UCLA
Recent Work

Title
The Arbitrage Pricing Theory: Estimation and Applications

Permalink
https://escholarship.org/uc/item/7nc1c6rx

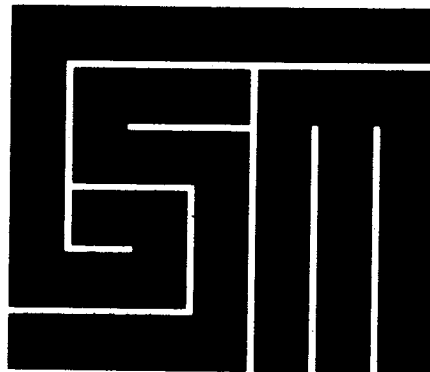**Author**
Chen, Nai-Fu

Publication Date
1980-06-01

October, 1979
(Revised June, 1980)

THE ARBITRAGE PRICING THEORY:

ESTIMATION AND APPLICATIONS

4-80

Nai-Fu Chen
Graduate School of Management
University of California
Los Angeles

**Graduate School of
Management UCLA**

ABSTRACT

The pricing equation of Ross' (1976) APT model is derived using estimable parameters. Estimation errors are discussed in the framework of elementary perturbation analysis. Theoretically, a simple link is provided among the mean-variance efficient set mathematics, mutual fund separations, discrete and continuous time CAPM, option pricing model, term structure of interest rate, capital budgeting, portfolio ranking, Modigliani Miller theorems with the APT.

Acknowledgement

I am greatly indebted to Richard Roll for many valuable comments and suggestions.

I. Introduction

Since the introduction of the arbitrage pricing theory (APT) by Ross [36, 37], the theory has gained theoretic importance because of its simplicity and its performance under empirical scrutiny.[1] Yet practical applications of its pricing equation have been partially hampered by difficulties in estimating the unknown "factors." In this paper, we shall address the estimation problem in the framework of elementary perturbation theory. In addition, a simple link is provided among the mean-variance efficient set mathematics, mutual fund separation, discrete and continuous time CAPM, option pricing model, term structure of interest rates, capital budgeting, portfolio ranking, the Modigliani-Miller theorems and the APT.

We begin by reviewing the main results of the APT. Assume asset markets are perfectly competitive and frictionless, and individuals believe that returns on assets are generated by a k-factor model, i.e., the return on the ith asset can be written as:

$$\tilde{r}_i = E_i + b_{i1}\tilde{\delta}_1 + \ldots + b_{ik}\tilde{\delta}_k + \tilde{\varepsilon}_i \qquad (1)$$

where E_i is the expected return; $\tilde{\delta}_j$, $j = 1, \ldots, k$, are the mean zero factors common to all assets; b_{ij} is the sensitivity of the return on asset i to the fluctuations in factor j; and $\tilde{\varepsilon}_i$ is the "unsystematic" risk component idiosyncratic to the ith asset with $E\{\tilde{\varepsilon}_i | \tilde{\delta}_j\} = 0$ for all j. In equilibrium, the expected return on the ith asset is given by:[2]

$$E_i = \lambda_0 + \lambda_1 b_{i1} + \lambda_2 b_{i2} + \ldots + \lambda_k b_{ik} \quad . \qquad (2)$$

If there exists a riskless (or a "zero beta") asset, then its return will be λ_0. Equation (2) implies that the expected return on any asset i can be written as a linear combination of expected returns on any arbitrary but fixed set of $k + 1$ linearly independent assets,[3] provided that the $k + 1$ expected returns of those assets can be independently estimated.

An unsuspecting practitioner who regresses returns of asset i on returns of the $k + 1$ assets to get E_i will find the results totally disastrous. The reason is that the variances of the idiosyncratic terms are not small[4] and they show up in the coefficient matrix of the normal equation when using simple least square regression. This problem can be partially alleviated by forming large portfolios, but unfortunately, the formation process in general would introduce a more subtle difficulty which we shall discuss in greater detail in section IV.

Nevertheless, our intuition is correct that E_i can be expressed in terms of expected returns of $k + 1$ linearly independent assets. In section II, we shall derive some basic results. In section III, we shall comment on the inheritability of the many applications of CAPM by the APT. Finally, in section IV, we shall analyze the sources of and the remedies for errors in estimating the pricing equation of the APT. Some mathematical derivations and related issues are discussed in the appendices.

II. The Pricing Equation

Suppose there are n assets and k factors in the economy and we possess the covariance matrix of m assets, where n >> m >> k. Assume further that within the m assets there exist k + 1 linearly independent assets. Then using the method of factor analysis and principle components,[5] we can determine the b_{ij}'s, i = 1, . . . ,m and j = 1, . . . ,k corresponding to a set of factors $\{\bar{\delta}_1, ..., \bar{\delta}_k\}$. The set $\{\bar{\delta}_1, ..., \bar{\delta}_k\}$ is not unique in the sense that any isomorphic linear transformation of the set can serve equally well as the k factors, but the b_{ij}'s that we get correspond to the same fixed set of facotrs. If the return generating process is given in terms of an arbitrary set of k factors, then there exists a unique linear transformation that takes this arbitrary set of factors into the fixed set that the obtained b_{ij}'s correspond to. Further, even though we do not know the identity of the factors, with proper normalization, the variances of the $\bar{\delta}_j$, j = 1, . . . ,k are known and given as the factor scores (eigenvalues) corresponding to the factor loadings (eigenvectors), and the covariance between any two distinct factors in the set is zero. Note that the converse is not true, i.e., it is not true generally that the variances of an orthogonal set of factors are eigenvalues. These assertions are proved in Appendix A. Define $\sigma_j^2 \equiv Var(\bar{\delta}_j)$. Aspects of obtaining clean estimates of the σ_j^2's and b_{ij}'s are discussed in section IV.

Obviously all our problems are solved if we can locate all the n assets in the economy and devise a fast enough algorithm to decompose the n x n covariance matrix. Since this is currently impossible,

we shall resort to an alternative scheme to manufacture "betas" from known "betas".

Theorem: Let $\bar{r}_i$, $i = 1, \ldots, k + 1$ be the returns of $k + 1$ linearly independent assets (portfolios) such that b_{ij}, $i = 1, \ldots, k + 1$, $j = 1, \ldots, k$ are known. Let r_p be the return of the pth asset, $p \neq i$, $i = 1, \ldots, k + 1$ with $\text{Cov}(\bar{\epsilon}_p, \bar{\epsilon}_i) = 0$.

(i) If $\text{Cov}(\bar{r}_p, \bar{r}_i)$ is known for $i = 1, \ldots, k$, then $b_{p1}, \ldots, b_{pk}$ can be uniquely determined.

(ii) If, further, E_i is known for $i = 1, \ldots, k + 1$, then the equilibrium E_p can be determined

Proof: (i) $\text{Cov}(\bar{r}_p, \bar{r}_i) = b_{p1}b_{i1} \text{Var}(\bar{\delta}_1) + \ldots + b_{pk}b_{ik} \text{Var}(\bar{\delta}_k)$

$$= b_{p1}b_{i1}\sigma_1^2 + \ldots + b_{pk}b_{ik}\sigma_k^2 \, ,$$

$$i = 1, \ldots, k \, .$$

The above k linear equations can be written as

$$
\begin{bmatrix}
b_{11}\sigma_1^2 & b_{12}\sigma_2^2 \cdots b_{1k}\sigma_k^2 \\
b_{21}\sigma_1^2 & \qquad\qquad\cdot \\
\cdot & \\
b_{k1}\sigma_1^2 & b_{k2}\sigma_2^2 \qquad b_{kk}\sigma_k^2
\end{bmatrix}
\begin{bmatrix}
b_{p1} \\
b_{p2} \\
\cdot \\
b_{pk}
\end{bmatrix}
=
\begin{bmatrix}
\text{Cov}(\bar{r}_p, \bar{r}_1) \\
\cdot \\
\cdot \\
\text{Cov}(\bar{r}_p, \bar{r}_k)
\end{bmatrix}
\qquad (3)
$$

Since the coefficient matrix is nonsingular, $(b_{p1}, \ldots, b_{pk})$ can be uniquely determined.

(ii) In equilibrium, all portfolios with the same set of betas will have the same expected returns (i.e. (2) holds). To construct a portfolio from the $k + 1$ assets that possess the same betas as the pth asset, we solve

$$
\begin{bmatrix} b_{11} & b_{21} & \cdots & b_{k+1,1} \\ \cdot & & & \cdot \\ b_{1k} & & \cdots & b_{k+1,k} \\ 1 & & \cdots & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ \cdot \\ x_k \\ x_{k+1} \end{bmatrix} = \begin{bmatrix} b_{p1} \\ \cdot \\ b_{pk} \\ 1 \end{bmatrix} \tag{4}
$$

The last equation ensures that the investment proportions $(x_1, \ldots, x_{k+1})$ add up to unity. Again since the coefficient matrix is non-singular, the solution $(x_1, \ldots, x_{k+1})$ is unique and

$$E_p = x_1 E_1 + \ldots + x_{k+1} E_{k+1}. \qquad \text{QED}$$

Take the $k + 1^{st}$ asset to be the riskless asset (or a zero beta portfolio). Define $(E - \lambda_0)'$ to be the k dimensional row vector whose ith component is $E_i - \lambda_0$, $\underline{Cov}(r_p, r)$ to be the column vector whose ith component is $Cov(\bar{r}_p, \bar{r}_i)$ and V to be the matrix whose (i, j) element is $b_{i1}b_{j1}\sigma_1^2 + \ldots + b_{ik}b_{jk}\sigma_k^2$. With some algebraic manipulations (carried out in Appendix A), the pricing equation of the APT can be written as:

$$
E_p = \lambda_0 + (E - \lambda_0)' \, V^{-1} \, \underline{Cov} \, (\bar{r}_p, \bar{r}) . \tag{5}
$$

Equation (5) is the general k-factor pricing equation in terms of any k + 1 linearly independent assets or portfolios. The last term on the right hand side, i.e., the inner product of "risk premium" with "risk factors," exhibits the linear risk-return relationship of the model.

It is immediately apparent from (1) and (5) that the market model (see Fama [10], p. 37) is a special case of the APT. Indeed, letting $k = 1$ and $\bar{\delta} = \bar{r}_m - \bar{E}_m$, we obtain the well known pricing equation:

$$E_p = \lambda_0 + (E_m - \lambda_0)(\sigma_m^2)^{-1} \, Cov(\bar{r}_p, \bar{r}_m) \, . \qquad (6)$$

Before proceeding to tie together various asset pricing models, it might be illuminating to digress and consider the following line of reasoning that leads to the traditional Sharpe [42] - Lintner [26] - Merton [28] CAPM, with the restriction being placed on the asset price distribution (rather than on the utility functions).

For any period t, there exists an ex ante mean-variance efficient frontier for risky assets. Let p be any portfolio and q be the portfolio on the efficient frontier that has the same expected return as p. Consider

$$\bar{r}_p = \bar{r}_q + (\bar{r}_p - \bar{r}_q) \equiv \bar{r}_q + \bar{h}.$$

we have

$$E(\bar{h}) = 0 \qquad (7a)$$

and from the efficient set mathematics (for a concise description, see Roll [34]), $\quad Cov(\bar{r}_p, \bar{r}_q) = \sigma_q^2 = Cov(\bar{r}_q, \bar{r}_q)$, hence

$$Cov(\bar{h}, \bar{r}_q) = 0 \qquad (7b)$$

However, h is not a noise term. To be a noise term (see Rothschild and Stiglitz [39], Ross [38]), we must have

$$E(\bar{h}|r_q) = 0 \quad \text{for all } r_q \qquad (7c)$$

which is weaker than independent (between r_q and h) but stronger than uncorrelated. If (7c) is true, then the mean-variance efficient frontier is the relevant investment opportunity set for all risk averse utility maximizers.[6] If a riskless asset exists, then the relevant frontier becomes a straight line. In either case, there is two fund separation.

Proposition: If all asset returns belong to a class of distributions that is closed under linear combination (with finite variances), and such that, (7a) and (7b) together imply (7c), then the economy exhibits two fund separation under perfect markets.

Corollary: CAPM holds under the multivariate normal distribution.[7]

Since the above analysis applies equally well to an instantaneous efficient frontier as to an ex ante frontier, it is intuitive to think of Merton's [28] continuous time CAPM in these terms with assets exhibiting a non-anticipating "locally multivariate normal distribution." Going a step further, it is obvious that Merton's [29] continuous time CAPM with k - 1 stochastic factors can be written in the same form as (5). In particular, Merton's equation (34) ([29], p. 882):

$$E_p = \lambda_0 + \frac{\sigma_p[\rho_{pm} - \rho_{pn}\rho_{nm}]}{\sigma_m(1 - \rho_{nm}^2)}(E_m - \lambda_0) + \frac{\sigma_p[\rho_{pn} - \rho_{pm}\rho_{mn}]}{\sigma_n(1 - \rho_{mn}^2)}(E_n - \lambda_0) \quad (8)$$

can be obtained by choosing $\bar{\delta}_1 = \bar{r}_m - E_m$ and $\bar{\delta}_2 = \bar{r}_n - E_n$ (see below) with r_n being the asset return that is perfectly negatively correlated with the shift in the riskless return. Since $\bar{\delta}_1$ and $\bar{\delta}_2$ may be correlated, V would contain some correlation terms that show up in (8).

Another model that is consistent with (5) is the Kraus and Litzenberger [23] skewness preference model. By taking $\bar{\delta}_1 = \bar{r}_m - E_m$ and $\bar{\delta}_2 = (\bar{r}_m - E_m)^2$, we obtain their equation (6) ([23], p. 1090).[8]

What is the relationship between the APT and the mean-variance efficient set mathematics? To answer this question, let us take Merton's CAPM with a stochastic interest rate as an example. From the point of view of the APT, equation (8) holds whether portfolio m is taken to be the "market portfolio" or the "first portfolio," call it portfolio A, defined by Merton in his Theorem 2 ([29], p. 880). This is obvious because portfolio A is a linear combination of the "market portfolio" and portfolio n, and vice versa. Portfolio A, with investment proportion in the ith asset given by the ith component of the vector

$$\ddagger^{-1}(E - \lambda_0) \,/\, \underline{e}' \,\ddagger^{-1}(E - \lambda_0)$$

where $\ddagger$ is the instantaneous covariance matrix and $\underline{e}$ is a vector of all one's, lies on the instantaneous efficient frontier of risky assets, and, together with the riskless asset, provides "the service to investors of an instantaneous efficient, risk-return frontier." Hence, if (8) is given with m being the portfolio A, then (8) and (6) simultaneously hold with the same m. This is true because m is then located on the instantaneous efficient frontier and the efficient set mathematics applies. Although the mathematics was developed for the discrete time case the continuous time case follows trivially. In fact, the pricing equation of a k factor multi-beta APT (eq. (5), discrete or continuous time) can always be reduced to a single beta equation (eq. (6)). The trouble is that the investment proportions of portfolio A as well as the betas computed against it

may change with time. In the discrete time case, a portfolio with
fixed investment proportions that is on the ex-ante efficient frontier
for a period may not be on the frontier for any subsequent period
or any subperiod. This problem can be partially alleviated if
portfolio A is "tracked" by another variable. Recently, Breeden [4]
has found that "real consumption" can replace portfolio A in the
single beta pricing equation, given appropriate assumptions.[9]

An important point to note here is that even though we are
always able to collapse a k factor pricing equation into a single
beta equation, we do not generally have two fund separation. In a
single period k-factor APT model, if we choose an orthogonal set of
k factors with one of them being the portfolio on the efficient
frontier, then the portfolios corresponding to the remaining k - 1
factors will have zero risk premium. If we then form k - 1 zero in-
vestment portfolios (from the k - 1 orthogonal portfolios) by
shorting the riskless asset, then we would have k - 1 portfolios
with no investment, zero expected return and uncorrelated with the
efficient portfolio. However, not every risk averse investor would
prefer investing only in the riskless asset and the efficient portfolio
because generally uncorrelated does not imply independent. (The
k - 1 zero investment, zero expected return portfolios behave like
h above). Algebraically, a portfolio p represented by $(b_{p1}, \ldots, b_{pk}, 1)$
(see footnote 3) can be decomposed into:

$$
\begin{aligned}
(b_{p1}, \ldots, b_{pk}, 1) &= (b_{p1}, 0, \ldots, 0, 1) \\
&\quad + (0, b_{p2}, 0, \ldots, 0) \\
&\quad + \ldots \\
&\quad + (0, \ldots, 0, b_{pk}, 0).
\end{aligned}
$$

Therefore, if we let $\bar{\delta}_1$ be on the mean variance efficient frontier and $\mathrm{Cov}(\bar{\delta}_1, \bar{\delta}_i) = 0$, $i = 1 \ldots, k$, then $(0, b_{p2}, 0, \ldots, 0), \ldots,$ $(0, \ldots, 0, b_{pk}, 0)$ are $k - 1$ zero investment, zero expected return portfolios uncorrelated with $(b_{p1}, 0, \ldots, 0, 1)$. However, not every investor would prefer $(b_{p1}, 0, \ldots, 0, 1)$ to $(b_{p1}, b_{p2}, \ldots, b_{pk}, 1)$. Therefore, we do not have two fund separation. A well known example is in Hakansson [17]. In the continuous time APT model with explicit state dependent utility functions, the $k - 1$ portfolios are used to provide hedges.

III. Applications

Many theorectical and practical applications of the CAPM will follow easily under the APT. In this section, we shall briefly comment on several obvious results in the area of corporate finance, term structure of interest rates and option pricing. The list is only for illustrative purposes and is by no means exhaustive.

Let us first consider the question of capital budgeting in a single period setting. If we can estimate subjectively $Cov(\bar{r}_p, \bar{r}_i)$, $i = 1, \ldots, k$ where r_p is the return on investment, and have estimates of E_i, $i = 1, \ldots, k + 1$ (in the same way that we estimate $Cov(\bar{r}_p, \bar{r}_m)$, E_m and λ_0 with the CAPM) then (5) can be used as a benchmark to distinguish profitable from unprofitable investment in the same way as the Security Market Line in the CAPM. The only difference is that we now have a k dimensional hyperplane. The investment is acceptable if its expected internal rate of return lies on or above the "Security Market Plane" in the k + 1 dimensional space with return on the "vertical" axis.

Theorem: A project's net present value is non-negative if and only if its internal rate of return $\geq$ R*, R* given by (5).

The result is intuitively clear. The proof and some comments on the multiperiod decision are provided in Appendix B. A closely related issue is ranking portfolios using the Security Market Plane. We shall return to this problem in the next section.

Other results in Rubinstein [40] follow under the APT with parallel arguments; specifically,

(i)   Proof of Modigliani-Miller proposition I with risky corporate
        debt and corporate taxation

(ii)   proof of Modigliani-Miller proposition II revised for risky
        corporate debt

(iii)   analysis of the separate effect of operating risk and financial
        risk on equity risk premium

(iv)   analysis of the components of operating risk.

Another well known application of the CAPM involves the option
pricing model. The derivation of the inverse parabolic differential
equation that governs the option price by Black and Scholes [1]
using the CAPM can be done using the APT in exactly the same
manner. The results in Galai and Masulis [14] follow by
considering the entire risk premium term rather than just beta in
their equations (8), (9) and (10).

It is not surprising that the option pricing equation holds
under a general equilibrium model such as the APT. Since the option
pricing model is derived by assuming partial equilibrium (i.e., no
arbitrage profit) among several securities, it must hold, in parti-
cular, with the stronger assumption of no arbitrage profit among all
securities. Consequently, prices obtained using the same type of
hedging arguments must be consistent with those obtained using the
APT. Important examples include the compound option formula of
Geske [16] and many recent theories of the term structure of inter-
est rates.

The derivation of option prices using a general equilibrium model such as CAPM, APT or Rubinstein's [41] model is important because not all underlying parameters are traded assets. Furthermore, using the APT to derive the option pricing differential equation is relatively easy and straightforward. The governing differential equation with stochastic interest rate in Merton ([30], equation (33), p. 164-165, in the interval when the bond price is following geometric Brownian motion) can be readily derived in a few lines. This technique can be used to model other variations in the basic Black-Scholes model assumptions. While not all those mixed initial-boundary problems have quasi-analytic solutions, many powerful numerical algorithms are available provided the boundary conditions can be meaningfully discretized and error propagation is stable.[10]

In the theory of the term structure of interest rates, it is well known that if both lenders and borrowers are risk averse, the forward rate will generally differ from the expected spot rate. If we assume that the liquidity premium arises solely form "risk," then we can use the APT to price "risk." This was the approach taken by Roll [33] when he used the CAPM to price "risk." From the term structure relationship, the liquidity premium of an N period bond is the average of the risk premiums of bonds of maturities from one to N periods.[11]

Recently, many authors have derived the term structure using hedging arguments. They differed in their assumptions about the number of state variables and the governing stochastic process. The first published paper along this line is the work of Vasicek [43] and the most comprehensive one is the yet unpublished work of Cox, Ingersoll and Ross [5]. As mentioned earlier, their conclusion must necessarily be consistent with the pricing of the APT whenever the APT holds, with perhaps the attendant simplifying assumptions. For example, the two factors assumption of Brennan and Schwartz [2] is equivalent to the assumption that there exists a linear transformation of the factors such that bond prices are sensitive to at most only two factors in a k factor world.

Before going on, let us reflect on a rather interesting point. Many theoretical applications of CAPM require only that (6) hold. But we know that (6) always holds as long as a mean variance efficient frontier exists. Can we conclude that, say, the MM proposition holds so long as we have an efficient frontier? Unfortunately, this is only an illusion. The efficient frontier depends on expected returns which depend on prices which we try to show are leverage irrelevant.

IV. Estimation of the pricing equation

In this section, we shall analyze the best way to estimate the expected return under the APT in the framework of perturbation analysis. These analyses can be found in most elementary textbooks on numerical analysis[12] and are relevant to the method described in section II as well as to any regression method. We shall begin by reviewing some of the observations made by von Neuman[13] in the late 1940's.

In solving a matrix equation $Ax = y$, we consider the effect of perturbations in y and in A on the computed solution $x + \delta x$, where δx denotes the error. If we let $A(x + \delta x) = y + \delta y$, where δy is the perturbation in y, then

$$||\delta x|| \leq ||A^{-1}|| \ ||\delta y|| \tag{9}$$

where $|| \cdot ||$ is any vector norm when applied to vectors and $|| \cdot ||$ is the subordinate matrix norm to the vector norm when applied to matrices. If we let $(A + \delta A)(x + \delta x) = y$, with δA as the perturbation in A, then

$$||\delta x||/||x + \delta x|| \leq \gamma ||\delta A||/||A|| \tag{10}$$

with $1 \leq \gamma \equiv ||A|| \ ||A^{-1}||$ which is called the condition number of the matrix A. The inequalities in (9) and (10) are sharp in the sense that equalities can hold. If we combine the above two cases and ignore error terms of second order, then the error bound for the case $(A + \delta A)(x + \delta x) = y + \delta y$ is given by

$$\frac{||\delta x||}{||x||} \leq \frac{\gamma}{1 - ||A^{-1}|| \, ||\delta A||} \left(\frac{||\delta A||}{||A||} + \frac{||\delta y||}{||y||} \right) \quad (11)$$

provided δA is so small that $||\delta A|| \leq 1/||A^{-1}||$. Equation (11) says that the relative error in the computed solution is bounded by a multiple of the sum of relative errors in y and A. This also points out the crucial role played by γ, the condition number of A. When γ is large, A is called ill conditioned, and the computed solution is not reliable because any error may be magnified by a factor of the order of γ.

In our case of computing (5), aside from sampling errors, $\underline{\delta y}$ will contain $Cov(\bar{\epsilon}_p, \bar{\epsilon}_i)$ whenever they are not all zeros. Also, δA will contain errors from the finite arithmetics carried out inside the computer. This component of δA will be ignored in our analysis because when k is small the cure is cheap and simple. If, instead of computing (5), we regress returns of asset p on returns of k + 1 assets, δA will contain the covariance matrix of $\bar{\epsilon}_i$, i = 1, . . . ,k + 1 when solving the normal equation, in addition to the perturbation mentioned above.

How large is γ before we call A ill conditioned? This is related to the size of $||\delta A||/||A||$ and $||\delta y||/||y||$. In most scientific applications in which A and y are measured quite accurately, matrices with $\gamma \leq 100$ are said to be well conditioned. But in social science, where A and y may contain relative errors of up to 1% or more, a $\gamma = 100$ may produce a computed solution with no single meaningful digit. What is worse is the fact that it is hard to judge from appearance how ill conditioned A is.

The innocent looking symmetric positive definite matrix

$$\begin{bmatrix} 5 & 7 & 3 \\ 7 & 11 & 2 \\ 3 & 2 & 6 \end{bmatrix}$$

has a condition number that is roughly equal to 1500. So some relative error may be magnified over a thousand times. A numerical example is constructed in Appendix C using the data in Miller and Scholes [31] where they considered the problem of why the assets' own variances have explanatory power in cross sectional regression.

Now let us specialize the perturbation theory of linear systems to our case. Suppose we decompose a reasonably sized covariance matrix of, say 200 x200 and obtain σ_j^2 and b_{ij}; then we can construct the matrix in (3) by forming portfolios from the 200 assets with the aim of making the matrix as diagonal as possible[14] (the same asset may enter into all k portfolios!). In addition, if all the weights of the assets in the portfolios are roughly of the same order, the process will reduce the random errors that may appear in individual b_{ij}'s (e.g. due to sampling errors). Further, if the $Cov(\varepsilon_p, \varepsilon_i)$ are randomly distributed about zero, then the covariance between the idiosyncratic terms will be much reduced. Hence we can reduce γ, δA, δy simultaneously (after knowing b_{ij}), and then use (5) to generate expected returns or use (3) to generate betas.

At this point, let us examine possible errors that may have crept into the computation of σ_j^2 and b_{ij}. There are two sources of errors: (i) sampling errors, (ii) possible contamination from idiosyncratic variances of individual assets when factoring the covariance matrix. The second source can be greatly reduced by factoring a large enough matrix, yet both sources can be reduced simultaneously if we use the following two stage process. In the first step, we factor a reasonably sized covariance matrix, determine the number of factors, aggregate to form the matrix in (3), and generate betas for a large class of assets. In the second step, we form portfolios of, say, 20 assets each, in such a way that (i) the idiosyncratic term of each portfolio is small (ii) none of the common factors will be diminished in the resultant portfolios. This can be done by successively emphasizing the jth common factor, $j = 1, \ldots, k$. Formation of these portfolios cannot be done before step 1 because we would not know the estimated b_{ij}, and arbitrary formation of portfolios may diminish the influence of a common factor. After forming the portfolios, we can compute the covariance matrix of the portfolios and repeat step 1. This two stage procedure can be iterated to achieve as accurate an answer as we wish subject of course to sampling errors of portfolios, but we suspect almost all benefits are realized in the first two or three passes.[15]

Obviously, the above procedure involves a nontrivial amount of computations. Can all these be avoided using simple regression on $k + 1$ portfolios? Such regression would be acceptable if

we know how to form $k + 1$ portfolios that are almost orthogonal and how to simultaneously reduce the idiosyncratic terms of the resultant portfolios without the knowledge of the b_{ij}'s of the individual assets. Generally, there will be a tradeoff between the two goals, and it is not clear which is the best way, if any.

Finally, let us turn to perhaps the most difficult problem associated with the empirical work -- the determination of the number of factors. Since the determination from factor analysis depends on some statistical assumptions, we are quite concerned about the impact of leaving out a factor. Fortunately, the penalty for carrying an extra factor or two is small, as long as we take care to isolate the "extra" factors from the regular ones. Therefore, in practice, a rule of thumb is when in doubt include an extra factor. As an additional safeguard, we can test if any important factor is left out by following the procedure Roll and Ross [35] used for testing the significance of own variances and append any factor, if necessary.

Having discussed the problems of estimating the pricing equation (and thus the Security Market Plane), we can look into the problem of ranking portfolios. If the pricing equation can be estimated without error, then the Jensen [22] index may be used. Unfortunately, it is clear from the above discussion that some errors will always be present. However, if we take care to keep such error small, bounds on the ambiguity when performance is measured by the Security Market Plane can be established. This would enable us to compute an incomplete ordering of performance. The calculation of the bounds involves only elementary numerical linear algebra; therefore, it is omitted here.

V.   Summary

The explicit pricing equation of the APT model is derived
using estimable parameters. Although the two stage method involves
rather substantial computations even with custom-designed algorithms
that can compute extreme eigenvalues and eigenvectors directly,
it seems to be worthwhile, especially if it can be carried out
on a vector machine or a super computer, and the results shared
among researchers. On the theoretical side, we have seen that many
asset pricing models are either special cases of the APT or implied
by the APT.

Appendix A

Suppose asset returns are generated by $\bar{r}_i = E_i + b_{i1}\bar{\delta}_1 + \ldots$ $b_{ik}\bar{\delta}_k + \bar{\epsilon}_i$. From factor analysis, with the normalization chosen below, we obtain b'_{ij}. We would like to show that there exists a unique linear transformation that will take b_{ij} into b'_{ij} and a related transformation that will take $\{\bar{\delta}_1, \ldots, \bar{\delta}_k\}$ into an orthogonal set $\{\bar{\delta}'_1, \ldots, \bar{\delta}'_k\}$ such that $var(\bar{\delta}'_j) = \Theta_j$, the eigenvalue from the factor analysis, and $\bar{r}_i = E_i + b'_{i1}\bar{\delta}'_1 + \ldots + b'_{ik}\bar{\delta}'_k + \bar{\epsilon}_1$.

All versions of factor analysis are designed to estimate the $m \times m$ covariance matrix Ω net of the idiosyncratic terms from the gross covariance matrix. Hence $\Omega = (\omega_{ij}) = Cov(\bar{r}_i - \bar{\epsilon}_i, \bar{r}_j - \bar{\epsilon}_j)$. Choose the normalization $\Omega B = B\Lambda$, $B'B = I$ where B is a $m \times k$ matrix of b'_{ij}'s and $\Lambda = diag(\Theta_1, \ldots, \Theta_k)$. Let T be the $k \times k$ matrix representation on the canonical basis of the linear operator.

$$T: R^k \rightarrow R^k$$

$$\begin{bmatrix} b_{i1} \\ \vdots \\ b_{ik} \end{bmatrix} \mapsto \begin{bmatrix} b'_{i1} \\ \vdots \\ b'_{ik} \end{bmatrix}$$

From linear algebra, T exists, and it is unique if there are k linearly independent assets. Now consider the transformation

$$\begin{bmatrix} \bar{\delta}'_1 \\ \vdots \\ \bar{\delta}'_k \end{bmatrix} = (T')^{-1} \begin{bmatrix} \bar{\delta}_1 \\ \vdots \\ \bar{\delta}_k \end{bmatrix} .$$

We have

$$\bar{r}_i - \bar{\epsilon}_i = (b_{i1}, \ldots, b_{ik}) \begin{bmatrix} \bar{\delta}_1 \\ \vdots \\ \bar{\delta}_k \end{bmatrix} = (b'_{i1}, \ldots, b'_{ik})(T')^{-1} \begin{bmatrix} \bar{\delta}_1 \\ \vdots \\ \bar{\delta}_k \end{bmatrix}$$

$$= (b'_{i1}, \ldots, b'_{ik}) \begin{bmatrix} \bar{\delta}'_1 \\ \vdots \\ \bar{\delta}'_k \end{bmatrix} ,$$

and

$$\Lambda = B' \Omega B = B'(\text{Cov}(\bar{r}_i - \bar{\epsilon}_i, \bar{r}_j - \bar{\epsilon}_j))B$$

$$= B'(B \{\text{covariance matrix of } \bar{\delta}'_1, \ldots, \bar{\delta}'_k\}B')B$$

$$= \{\text{covariance matrix of } \bar{\delta}'_1, \ldots, \bar{\delta}'_k\} .$$

Therefore, we have shown the existence of a linear operator that accomplishes the goal. It is also clear that if $\Omega = C \Gamma C'$ where columns of C are not orthonormal, then even though Γ may be diagonal, its elements are not eigenvalues of Ω. Hence, the converse is not true.

To derive equation (5) from the theorem in section II, we first combine the two parts of the theorem into the matrix equation:

$$E_p = \underline{E}' F_2^{-1} \begin{bmatrix} F_1^{-1} \underline{Cov}(\tilde{r}_p, \tilde{r}) \\ 1 \end{bmatrix} \tag{A.1}$$

where $\underline{E}' = (E_1, \ldots, E_{k+1})$, F_1, F_2 are the coefficient matrices in equations (3) and (4) respectively and $\underline{Cov}(\tilde{r}_p, \tilde{r})$ is the vector $(Cov(\tilde{r}_p, \tilde{r}_1), \ldots, Cov(\tilde{r}_p, \tilde{r}_k))'$.

To reduce (A.1) to a form that resembles the CAPM equation, we note that the riskless (or a zero beta) asset also satisfies (A.1), hence

$$\lambda_0 = \underline{E}' F_2^{-1} \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

therefore

$$E_p - \lambda_0 = \underline{E}' F_2^{-1} \begin{bmatrix} F_1^{-1} \underline{Cov}(\tilde{r}_p, \tilde{r}) \\ 0 \end{bmatrix}. \tag{A.2}$$

If we let the $k+1^{st}$ asset be the riskless asset, then the last column of F_2 becomes $(0, \ldots, 0, 1)'$, therefore

$$F_2^{-1} = \begin{bmatrix} F_3^{-1} & 0 \\ -\underline{e}' F_3^{-1} & 1 \end{bmatrix}$$

where F_3 is the principal $k \times k$ minor of F_2 and $\underline{e}'$ is the k component vector $(1, \ldots, 1)$. The product $(E_1, \ldots E_k, \lambda_0) F_2^{-1}$ is a row vector whose first k elements are given by

$$(E_1 - \lambda_0, \ldots, E_k - \lambda_0) F_3^{-1} \equiv (\underline{E} - \underline{\lambda_0})' F_3^{-1} .$$

Hence equation (A.2) can be written as

$$E_p - \lambda_0 = \underline{(E - \lambda_0)}' \, F_3^{-1} \, F_1^{-1} \, \underline{Cov}(\tilde{r}_p, \tilde{r}) \ .$$

Defining $V \equiv F_1 F_3$, we have equation (5)

$$E_p = \lambda_0 + \underline{(E - \lambda_0)}' \, V^{-1} \, \underline{Cov}(\tilde{r}_p, \tilde{r}) \ .$$

Appendix B

To prove the theorem in section III, let V_0 be the equilibrium present value of the project, I be its investment cost, $\tilde{V}_1$ be the end of period value, $E_p = E(\tilde{V}_1)/V_0 - 1$ be the equilibrium return, and $\tilde{r}_p = \tilde{V}_1/I - 1$ be the internal rate of return. By (5), in equilibrium

$$E_p = (E - \lambda_0)' \ V^{-1} \ \underline{Cov}(\tilde{r}_p I/V_0, \ \tilde{r}) \ .$$

Solving for V_0 yields

$$V_0 = \frac{E(\tilde{V}_1) - (E - \lambda_0)' \ V^{-1} \ \underline{Cov}(\tilde{r}_p, \tilde{r}) \ I}{1 + \lambda_0} \ .$$

Hence the net present value of the project is

$$V_0 - I = \frac{E(\tilde{V}_1) - (E - \lambda_0)' \ V^{-1} \ \underline{Cov}(\tilde{r}_p, \tilde{r}) I - I - \lambda_0 I}{1 + \lambda_0}$$

$$= \frac{I}{1 + R_0} \left\{ \left[\frac{E(\tilde{V}_1)}{I} - 1 \right] - \left(\lambda_0 + (E - \lambda_0)' V^{-1} \underline{Cov}(\tilde{r}_p, \tilde{r}) \right) \right\}$$

$$= \frac{I}{1 + \lambda_0} (IRR - R*) \ . \tag{B.1}$$

Therefore, net present value is nonnegative if and only if $IRR \geq R*$. Q.E.D.

Equation (B.1) gives an explicit formula relating NPV to I, IRR and $R*$; therefore, it can be used to rank mutually exclusive

projects too. A point that is worth repeating here is that whenever IRR $\neq$ R*, the proper discount rate, call it d, that discounts end of period $E(\tilde{V}_1)$ to V_0 (i.e., $V_0 = E(\tilde{V}_1)/(1 + d)$) is neither IRR nor R*. From (B.1), d can be solved and is given by

$$d = \frac{\lambda_0 - (E - \lambda_0)'V^{-1} \underline{Cov}(\tilde{V}_1/E(\tilde{V}_1), \tilde{r}_i)}{1 - (E - \lambda_0)'V^{-1} \underline{Cov}(\tilde{V}_1/E(\tilde{V}_1), \tilde{r}_i)} \ . \tag{B.2}$$

Equation (B.2) is especially important in a multiperiod setting because we can no longer define IRR properly in general and there is no multiperiod analog of (B.1). Since there is no theoretical reason why the APT is limited to a single period model, extension to the multiperiod case can be naturally accomplished following the reasoning and techniques in Fama [9].

Another interesting point was considered by Fama on what types of uncertainty are allowed in the parameters of multiperiod asset pricing models. If the intertemporal stochastic shifts can be perfectly hedged (e.g. Black and Scholes [1], Merton [29], Cox, Ross and Rubinstein [7]), then all market parameters may be assessed as if they were deterministic.

Appendix C

In this appendix we shall discuss briefly the problem of errors in variables when two or more explanatory variables are involved. Let us first consider the regression equation in Miller and Scholes [31]:

$$R_i = a_0 + a_1 b_i + a_2 s^2(e_i) + \epsilon_i$$

where R_i is the annual return, b_i is an estimate for beta and $s^2(e_i)$ is the residual variance (defined in detail in Miller and Scholes). Reconstructing the data given on p. 53, table 1 of their paper, the solution $(a_0, a_1, a_2) = (.127, .042, .310)$ satisfies approximately the matrix equation $(X'X\beta = X'y)$:

$$\begin{bmatrix} 631 & 631 & 47.96 \\ 631 & 820.5 & 64.67 \\ 47.96 & 64.67 & 12.58 \end{bmatrix} \begin{bmatrix} .127 \\ .042 \\ .310 \end{bmatrix} = \begin{bmatrix} 121.8 \\ 134.6 \\ 12.71 \end{bmatrix} .$$

Now suppose the true beta is approximated by b_i with error; we want to know how much error would have caused a_2 (the slope coefficient of the residual variance) to go from zero to .310 $(t = 11.76)$. Suppose the error in b_i has zero mean; then the only numbers that contain errors in the above matrix equation would be those in the boxes. We shall ignore the (3,3) element in the matrix since it is irrelevant when $a_2 = 0$. The elements (2,3) and (3,2) must be the same. Therefore, we are free to adjust only four entries.

We changed the two entries in the matrix by 5% and we were able
to come up with the following solution

$$
\begin{bmatrix} 631 & 631 & 47.96 \\ 631 & 779 & 67.9 \\ 47.96 & 67.9 & 12.58 \end{bmatrix} \begin{bmatrix} .053 \\ .140 \\ 0 \end{bmatrix} = \begin{bmatrix} 121.78 \\ 142.5 \\ 12.05 \end{bmatrix} \quad . \quad .
$$

The two entries on the right hand side were changed by 5.8% and 5.2%
respectively. The solution $a_1 = .140$ and $a_2 = 0$ was much closer
to the expected value if the CAPM was correct.

We can see from the numerical example constructed above that
a change of 5% - 6% would produce a drastic change in the solution
of the regression. However, the numbers in this example should not
be taken at face value because of the artificial nature of the revised
matrix. Also, one reason why the original regression coefficients
looked so bad is probably the existence of missing factors (and
that is why we have APT) and of skewness-induced spurious dependence
as discussed by Miller and Scholes, and Roll and Ross [35]. Never-
theless, this example contains a message:

(i) Although it is well known, it might be worthwhile to emphasize
     again that the usual "bias towards zero" discussion of the
     errors-in-variables problem applies generally only in the case
     where there is one explanatory variable. It is no longer true
     when more than one explanatory variables are involved.

(ii) The errors-in-many-variables problem is quite difficult.
     Equation (11) provides a highly exaggerated "average

bound." It is exaggerated because of the special form of error that can occur. It is an "average bound" because it is in norm form rather than in component form. In the previous example, if the true a_3 is zero, then in component form, the relative error is infinite.

(iii) If we are willing to do more work, then in some special cases, a tighter bound is possible. See Leamer ([25], pp. 254-255).

(iv) In general, if we have many explanatory variables, we should be very concerned about measurement error. The answer to many puzzling results might be simpler than we thought.

FOOTNOTES

1/ See Gehr [15], Roll and Ross [35].

2/ This statement of the APT is best regarded as an intuitive result. See Ross [36, 37, 38], Huberman [19].

3/ Associated with every asset i (or portfolio of assets) is a k + 1 - vector (bil,...,bik, n) where bil,...,bik are the coefficients in equation (1) and (2) and n is the number of units (say, in dollars) of investment in i. The linear independence thus refers to the relationship in the k + 1 dimensional vector space.

4/ The size of the idiosyncratic terms was observed by Roll and Ross.

5/ See Kruskal [24] for a survey and discussion of the many variants of the basic method.

6/ In the multiperiod setting or the continuous time setting, we shall make enough assumptions so that the derived utility has the characteristics of risk aversion. . .See Fama [11], Merton [28], Cox, Ingersoll and Ross [5].

7/ An alternative proof of the CAPM is to note that the multivariate normal distribution assumption implies, via the above analysis, the market model (i.e., equation (1) with $k = 1$, $\delta_1 = r_m - E_m$), hence equation (6).

8/ Their derivation of their equation (3) did not need their equation (6). A weaker test of their (3) or (6) that avoids the collinearity problem that they faced would be to determine if there are only two factors.

9/ See also Cornell [3].

10/ A well known numerical disaster is solving the option pricing type differential equation with an algorithm that violates a von Neuman stability condition.

11/ Of some interest is Macaulay's duration as a measure of risk. See Macaulay [27], Hicks [18], and Ingersoll, Skelton and Weil [20], and Cox, Ingersoll and Ross [6].

12/ See, for example, Dahlquist and Björck [8], pp. 174-177, Forsythe and Moler [13], Isaacson and Keller [21].

13/ von Neuman and Goldstine [32].

14/ If the matrix is exactly diagonal, inverting the matrix involves only simple divisions and (11) does not apply. In general, if the matrix is nearly diagonal and symmetric, the condition number will be small after scaling.

15/ This construction was used by Marc Reinganum in a recent study of the APT [44]. A more extensive empirical study is currently under way by the author.

BIBLIOGRAPHY

1. Black and Scholes, 1978, "The Pricing of Options and Corporate Liabilities," Journal of Political Economy.

2. Brennan and Schwartz, 1978, "A Model for the Valuation of Savings Bonds and the Determination of the Optimal Redemption Strategy," manuscript, University of British Columbia.

3. Bradford Cornell, "The Consumption Based Asset Pricing Model: Potential Tests and Applications." Working Paper, UCLA.

4. Douglas T. Breeden, 1979, "An Intertemporal Asset Pricing Model with Stochastic consumption and Investment Opportunities," Journal of Financial Economics.

5. Cox, Ingersoll and Ross, 1978, "A Theory of the Term Structure of Interest Rates," manuscript, Graduate School of Business, Stanford University.

6. Cox, Ingersoll and Ross, 1979, "Duration and the Measurement of Basis Risk," Journal of Business.

7. Cox, Ross and Rubinstein, 1979, "Option Pricing: A Simplified Approach," Journal of Financial Economics.

8. Dahlquist and Björk, 1974, "Numerical Methods," Prentice-Hall.

9. Eugene Fama, 1977, "Risk Adjusted Discount Rates and Capital Budgeting under Uncertainty," Journal of Financial Economics.

10. Eugene Fama, 1968, "Risk, Return and Equilibrium: Some Clarifying Comments," Journal of Finance.

11. Eugene Fama, 1970, "Multiperiod Consumption - Investment Decisions," The American Economic Review.

12. Lawrence Fisher, 1966, "An Algorithm for Finding Exact Rates of Return," Journal of Business.

13. Forsythe and Moler, 1967, "Computer Solution of Linear Algebraic Systems," Prentice-Hall.

14. Galai and Masulis, 1976, "The Option Pricing Model and the Risk Factor of Stock," Journal of Financial Economics.

15. Adam Gehr, Jr., 1976, "Some Tests of the Arbitrage Pricing Theory," Journal of the Midwest Finance Association.

16. Robert Geske, 1979, "The Valuation of Compound Options," Journal of Financial Economics.

17. Nils H. Hakansson, 1971, "Capital Growth and the Mean-Variance Approach to Portfolio Selection," _Journal of Financial and Quantitative Analysis_.

18. J. Hicks, 1938, "Value and Capital," Oxford: Clarendon Press.

19. Gur Huberman, 1979, "Arbitrage Pricing Theory: A Simple Approach," Working Paper, Yale University.

20. Ingersoll, Skelton and Weil, 1978, "Duration Forty Years Later," Working Paper, University of Chicago.

21. Isaacson and Keller, 1966, "Analysis of Numerical Methods," New York: John Wiley and Sons.

22. M. Jensen, 1968, "The Performance of Mutual Funds in the Period 1945-1964," _Journal of Finance_.

23. Kraus and Litzenberger, 1976, "Skewness Preference and the Valuation of Risk Assets," _Journal of Finance_.

24. Kruskal, 1976, "Factor Analysis: Bilinear Methods," in International Encyclopedia of Statistics, pp. 307-329.

25. Edward Leamer, 1978, "Specification Searches," (John Wiley & Sons, 1978).

26. John Lintner, 1965, "The Valuation of Risk Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets," _Review of Economics and Statistics_.

27. F. Macauley, 1938, "Some Theoretical Problems Suggested by the Movements of Interest Rates, Bond Yields, and Stock Prices in the United States Since 1856," (New York: Columbia University Press).

28. Robert C. Merton, 1971, "Optimum Consumption and Portfolio Rules in a Continuous - Time Model," _Journal of Economic Theory_.

29. Robert C. Merton, 1973, "An Inter-Temporal Capital Asset Pricing Model," _Econometrica_.

30. Robert C. Merton, 1973, "Theory of Rational Option Pricing," _Bell Journal of Economics and Management Science_.

31. Miller and Scholes, 1972, "Rates of Return in Relation to Risk: A Re-examination of Some Recent Findings," in Michael Jensen, ed., Studies in the Theory of Capital Markets (New York: Praeger, 1972).

32. von Neuman and Goldstine, 1947, "Numerical inverting of matrices of high order." Bulletin America Mathematics Society.

33. Richard Roll, 1971, "Investment Diversification and Bond Maturity," Journal of Finance.

34. Richard Roll, 1977, "A Critique of the Asset Pricing Theory's Tests," Journal of Financial Economics.

35. Roll and Ross, 1979, "An Empirical Investigation of the Arbitrage Pricing Theory," Working Paper, UCLA.

36. Stephen A. Ross, 1973, "Return, Risk and Arbitrage," Working Paper, University of Pennsylvania.

37. Stephen A. Ross, 1976, "The Arbitrage Theory of Capital Asset Pricing," Journal of Economic Theory.

38. Stephen A. Ross, 1978, "Mutual Fund Separation in Financial Theory -- The Separating Distributions," Journal of Eocnomic Theory.

39. Rothschild and Stiglitz, 1970, "Increasing Risk I. A Definition," Journal of Economic Theory.

40. Mark Rubinstein, 1973, "A Mean-Variance Synthesis of Corporate Financial Theory," Journal of Finance.

41. Mark Rubinstein, 1976, "The Valuation of Uncertain Income Streams and the Pricing of Options," Bell Journal of Economics.

42. William F. Sharpe, 1964, "Capital Asset Prices: A Theory of Market Equilibrium Under Conditions of Risk," Journal of Finance.

43. Oldrich Vasicek, 1977, "An Equilibrium Characterization of the Term Structure," Journal of Financial Economics.

44. Marc Reinganum, 1980, "A Simple Test of the Arbitrage Pricing Theory," Working Paper, University of Southern California.