is important is that the factors be separately identifiable from asset returns, i.e. that each factor be approximated closely by some linear combination of asset returns. In Connor's model, this assumption is combined with an assumption that no asset's idiosyncratic noise plays a significant role in aggregated wealth.

[4] Suppose that all the $\tilde{\varepsilon}_i$'s are identically zero. If the linearity condition in (2) did not hold, there would be two portfolios with the same factor loadings but different expected returns. Consequently, going long in the portfolio with higher expected return, financed by shorting the other portfolio, would be an arbitrage opportunity.

[5] This has led some researchers to conclude that empirical investigations of the APT should account explicitly for the error term arising from one or another theoretical model. However, including the error term in an empirical test defeats the whole purpose of the theoretical models, which are intended to demonstrate that the error term is negligible. Furthermore, including the error term would require a new empirical test for each such version of the model. Fortunately, the notion of including the error term in the test appears in methodological discussions but not in the actual empirical tests.

[6] Specifically, the first papers show that for whatever factor structure we extract, then $(E - re - \lambda\beta)' \Sigma^{-1}(E - re - \lambda\beta)$ is uniformly bounded as more assets are added, where Σ is the covariance matrix for the residual "idiosyncratic" noise and e is a vector of ones. Note that Σ could even be the original covariance matrix. Connor shows that the sum across assets of squared deviations, $(\mu - re - \lambda\beta)' (\mu - re - \lambda\beta)$, goes to zero as assets are added. Connor also admits a general residual idiosyncratic noise matrix, but makes assumptions to ensure that the APT is valid. (One of Connor's assumptions is that $\|(\beta'\Sigma^{-1}\beta)^{-1}\| \rightarrow 0$ as the number of assets increases.)

differing approximations.[7] Some models even have assumptions sufficiently strong to ensure that the APT will hold exactly (Chen and Ingersoll [8], Dybvig [16], and implicitly Cragg and Malkiel [14]). The important point about all these papers is that more assumptions are required to move from (1) to (2) than we have made so far.

Besides more technical assumptions, each paper makes some subset of the following assumptions:

1. Each asset has small idiosyncratic variance, i.e., var($\tilde{\varepsilon}_i$) is small.
2. Each asset has small supply in the economy (at least in the limit).
3. There is a portfolio which, up to a constant, mimics factor j (at least approximately, perhaps in the limit).
4. Some agent holds a well-diversified portfolio that does not contain any idiosyncratic risk.
5. There is no arbitrage (directly or in some asymptotic sense).
6. There is Pareto efficiency and/or aggregation.
7. There are many assets.
8. All assets are in positive supply.

This is not an exhaustive list of the important economic assumptions made in these models. The point we are making is that in order to get (2) as a conclusion, it is necessary to make additional nontrivial assumptions besides (1).

We cannot expect the APT to hold unless sufficiently many of the above assumptions are good approximations. Since these assumptions are reasonable assumptions to make about fundamental traded assets in our economy (e.g., stocks and bonds), the APT can be expected to hold for these assets. However, these are not good assumptions to make about *arbitrary* portfolios of assets. Consequently, it is an error to apply the APT to arbitrary portfolios.

By way of illustration, Connor [11], Ross [38], and most usefully, Dybvig [17] and Grinblatt and Titman [21], have all derived specific expressions for the magnitude of the pricing error. Ross [38] displayed a simple expression for the error in an intertemporal diffusion model. Grinblatt and Titman [21] derive essentially the same bound based on multivariate normality, and Dybvig [17] has found a very similar formula for the error bound assuming efficiency. Generally, the pricing error for any asset i is bounded by a measure of representative relative risk aversion, R, multiplied by the asset's idiosyncratic variance, σ^2, and the proportion of total wealth represented by the asset, α_i. That is, the pricing error for asset i is bounded by $R_i\sigma^2\alpha_i$. It might be thought that since this bound depends on the market proportion of the asset, that it suffers from the same problem of the unobservability of the market portfolio that plagues the Capital

[7] Cragg and Malkiel [14] give an intuitive argument why the deviations from APT pricing should be small. Dybvig [17] gives a bound in terms of the harmonic mean of individual absolute risk aversion, per capita asset supplies, the riskless discount factor, and the variance of idiosyncratic noise. Grinblatt and Titman [21] have two similar bounds. One bound is in terms of a uniform bound across agents and wealth levels of $u''/E(u')$, and the other bound is based on multivariate normality. Connor [11] has utilized similar bounds on preferences to those of Grinblatt and Titman [21] to derive equilibrium pricing results in sequence economies. Ross' [38] continuous time bound is based on the continuous time intertemporal asset pricing model.

Asset Pricing Model (CAPM). (See Roll [32] for this view of the CAPM, and Shanken [40] for an argument that the view is valid as well for the APT.) However, it is incorrect to make this argument for the APT. First, the proportion of asset i in all assets is always smaller than the proportion of asset i in observed assets (e.g., the proportion of a stock's value in NYSE stocks). This means that the pricing error will be even smaller than that indicated by the bound, i.e., the bound is conservative and potentially observable. Second, the assertion that the proportion of asset i is small is an *assumption* of the APT, and is therefore not an appropriate object for an empirical test. This is analogous to the treatment of normality or quadratic utility in the CAPM. Observed skewness of security returns is not cause for rejecting the CAPM, since normality is not part of the *conclusion* of the CAPM that expected returns are linear in beta coefficients.

These formulas provide us with a quick estimate of how large a pricing error we can expect. Dybvig's back of the envelope calculations, based on stock returns alone, produce an error bound of 0.04% or 4 basis points (assuming risk aversion of 5, market value proportion of 0.2%, and standard deviation of 20%). This error is small relative to the measurement error in expected returns, and the bound is valid even though not all wealth is in stocks. If, for example, the wealth in stocks is only 10% of aggregate wealth, a valid tighter bound would be 10% of 0.04 or 0.004%. In any case, this calculation provides a simple answer to the question, "How accurate should I expect the approximation to be?"

To put the matter somewhat differently, the APT is based on the view that assets are close substitutes. Consequently, slight deviations in pricing lead to large deviations in demand. For example, relying again on Dybvig [17], a 1% increase in the deviation from APT pricing will lead to an increased demand for an asset i on the order of $1\%/\sigma_i^2 R$, or 5% of world wealth (again choosing $R = 5$ and $\sigma_i = 0.2$). This is not exact arbitrage, but it is clearly akin to arbitrage.

Part of the confusion in methodological discussions of testing has been the tendency to interpret the APT as including the approximation error, or similarly, to interpret the APT as exact, rejecting inexact derivations. (Fortunately, this confusion has not invaded the actual testing.) By contrast, in the physical sciences it is customary to go through a series of calculations and approximations and test the final result as an equality, with the understanding that the model would be applied only when the approximations can reasonably be expected to make sense. In the social sciences, as in the physical sciences, the question of testability of a theory is independent of the question of the accuracy of the approximations used in its derivation.

I. Shanken's Critique

Shanken [40] has taken issue with a view that he characterizes as the "empirical version of the APT."[8] The "empirical APT" differs from the usual APT in that

[8] See p. 1132. In what follows, we will use Shanken's term as he used it. His definition (5) refers specifically to "assets," but he refers to portfolios as "transformed assets" and treats them as assets in applying his "empirical APT." This is a misleading feature of Shanken's paper, since his definition (5) would be reasonable if it referred only to actual assets, but is not reasonable as he used it.

the assertion that (1) implies (2) is held to be valid, independent of other assumptions, not only for assets, but also for arbitrary portfolios.[9] Of course, empirical tests are performed using actual assets (e.g., stocks), not on arbitrary recombinations, i.e., the "empirical APT" is unrelated to existing empirical work and is, therefore, something of a straw man.

Shanken bases his analysis on a two-step procedure involving a transformation of the original assets. In the first step, the original covariance matrix[10] of returns, V, is reduced to a diagonal matrix, and the "empirical version of the APT" is applied to "prove" that the transformed assets all have the same expected return. In the second step, the transformation is inverted to "prove" that the original securities have the same expected return.

Let's begin our discussion with a close look at Shanken's two-asset example, warning the reader that, as we will show, the two-asset example is degenerate. As Shanken says, the APT "exploits the concept of a 'large' (many assets) security market," a concept which cannot be captured with only two assets. In the example,

$$\tilde{R}_1 = E_1 + \tilde{\delta} + \tilde{\varepsilon}_1,$$

and

$$\tilde{R}_2 = E_2 - \tilde{\delta} + \tilde{\varepsilon}_2, \tag{3}$$

where R_i is the return on asset i, E_i is the expected return on asset i, δ is a common factor, and ε_i is asset i's idiosyncratic noise term. It is assumed that ε_i has variance σ^2, and that δ has variance 1.

Shanken wishes to find two portfolios that span the original assets (and therefore represent the same feasible investment opportunities) but have a degenerate factor structure. The first of his two portfolios is the original first asset. The second is composed of α shares of the first asset and $1 - \alpha$ shares of the second asset, where $\alpha \equiv 1/(1 + \sigma^2)$. Therefore, indicating the expected returns of the portfolios (or transformed assets) using *'s we have that

$$\tilde{R}_1^* = \tilde{R}_1,$$

and

$$\tilde{R}_2^* = \alpha\tilde{R}_1 + (1 - \alpha)\tilde{R}_2. \tag{4}$$

The choice of α guarantees that R_1^* and R_2^* are uncorrelated, i.e., that R_1^* and R_2^* have a zero-factor structure. By applying the "empirical APT," Shanken

[9] Many people we have talked with have thought that the APT "should" be robust to linear transformations or else its validity would be affected by mergers or spin-offs. However, mergers and spin-offs are not arbitrary transformations; they represent the adding together of returns of two stocks or the separation of a single security into two parts. Neither case is likely to affect the validity of the APT. This is in contrast to the type of extreme transformation required for Shanken's argument, which might create a new security which is long 1000 shares of security one and short one share each of 999 other securities. There is no reason to require or expect the APT's distributional assumptions to be robust to such transforms.

[10] We will work directly with the case $k = 0$ described by Shanken on p. 1140 to show a paradox, so that the residual covariance matrix, Σ, is the covariance matrix, V.

concludes that the transformed expected returns are equal, i.e., that $E_1^* = E_2^*$. But, the transformed expected returns can be equal only if the unconstrained expected returns are equal, i.e., if $E_1 = E_2$. In this case, and in any two-asset case for which the ε's have positive variance, the returns must be equal.

As Shanken pointed out, this is a contradiction. It is in the interpretation of this contradiction that disagreement sets in. Shanken argues that the contradiction is due to a logical error intrinsic to the APT as it is usually formulated for empirical purposes. Shanken concludes from the example that empirical tests should take into account the assumptions used in deriving the theory and the resultant approximation error. In fact, the contradiction is due to the error of applying the APT to transformed securities. In Shanken's paper, this error is hidden in the definition of the "empirical APT." Saying that pricing is explained by the factor structure in assets is much different than saying that pricing is explained by the factor structure in arbitrary portfolios (which Shanken calls "transformed assets"), especially when the portfolios have been chosen to destroy the original factor structure.

The two-asset example is a special case of Shanken's general procedure for transforming asset returns to obtain a logical contradiction. Our analysis of the general case is contained in the Appendix. The conclusion of our analysis is that in situations for which the APT makes sense (when there are many assets, all factors are significant, and idiosyncratic variances are not too large), either Shanken's transformation grotesquely magnifies the variances of the transformed assets or its inverse blows up, thus making the procedure sensitive to even very small errors in pricing.[11]

It is useful to consider a simple example of what happens to Shanken's transformation when there are many assets. Suppose that there are 1000 assets satisfying the one-factor model

$$\tilde{R}_i = E_i + \beta_i\tilde{\delta} + \tilde{\varepsilon}_i, \tag{5}$$

where we take δ and all of the ε_i's to be mutually independent and each to have variance σ^2. For simplicity, we take all the β_i's to be 1. For example, with stock price data we might think of $\sigma^2 \approx 0.04$ per year, corresponding to an annual standard deviation of about 20% for the common factor and about 28% for individual stocks. (The assumption of equal betas makes the example somewhat degenerate, but it is easy to show that the transformation is algebraically messier but not significantly different if, e.g., half the stocks had betas of 0.5 and the other half had betas of 1.5.[12])

[11] In the degenerate case where the original assets satisfy a zero-factor structure, the transformation is well-behaved, but there is no contradiction since the original assets *are* priced correctly by equating expected returns.

[12] Assume that returns are generated by a one-factor structure $V = \sigma^2(I+\beta\beta')$. Then the matrix Q in Shanken's procedure is proportional to $I - c\beta\beta'$, where $c = (1 + (1/\sqrt{1 + \beta'\beta}))/\beta'\beta$. Using this formula, it is easy (but tedious) to go through the example in the text for which different assets have different betas. More generally, if β is a matrix with orthogonal columns (corresponding to choosing the factors uncorrelated), then Q is proportional to $I - \beta c\beta'$, where c is now a diagonal matrix having diagonal entries that look like c above.

As is generally the case for Shanken's transform, the transform is not uniquely identified in this case. A natural choice is the following.[13]

$$R_i^* = \alpha_1 R_i + \textstyle\sum_{j\neq i} \alpha_2 R_j, \tag{6}$$

where $\alpha_2 \equiv 1/(\sqrt{1001} - 1) \approx 0.03$ and $\alpha_1 \equiv -\sqrt{1001} + \alpha_2 \approx -31.6$. It is easy to verify that the transformed assets are uncorrelated. Clearly, the effect of the transformation is to emphasize the idiosyncratic error terms (through leverage) and therefore deemphasize the factor. Each transformed asset has variance 1001 times σ^2, which is about 500 times the variance of the original assets. This violates the assumption of the APT that idiosyncratic variances should not be too large. If, as discussed above, the original assets have annual standard deviations of 20%, the transformed assets would have annual standard deviations of about 630%.

Our example with 1000 assets seems qualitatively different than Shanken's two-asset example. In the two-asset example, the transformation involves leaving one asset unchanged and replacing the other asset with a convex combination of the two assets, with little change in the overall variability of either asset. The 1000 asset example replaces each asset with a big short position in itself, investing the proceeds in 999 small long positions in the other assets, resulting in a 500-fold increase in variance. Why such a difference? The answer is that the two-asset example is degenerate.

First of all, as mentioned before, the transformation behaves much differently on two assets than on many assets. If there were two assets with security returns described by (3), then the transformation would still be given by (4), but with modest values $\alpha_2 \equiv 1/(\sqrt{3} - 1) \approx 1.37$ and $\alpha_1 \equiv -\sqrt{3} + \alpha_2 \approx -0.37$. Second, it is odd that the two assets have opposite factor loadings, since we would expect that in a one-factor world, virtually all assets would have positive loadings on the factor (just as virtually all stocks have positive CAPM betas). In any example with positive factor loadings, the transformation will require short sales (otherwise, all sources of correlation would contribute positively). Furthermore, as is shown in the Appendix, in a one-factor world, all but two of the assets will require short sales, regardless of the loadings. Finally, the factor structure in Shanken's example is ill defined.[14] In general, the factor structure is well defined if the number of assets exceeds twice the number of factors.

The Appendix verifies that either Shanken's transformed variances or the inverse of the transformation blows up. As we have seen in the above example,

[13] This choice is natural because the transformation inherits the eigenvectors of the original V matrix, i.e., it uses a canonical choice of $\sqrt{V}$.

[14] In Shanken's two-asset example,

$$V = \begin{bmatrix} 1+\sigma^2 & -1 \\ -1 & 1+\sigma^2 \end{bmatrix},$$

which is consistent with the one-factor structure with loadings

$$\beta = \begin{bmatrix} x \\ -x^{-1} \end{bmatrix}$$

if the variances blow up, then the assumptions used to motivate the APT (such as small idiosyncratic variance) will not be preserved when moving from the original assets to the transformed assets.

Another simple example illustrates the pathology of having the transform become singular, i.e., having its inverse explode. If the inverse blows up, even small errors in the pricing of the transformed assets can correspond to large errors in pricing the original assets. In other words, when the inverse blows up, it is consistent to have E_i^*'s all approximately the same without having the E_i's all approximately the same. As a simple example with this feature, suppose that $R_i^* \equiv \alpha R_i + (1 - \alpha)R_1$, where α is very close to zero. In this case,

$$\begin{aligned} E_i^* &= \alpha E_i + (1 - \alpha)E_1 \\ &= E_1 + \alpha(E_i - E_1) \\ &\approx E_1 \end{aligned}$$

for all i, even if the untransformed expected returns are not nearly equal. In other words, even though the E_i^*'s are all approximately the same, the E_i's can all be different, and in particular, the original assets can all satisfy the APT.

We have seen that it is an error to apply the APT to arbitrary portfolios of assets. To reconcile this observation with Shanken's apparently reasonable two-asset example, we have contrasted the two-state example with a more realistic example with many assets. The comparison tells us that the two-asset example makes the transformation seem reasonable only because it is a degenerate case. In the more realistic example, the transformation involves big short positions and large idiosyncratic variances after transformation. Furthermore, results in the Appendix tell us that whenever the assumptions underlying the APT make sense as applied to the transformed assets, the transformation is ill behaved and the original assets can also satisfy the APT. Together with the remarks in the introduction, these observations form the core of our response to Shanken. For the rest of the paper, we will touch on Shanken's analysis only briefly and occasionally. Instead, we will work on clarifying the relationship between the CAPM and the APT (Section II) and the sense in which the APT is testable on subsets, but the CAPM is not (Section III).

II. The APT and the CAPM

Since we have been examining the testability of the APT, it is natural to contrast the APT with the CAPM. The first question we want to ask is what it means for

---

and idiosyncratic variances

$$D = \begin{bmatrix} 1 + \sigma^2 - x^2 & 0 \\ 0 & 1 + \sigma^2 - x^{-2} \end{bmatrix},$$

i.e.,

$$V = D + \beta\beta',$$

for every x such that $1 + \sigma^2 > x^2 > 1 + \sigma^2$.

the CAPM to be true when asset returns are generated by the factor structure (1). The CAPM states that assets' expected returns are explained in cross-section by covariances with the market return, $\tilde{R}_m$:

$$E_i = r + \lambda \operatorname{cov}(\tilde{R}_i, \tilde{R}_m). \tag{7}$$

If we let α_m be the market portfolio, the factor structure of returns (1) yields,

$$\begin{aligned} \tilde{R}_m &= \textstyle\sum_{i=1}^{n} \alpha_{mi}\tilde{R}_i \\ &= \textstyle\sum_{i=1}^{n} \alpha_{mi}E_i + \sum_{i=1}^{n} \alpha_{mi} \sum_{j=1}^{k} \beta_{ij}\tilde{\delta}_j + \sum_{i=1}^{k} \alpha_{mi}\tilde{\varepsilon}_i, \\ &= E_m + \textstyle\sum_{j=1}^{k} \beta_{mj}\tilde{\delta}_j + \sum_{i=1}^{n} \alpha_{mi}\tilde{\varepsilon}_i, \end{aligned} \tag{8}$$

where

$$E_m \equiv \textstyle\sum_{i=1}^{n} \alpha_{mi}E_i,$$

and

$$\beta_{mj} \equiv \textstyle\sum_{i=1}^{n} \alpha_{mi}\beta_{ij}.$$

Consequently, the CAPM relation (7) can be written as

$$\begin{aligned} E_i &= r + \lambda[\textstyle\sum_{j=1}^{k} \beta_{ij}\beta_{mj} + \alpha_{mi}\operatorname{var}(\tilde{\varepsilon}_i)] \\ &\approx r + \textstyle\sum_{j=1}^{k} \beta_{ij}[\lambda\beta_{mj}], \end{aligned} \tag{9}$$

where we have assumed without loss of generality that the factors' covariance matrix is I, and where the approximation is good provided $\operatorname{var}(\tilde{\varepsilon}_i)$ is not too large and α_{mi} is small. Consequently, the CAPM implies the APT, provided that assets satisfy the factor structure. (What we have just done is a simple version of the computation of exact deviations from the APT under conditions that imply the CAPM.)

While the CAPM implies the APT in the presence of a factor structure (provided that relative asset supplies and idiosyncratic variances are small), it should be clear from (9) that the APT is more general than the CAPM in this context. To be specific, the factor prices in the APT do not have to bear any special relation to market factor loadings beyond perhaps having the same sign.[15] Within a k-factor model, the CAPM requires factor prices to be proportional to the loadings of the market portfolio, $\lambda_j = \lambda\beta_{mj}$, while the APT does not. (With a single factor, however, if the APT holds then the market portfolio will be a proxy for the single factor and betas with respect to the market and the factor will be approximately the same. In other words the "market model" and the one-factor APT are empirically equivalent.)

Nonetheless, it is critical to understand that the CAPM does not imply any restriction to a factor structure of the sort indicated by (1), neither one-factor nor k-factor, for $k < n/2$. Therefore, unlike the CAPM, the APT imposes a dimensionality restriction on the risky part of asset returns.[16]

In this section, we have shown that a factor structure plus the CAPM, small

[15] To see why the factor prices might be expected to have the same sign as their loadings in the market even if the CAPM does not hold, see Dybvig [17].

[16] The dimensionality restriction is somewhat blurred in the derivations with correlated idiosyncratic noise terms, since in these models the factor structure itself is approximate.

idiosyncratic noise, and small asset supplies together imply the APT. In fact, a similar argument will show that any of the familiar equilibrium theories can be used together with the factor structure and a marginal utility of income uncorrelated with idiosyncratic noise to imply the APT. For example, suppose that excess returns are proportional to covariances with a consumption index, $\tilde{C}$. Then,

$$\begin{aligned} E_i - r &= \lambda \operatorname{cov}(\tilde{R}_i, \tilde{C}) \\ &= \lambda \textstyle\sum_{j=1}^{k} \beta_{ij} \operatorname{cov}(\tilde{\delta}_j, \tilde{C}) + \lambda \operatorname{cov}(\tilde{C}, \tilde{\varepsilon}_i). \end{aligned}$$

Assuming that consumption is uncorrelated with the idiosyncratic noise terms says that the last term is zero, and this verifies the APT.

At this point, we return to a related minor issue with Shanken's analysis. Shanken observes that some of the factors might be highly correlated with a mean-variance efficient portfolio which would, he argues, lend spurious support to the APT. In fact, such support would not be spurious, since, under these conditions, the APT is true. The proof is the argument given above, with α_m replaced by the highly-correlated mean-variance efficient portfolio.

III. Testability of Subsets

"The APT is testable on subsets of assets." This statement has been widely noted as the reason why the APT is testable and does not suffer from the shortcomings of the CAPM which were pointed out by Roll [32, 33, 34]. Since this paper is devoted to resolving potential confusion about issues pertaining to the testability of the APT, we include this section on what it means for an asset pricing theory to be testable on subsets, with a brief characterization of the sense in which the APT is testable on subsets.

Before moving to the APT, it is useful to review what it means to say that the CAPM is not testable on subsets. For the CAPM, knowing the joint distribution of asset returns and the market portfolio weights would comprise essentially complete information, and would allow a test of the CAPM relation (7). Since means and covariances are identifiable from time series, (7) can be tested if we know a sequence of returns for all the assets and the composition of the market portfolio. (Throughout, we will assume that returns are i.i.d. over time or that they change in a way that does not affect identification.) Alternatively, the CAPM is testable if we have a time series of returns for all or a subset of the assets and a time series for the actual market portfolio. However, the CAPM is not testable if we have a time series for a subset of the assets and knowledge of their market portfolio weights, unless we are willing to make very strong assumptions about the relationships among the omitted assets and the included assets (see Kandel [26]). This is the sense in which the CAPM is not testable on subsets.

Now consider the APT. In order for the factor structure in (1) to make sense in an identifiable way, it is necessary for each factor to be present in a number of assets, so it will not be mistaken for idiosyncratic noise. In an economy with many assets, this is basically a requirement that one can approximately duplicate the factors with linear combinations of the assets, which is item 3 on our list of assumptions, or the requirement that $\beta\beta'$ must have k very large eigenvalues. Clearly, if this condition holds for the restriction of β to the subset, the APT is testable on that subset. This result is useful, since this is a property that will be valid on subsets drawn at random from the same population as the original assets.[17] This is a strong sense in which the APT is testable on subsets.

An example will lead to a generalization of this result. Suppose that there is a factor that is not contained at all in any of the assets in the subset we are considering. In this case, $\beta\beta'$ is singular, but there is still not any possibility of misinterpreting the factor noise of the missing factor as idiosyncratic noise, since the missing factor does not appear at all. Therefore, so long as the remaining factors satisfy the condition that the eigenvalues of $\beta\beta'$ are large in the subset, the tests will still be valid. Omitting the factor in (1) does not distort the test, since in that case it does not appear in (2) either. The generalization of this result is that for the APT to be testable on a subset, it is sufficient for the matrix $\beta\beta'$, for β restricted to the subset, to have some very large eigenvalues with the remaining eigenvalues zero. The relation to the large eigenvalues characterization of significant factors (Chamberlain and Rothschild [5] and Ingersoll [24]) is simple. On the subset, there is essentially a q-factor model where q is the number of large eigenvalues of $\beta\beta'$.

There is in fact an intermediate case in which performing an APT test on a subset of assets is biased towards rejecting the APT. Suppose, for example that there is a factor, call it land value, which occurs in only one asset. Also suppose someone is doing an APT test using stocks (which do not include the land value factor) and a single other asset which is a real estate index. In this case, the land value factor variance in the real estate index will be mistakenly identified as idiosyncratic variance, and the real estate index will appear not to conform to the APT when in fact it does. This example shows how pooling can do damage in tests of the APT. The test would have been unbiased if several different assets including real estate risk had been included, or if the real estate index had been omitted entirely (yielding a weaker test). Generally speaking, if all assets in a subset are taken from the same cross-section, or if many assets are taken from each of several cross-sections, then testing the APT is unbiased even though some assets are not included. In any case, this problem can cause spurious rejection but not spurious acceptance. Therefore, the APT will be rejected on spurious grounds only when it is not of much practical use, that is, when not all of the factors have been found.

One final note on testability on subsets is in order. The tests we have been talking about extracting the factor structure from the asset returns. If the factors are estimated from exogenous economic data (as when the factors are macroeconomic variables whose choice is motivated by economic arguments), the APT is always testable on subsets, and pooling assets will still yield an unbiased

[17] Suppose that β is drawn in cross-section from any probability distribution with a density. Then, with probability one, every $k \times k$ submatrix of $\beta\beta'$ has full rank k. Using this observation, it is easy to show that $\beta\beta'$ is determined from the covariance matrix, whenever we are considering a subset of n assets with $n/2 > k$.

(although less powerful) test. See Chen, Roll, and Ross [9] for an example of a test using macroeconomic variables as factors.

IV. Summary

The Arbitrage Pricing Theory uses a factor model for asset returns to capture the intuition that there are many close substitutes in asset markets. The word "arbitrage" in the name comes from the limiting case in which there is no idiosyncratic noise. In this case, the linearity of expected returns in factor loadings is a direct consequence of the absence of arbitrage, since in this case portfolios with identical factor loadings are perfect substitutes. More generally, the APT follows in theoretical models with assumptions ensuring that portfolios with identical factor loadings are close substitutes. Empirically, the APT should be tested as an equality. Understanding these definitions lies at the heart of the relationship between the theory and empirics of the APT.

Shanken's analysis of the APT has little relevance for actual empirical tests. His "empirical APT" is a straw man that applies the APT to arbitrary portfolios that should never be the subject of empirical tests. Furthermore, he treats a good approximation (which can and should be treated empirically as an equality) as an equality that can be manipulated arbitrarily. (While for many purposes 0.00001 may be approximately zero, if 0.00001 is multiplied by a large enough constant, perhaps one million, the result is not approximately equal to zero times the large constant.) In his transformation of the original assets, Shanken creates portfolios chosen specifically to view all randomness, systematic factors, and idiosyncratic noise alike, as idiosyncratic noise. This is the opposite of diversification, and it is not surprising that applying the APT to these portfolios gives a paradox. The apparently reasonable two-asset example is particularly misleading, since by design it avoids the perverse behavior of Shanken's transformation in examples with many assets and a significant factor structure.

For the original marketed assets, assuming a factor structure with small idiosyncratic noise terms is not errant whimsy. In contrast, if we arbitrarily repackage the securities into portfolios that offer investors equivalent opportunities, we have no intuition concerning the joint distribution of returns for these portfolios. It is conceivable, though, that we may have some ideas concerning the joint distribution of the returns on the original assets. Specifically, firms' returns are affected by macroeconomic variables (such as GNP or inflation) and by the idiosyncratic variables (such as actions of individual workers and managers). However, when a portfolio is constructed to confound systematic and idiosyncratic noise terms, it is not surprising to discover that the factor structure is no longer useful. It is not the assumption of a factor structure which is lacking in economic content. Rather, there is no content in a transform that produces portfolios defined only by their relation to the given covariance structure.

We have examined the relationship between the CAPM and the APT, with special emphasis on testability. Both the CAPM and the APT are closely linked to separation theory—two-fund separation and k-fund separation, respectively. If asset returns are generated by a factor model and no asset is in large aggregate supply, the CAPM implies the APT, but the converse is not true. Still, the models are not strictly nested, since the validity of the CAPM does not rely on any factor structure. Finally, we explored what is meant when it is said that the APT is testable on subsets of the assets while the CAPM is not. We argued that testing the APT on a subset is typically valid. In the degenerate cases for which testability on subsets is biased, the bias is towards rejection, so there is little danger of spurious acceptance of the APT.

Appendix

The following theorem verifies that if there is a significant factor structure, then, as the number of assets grows, either the transformed variances explode or the inverse transform explodes. The assumption of the theorem is that there is a nontrivial factor structure, as defined by Chamberlain and Rothschild [5] and Ingersoll [24]. Specifically, they take a nontrivial k-factor structure to be one for which the k largest eigenvalues of V tend to infinity as the number of assets increases. This definition is consistent with the simple factor structure (1) with each factor appearing significantly in many assets.

THEOREM 1. *Let*

$$V = D + \beta\beta',$$

where D is diagonal and β is $(n \times k)$. Let P transform V to a diagonal $V^* \equiv PVP'$, and let $\|\cdot\|$ denote the matrix norm given by the absolute value of the absolutely largest eigenvalue. If, as the number of assets increases, the factor structure remains nontrivial ($\|V\| \to \infty$), then either the transformed variances explode ($\|V^*\| \to \infty$), or else the inverse transform explodes ($\|P^{-1}\| \to \infty$).

Proof: The proof is a simple application of the triangle inequality applied to the matrix norm. Now,

$$\begin{aligned} \|V\| &= \|P^{-1}V^*P^{-1\prime}\| \\ &\leq \|P^{-1}\|\ \|V^*\|\ \|P^{-1\prime}\| \\ &= \|P^{-1}\|^2\ \|V^*\|. \end{aligned}$$

Since $\|V\| \to \infty$, we are done. Q.E.D.

(Note: If we were to allow correlated error terms, as do Chamberlain and Rothschild [5] and Ingersoll [24], then the theorem would still be valid, with the same proof. Theorem 1 would be stated as follows: If there is a significant factor structure before the transform, then either there is a significant factor structure afterwards, or the inverse transformation explodes.)

It follows immediately from Theorem 1 that any transform that destroys a significant factor structure must be pathological. Suppose that the transformed assets satisfy the assumptions of the zero-factor APT, including specifically that V^* is bounded and diagonal. Let E denote the original expected returns and $E^* \equiv PE$ the transformed returns. If $\|P\| \to 0$, then

$$\|E^*\| = \|PE\| \leq \|P\|\ \|E\| \to 0,$$

whenever the original assets satisfy the APT and $\|E\|$ is bounded. (Here, $\|E\|$ is the usual Euclidian norm of E.)

In other words, it is perfectly consistent to have *both* the original assets *and* the transformed assets satisfy the APT. To put the matter in the negative, it is an error to conclude that if the transformed assets satisfy the zero-factor APT the original assets cannot satisfy their k-factor structure APT.

To the contrary, if

$$\|E^* - re\| \approx 0,$$

then

$$\|E - re\| = \|P^{-1}(E^* - re)\|$$

can be quite different from zero and

$$E \approx r \cdot e + \beta\lambda$$

is perfectly possible. In fact, one interpretation of Theorem 1 is that whenever the transformed assets satisfy the zero-factor APT, the original assets can satisfy the k-factor APT.

The next result shows that in a one-factor world, all but two of the transformation portfolios require short sales.

THEOREM 2. *Let*

$$V = D + \beta\beta'$$

where D is diagonal with positive entries and β is $(n \times 1)$. If

$$V^* = PVP'$$

is diagonal, then no more than two of the transformation portfolios can be positive. That is, no short sales is possible only if $n = 2$.

Proof: Restrict attention to that subset of transformation portfolios not allowing short sales. Within this set, $P > 0$.
For $i \neq j$,

$$V^*_{ij} = P_iDP_j + (P_i\beta)(P_j\beta) = 0.$$

Since $P > 0$, $P_iDP_j > 0$, and therefore $(P_i\beta)(P_j\beta) < 0$. But this means that $P_i\beta$ and $P_j\beta$ have opposite signs. But then if there is a third asset, $k \neq i$ or j, then either $(P_k\beta)(P_i\beta) > 0$ or $(P_k\beta)(P_j\beta) > 0$. This is a contradiction [for the same reason that $(P_i\beta)(P_j\beta) < 0$], and therefore there must not exist a third portfolio, k, not admitting short sales. Q.E.D.

REFERENCES

1. W. Brock. "An Integration of Stochastic Growth Theory and the Theory of Finance." Economics Department Working Paper, University of Chicago, 1978.
2. ———. "Asset Prices in a Production Economy." In J. McCall (ed.), *The Economics of Information and Uncertainty.* Chicago, IL: University of Chicago Press, 1–43, 1982.
3. S. Brown and M. Weinstein. "A New Approach to Testing Asset Pricing Theories: The Bilinear Paradigm." *Journal of Finance* 38 (June 1983), 711–43.

4. G. Chamberlain. "Funds, Factors, and Diversification in Arbitrage Pricing Models." *Econometrica* 51 (September 1983), 1305–23.
5. ——— and M. Rothschild. "Arbitrage, Factor Structure, and Mean-Variance Analysis on Large Asset Markets." *Econometrica* 51 (September 1983), 1281–1304.
6. K. Chan, N. Chen, and D. Hsieh. "An Exploratory Investigation of the Firm Size Effect." CRSP Working Paper, University of Chicago, 1983.
7. N. Chen. "Some Empirical Tests of the Theory of Arbitrage Pricing." *Journal of Finance* 38 (December 1983), 1393–414.
8. ——— and J. Ingersoll, Jr. "Exact Pricing in Linear Factor Models with Finitely Many Assets: A Note." *Journal of Finance* 38 (June 1983), 985–88.
9. ———, R. Roll, and S. Ross. "Economic Forces and the Stock Market: Testing the APT and Alternative Asset Pricing Theories. Yale School of Management Working Paper, 1983.
10. D. Cho, E. Elton, and M. Gruber. "On the Robustness of the Roll and Ross Arbitrage Pricing Theory." *Journal of Financial and Quantitative Analysis* 19 (March 1984), 1–10.
11. G. Connor. "A Factor Pricing Theory for Capital Assets." Kellogg School Working Paper, Northwestern University, 1981.
12. G. Constantinides. "Intertemporal Asset Pricing with Heterogeneous Consumers and without Demand Aggregation." *Journal of Business* 55 (April 1982), 253–67.
13. J. Cox, J. Ingersoll, Jr., and S. Ross. "An Intertemporal General Equilibrium Model of Asset Prices." *Econometrica* 53 (March 1985), 363–84.
14. J. Cragg and B. Malkiel. "Expectations and the Structure of Share Prices." University of Chicago Press, 1982.
15. P. Dhrymes, I. Friend, and B. Gultekin. "A Critical Reexamination of the Empirical Evidence on the Arbitrage Pricing Theory." *Journal of Finance* 39 (June 1984), 323–46.
16. P. Dybvig. "A Practical Framework for the Capital Budgeting Problem." Princeton Financial Research Center Working Paper, 1981.
17. ———. "An Explicit Bound on Deviations from APT Pricing in a Finite Economy." *Journal of Financial Economics* 12 (December 1983), 483–96.
18. M. Garman and J. Ohlsen. "A Dynamic Equilibrium for the Ross Arbitrage Model." Department of Business Administration Working Paper, University of California, Berkeley, 1978.
19. A. Gehr. "Some Tests of the Arbitrage Pricing Theory." *Journal of the Midwest Finance Association* (1975).
20. M. Gibbons. "Empirical Examination of the Return Generating Process of the Arbitrage Pricing Theory." Stanford Business School Working Paper, 1981.
21. M. Grinblatt and S. Titman. "Factor Pricing in a Finite Economy." *Journal of Financial Economics* 12 (December 1983), 497–507.
22. G. Huberman. "A Simple Approach to Arbitrage Pricing Theory." *Journal of Economic Theory* 28 (October 1982), 183–191.
23. P. Hughes. "A Test of the Arbitrage Pricing Theory." University of British Columbia Working Paper, 1982.
24. J. Ingersoll, Jr. "Some Results in the Theory of Arbitrage Pricing." *Journal of Finance* 39 (September 1984), 1021–39.
25. J. Jobson. "A Multivariate Linear Regression Test for the Arbitrage Pricing Theory." *Journal of Finance* 37 (September 1982), 1037–42.
26. S. Kandel. "On the Exclusion of Assets from Tests of the Mean Variance Efficiency of the Market Portfolio." *Journal of Finance* 39 (March 1984), 63–75.
27. G. Oldfield and R. Rogalski. "Treasury Bill Factors and Common Stock Returns." *Journal of Finance* 36 (May 1981), 337–49.
28. P. Pfleiderer. "A Short Note on the Similarities and the Differences between the Capital Asset Pricing Model (CAPM) and the Arbitrage Pricing Theory (APT)." Graduate School of Business Working Paper, Stanford University, 1983.
29. ——— and P. Reiss. "Testing the APT with Arbitrage Portfolios." Stanford Business School Working Paper, 1983.
30. E. Prescott. "Comment on 'Asset Prices in a Production Economy' by W. Brock." In J. McCall (ed.), *The Economics of Information and Uncertainty*. University of Chicago Press, Chicago, IL: 43–46, 1982.

31. M. Reinganum. "The Arbitrage Pricing Theory: Some Empirical Results." *Journal of Finance* 36 (May 1981), 313–22.
32. R. Roll. "A Critique of the Asset Pricing Theory's Tests." *Journal of Financial Economics* (March 1977), 129–76.
33. ———. "Ambiguity When Performance is Measured by the Securities Market Line." *Journal of Finance* 33 (September 1978), 1051–69.
34. ———. "Testing a Portfolio for Ex Ante Mean Variance Efficiency." In E. Elton and M. Gruber (eds.), *TIMS Studies in the Management Sciences.* North-Holland, Amsterdam: Elsevier, 1979.
35. ——— and S. Ross. "An Empirical Investigation of the Arbitrage Pricing Theory." *Journal of Finance* 35 (December 1980), 1073–1104.
36. S. Ross. "Return, Risk, and Arbitrage." In I. Friend and J. Bicksler (eds.), *Risk and Return in Finance.* Cambridge, Massachusetts: Ballinger, 1976.
37. ———. "The Arbitrage Theory of Capital Asset Pricing." *Journal of Economic Theory* 13 (December 1976), 341–60.
38. ———. "On Intertemporal Asset Pricing Theories." School of Management Working Paper, Yale University, 1982.
39. R. Schmalensee. "Risk and Return on Long-Lived Tangible Assets." *Journal of Financial Economics* 9 (June 1981), 185–206.
40. J. Shanken. "The Arbitrage Pricing Theory: Is it Testable?" *Journal of Finance* 37 (December 1982), 1129–40.
41. ———. "Equilibrium, Factors and Arbitrage Pricing." Working Paper, University of California, Berkeley, 1984.
42. R. Stambaugh. "Arbitrage Pricing with Information." *Journal of Financial Economics* 12 (November 1983), 357–69.