NBER TECHNICAL PAPER SERIES

ASSET PRICING THEORIES

Michael Rothschild

Technical Working Paper No. 44

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
March 1985

I am grateful to Gary Chamberlain and Bernd Luedecke for many helpful conversations and to the National Science Foundation for research support. A preliminary version of this paper was presented as an invited address at the August 1984 meetings of the Econometric Society in Sidney, Australia. The research reported here is part of the NBER's research program in Financial Markets and Monetary Economics. Any opinions expressed are those of the author and not those of the National Bureau of Economic Research.

NBER Technical Paper #44
March 1985

Asset Pricing Theories

ABSTRACT

This article compares two leading models of asset pricing: the capital asset pricing model (CAPM) and the arbitrage pricing theory (APT). I argue that while the APT is compatible with the data available for testing theories of asset pricing, the CAPM is not. In reaching this conclusion emphasis is placed on the distinction between the unconditional (relatively incomplete) information which econometricians must use to estimate asset pricing models and the conditional (complete) information which investors use in making the portfolio decisions which determine asset prices.

Empirical work to date suggests that it is unlikely that the APT will produce a simple equation which explains differences in risk premium well with a few parameters. If the CAPM were correct, it would provide such an equation.

Michael Rothschild
Department of Economics, D-008
University of California
at San Diego
La Jolla, CA 92093

1. The Asset Pricing Problem.

A central problem for the analysis of private economies is determining the prices of commodities which are risky. General equilibrium theory provides a complete but distressingly general answer to these questions. Arrow-Debreu theory, and even when modified to deal with imperfect and heterogeneous information, can do little more than state conditions for the existence of equilibrium with various measures of efficiency. Not surprisingly, it is difficult to think of any empirical work which tests or otherwise exposes this theory to data. In sharp contrast, modern theories of asset pricing offer very sharp predictions about relative asset prices; these theories have spawned hundreds of empirical papers which purport to confirm or reject various aspects of the theory. In this paper I will try to offer a critical perspective on two leading theories of asset pricing -- the Capital Asset Pricing Model (henceforth CAPM[1]) -- and the Arbitrage Pricing Theory (APT) due to Steven Ross (1976). I shall largely focus on the theoretical properties of the models. Indeed one major purpose is demonstrate a simple analytical structure which makes developing and comparing the two theories straightforward. However, since this is an empirical subject, much of what I have to say will be directed toward empirical work. On occasion I will, forsaking comparative advantage, comment on things empirical.

1.1 The Context

The question which this theory attempts to answer may be stated simply: Given a bunch of random variables, what determines their prices? In the simplest case, and the only one discussed in this paper, the random variables are being traded this period for money and yield random payoffs next period[2]. A natural choice of units for this problem would be expected value. We could specify z_i as the amount of asset i needed to provide an expected next period payoff of one. The literature generally chooses a different normalization. Following it we specify that x_i is an asset which can be bought for one dollar. This specification is strange to an equilibrium theorist for it assumes that prices have been determined already. In most of economics one starts out with preferences, technology and supply and asks whether or not equilibrium, and thus equilibrium prices, can exist. In asset pricing theory, for the most part, one starts with a set of equilibrium prices and asks what conditions they must satisfy.

Let

1) $E(x_i) = \mu_i$

be the mean return on asset i. If there is a riskless asset, of which more later, and it has a rate of return of ρ , (that is for \$1 one can purchase enough of the riskless asset to ensure a return of \$$\rho$ next period) the **risk premium** for asset i is $\mu_i - \rho$. The aim of asset pricing theories is to explain why different assets have different risk premia.

1.2 Internal Theories

Given that the problem is to explain what determines mean returns or risk premia, we can distinguish between two kinds of theories. **Internal** theories explain expected returns in terms of the characteristics of the probability distribution of returns on all assets. For the most part internal theories are second moment theories. That is they seek to explain means in terms of

1. The CAPM is due to Treynor (1961), Sharpe (1964), Lintner (1965), and Mossin (1966).
2. I deal exclusively with one period models in this paper.

variances and covariances. The CAPM is a theory of this sort. It's pricing equation is

2) $\mu_i = \rho + \beta_i (\mu_M - \rho)$

where p_M is the market portfolio (normalized to cost one dollar), $\beta_i = cov(x_i, p_M)$, and $\mu_M = E(p_M)$.

1.3 External Theories

In contrast **external** theories, use information beyond the distribution of returns to explain asset prices. A good example of such a theory is the paper by Chen, Roll, and Ross (1983) which uses macroeconomic variables to explain asset prices. Of course, internal and external theories are not exclusive alternatives. One of the virtues of a tight internal theory is that it specifies the way in which external forces influence asset prices. If a macroeconomic variable is to influence the structure of asset prices, and if the CAPM holds, then equation (2) states how that variable will affect asset prices.

1.4 Notation

At this point it is appropriate to introduce some notation and assumptions. There is a large number of assets, potentially a countable infinity. As mentioned x_i is the return from a \$1 investment in asset i. We assume that means and variances are uniformly bounded and denoted by

3) $\mu_i = E(x_i), \quad \sigma_{ii} = V(x_i), \quad \sigma_{ij} = cov(x_i, x_j).$

We suppose that the assets are arranged in order and we shall often have occasion to refer to the vector of the means of the first N assets which we denote μ_N. The corresponding variance-covariance matrix is Σ_N. We assume that Σ_N is nonsingular for all N. The matrix of second moments is

4) $\Gamma_N = \Sigma_N + \mu_N \mu'_N$.

Throughout this paper we will be concerned with linear combinations of assets, which we will call portfolios. In parts of the sequel, particularly those dealing with conditional inference, it will be useful to maintain the pedantic distinction between a portfolio which is a random variable and the distribution of that random variable. That is, included in the definition of the random variable x_i are the events in which they give payoffs of various amounts. The probability of these events cannot be specified without specifying a probability distribution. This distinction is important because, when we deal with conditional inference we shall be concerned with the relationship of different probability assessments of the same events. Thus if $p = \Sigma_i \alpha_i x_i$ and $q = \Sigma_i \gamma_i x_i$ are portfolios and if we specify first and second moments of returns as in (3) and (4) we can write

5) $cov(p, q) = \alpha' \Sigma \gamma$

and

6) $E(pq) = \alpha' \Gamma \gamma$.

2. Empirical setting

The goal of asset pricing theory is to explain expected returns, μ_N, possibly using the second moments of returns, Σ_N. Both μ_N and Σ_N are expectations. They cannot be directly observed. For testing and verification empirical counterparts of μ_N and Σ_N are required. In the U.S. rich data on **actual** returns are available which can be used to estimate expected returns. Despite the quality and volume of data, the estimates are not very precise. However, for empirical work, actual return data seems more promising than any alternative[3]. It is worth spending some time describing the characteristics of returns data. The theories we will discuss must be evaluated, at least in part, in terms of their ability to use, and be validated by, this data.

2.1 The CRSP data

The Center for Research on Security Prices (CRSP) of the University of Chicago has on tape the daily returns from holding any stock listed on the New York or American Stock Exchanges from 1962 through 1983. Monthly returns are available for stocks traded on the New York Stock Exchange from 1925 to 1983. Two characteristics of this data are noteworthy. First, a large number of assets is included. There are over four thousand assets in the daily returns file and almost three thousand assets in the monthly return file. This data is appropriate for theories which assume a large number of assets.

Secondly, it is not **comprehensive**. Even though the CRSP files have information on a huge number of assets, many investors hold assets in their portfolios which are not included in the CRSP data. While CRSP does collect data on the returns from bonds which could be used together with the stock returns, comparable data for many marketed assets do not exist. Commodities, foreign exchange, options, futures contracts, and stocks and bonds traded outside the U.S. are examples. Furthermore, many important assets are not traded on organized exchanges. The classic examples are human capital[4] and real estate. A theory which explains relative asset prices on the assumption that investors portfolios consist exclusively of the assets studied, cannot be rejected using the CRSP data alone. Given this, it is not clear in what sense the CRSP data can verify such a theory. The CAPM is a theory of this sort.

---

3. Not all researchers have restricted themselves to actual returns data. Cragg and Malkiel (1982) attempted to collect data on expectations which was sufficiently rich that it could be used to examine the determination of assets' prices. They collected nine years of forecasts (by financial institutions) of growth and earning for 260 of the companies whose stocks are traded on the New York Stock Exchange. While rich, this data is quite meager compared to the CRSP data on actual returns (described in the next subsection). The Cragg and Malkiel data is annual while the CRSP data is daily or monthly. Cragg and Malkiel cover less than a tenth of the number of firms included in the CRSP files. Furthermore its not clear that the financial institutions which provided the raw data to Cragg and Malkiel would be willing to do so again. One of Cragg and Malkiel's findings was that analysts' forecasts didn't contain much information.

While expectations data is unlikely to be used in preference to data on actual returns, it does reveal some interesting things. Cragg and Malkiel convincingly demonstrated that expectations are not homogeneous. Indeed they found that one of the best measures of risk for the stock of a firm was the extent to which forecasts of the firm's growth differed.

4. The current U. S. administration is working hard to improve the market for human capital. However, one doubts that CRSP will soon have good data on the return to human capital. It is unfortunate that the period when data on the return to liquid human capital was relatively good (Fogel and Engerman 1974) was a period when asset markets were relatively undeveloped.

2.2 Estimating μ_N and Σ_N using the CRSP data.

The most straightforward way to estimate expected returns using CRSP data is to form the sample mean. That is, given observations on the random returns on N assets for T periods, calculate

$$7) \quad \mu_N = \sum_{t=1}^{T} \tilde{x}_{Nt} .$$

This procedure is justified by some assumption like

$$8) \quad \tilde{x}_{Nt} = \mu_N + \tilde{\omega}_{Nt}$$

where the

9) $\tilde{\omega}_{Nt}$ are i.i.d. random variables.

There are, at least, two problems with this approach. First, the assumption that the $\tilde{x}_t$ have a distribution which is constant over time seems contradicted by the data. Stock returns behave differently in different periods. For, example, Luedecke (1984) reports that during the period 1971 to 1974 the daily mean return on 392 stocks was -.00015% while during the period 1975 to 1978 the mean return on the same 392 stocks was .0009%. The first figure corresponds to an annual loss of about 5%; the second to an annual gain of about 30%. It is hard to believe that investors expected the same returns in 1973 as they did in 1977. Thus, it is common practice for investigators using the CRSP daily data to split the approximately twenty years of data into a number of subperiods each no longer than six years in length.

Second actual returns are quite volatile. As a result it is hard to estimate μ_N in (7) precisely using a short time series. The many observations which CRSP daily data files contain do not alleviate this problem. The accuracy of the estimate of the mean is determined by the length of time which the process is observed, and not by the frequency of the observations. Most models of stock returns suppose that the returns process is some kind of (possibly geometric) Brownian motion. Suppose that x_t evolves according to

$$10) \quad dx = \alpha dt + \sigma dW$$

where W is a Weiner process. Consider trying to estimate the parameter α from observations of the process over the interval [0,T]. The accuracy of the estimate of α is a function of T and is independent of how often x_t is observed. Suppose that we have observations of x_t at $t = 0, 1/n, 2/n, \ldots, Tn/n$. Then $y_i = x_i - x_{i-1}$ is a normal random variable with mean α/n and variance σ^2/n. The sample information may as well be considered as consisting of the observations y_i, $i = 1, \ldots, nT$. A sufficient statistic for the mean is then $\Sigma_i y_i = x_T - x_0$. Thus the accuracy of the estimate of the mean can depend only on T and not on n.

In contrast, more frequent observations do permit more accurate estimates of the variance term σ^2. The variance of most estimates of σ^2 are of order $1/nT$; frequent observations of the x_t process let one see how much it fluctuates.

Both of these results carry over to the multivariate case. Accuracy of estimates of the mean vector μ are not improved by more frequent sampling while accuracy of estimates of the variance-covariance matrix Σ is of the order of $1/nT$; the variance of each of the terms of Σ is of order $1/nT$. (Anderson 1958 page 161). The fact that variances can be well estimated while means cannot is one of the great virtues of the Black-Scholes option pricing theory. It requires as data, and makes predictions using, variances of stock prices. It does not use information about expected returns.

2.3 Conditional and Unconditional Expectations

When combined with the fact that it is hard to estimate expected returns over intervals sufficiently short to make the assumption of constancy plausible, the probability that the distribution of asset returns is not constant over time suggests an alternative model. Suppose that instead of (8) and (9) we assume that there is a state variable s such that (x_{Nt}, s_t) is jointly distributed. Conditional on s, x_{Nt} is independent of t. That is the distribution of $(x_{Nt} \mid \tilde{s}_t = s)$ is independent of t. Suppose also that s_t has a stationary distribution.

This specification invites many interpretations. For example, different values of s could represent different information available to different traders; alternatively s could be an index representing general economic conditions in different time periods. In either case investors have information about the conditional distribution of returns, $(x_{Nt} \mid s)$. They use this information when making investment decisions. In so far as their choices depend on the first two moments of returns the portfolios they choose to hold will be functions of

11) $\mu_{Ns} = E(x_N \mid \tilde{s} = s)$

and

12) $\Sigma_{Ns} = E[(x_N - \mu_{Ns})(x_N - \mu_{Ns})' \mid \tilde{s} = s]$.

If the variable s is unobservable, the analyst cannot hope to estimate μ_{Ns} or Σ_{Ns}. However, if s has a stationary distribution he can estimate unconditional first and second moments. Let $\Gamma_{Ns} = \Sigma_{Ns} + \mu_{Ns}\mu'_{Ns}$ Then the analyst can estimate

13) $\overline{\mu}_N = E(\mu_{Ns})$, $\overline{\Sigma}_N = E(\Sigma_{Ns}) + V(\mu_{Ns})$ and $\overline{\Gamma}_N = E(\Gamma_{Ns})$

using (7) for $\overline{\mu}_N$ and an analogous procedure for $\overline{\Sigma}$ and $\overline{\Gamma}_N$. Suppose for simplicity that there are a finite number (S) of states and that the stationary or ergodic distribution of s is given by

14) $\pi_s = Pr\{s_t = s\}$.

Then unconditional means and covariances are

15) $\overline{\mu} = \Sigma_s \pi_s \mu_{Ns}$

and

16) $E_N(\Sigma_{Ns}) = \Sigma_s \pi_s \Sigma_{Ns}$.

Both $\overline{\mu}_N$ and $\overline{\Sigma}_N$ can be estimated from data on actual returns. With this specification, it is legitimate to use the entire time series of observations to estimate expected returns. However, the mere fact that unconditional expectations can be estimated doesn't mean that they are of any interest. For them to be useful, it is necessary to have a theory which makes predictions about unconditional expectations. As we will see, it is not clear that the CAPM does this.

2.4 Summary

The data most likely to be used for testing asset pricing theories is very rich. However, it is not perfect. In particular it is not comprehensive and it is unlikely that the asset pricing process is stationary over time periods short enough to estimate expected returns with much accuracy. It may be possible to estimate unconditional expectations accurately.

3. Mathematical Setting [5]

We will be considering many assets; we need a mathematical structure which allows us to let $N \to \infty$. Let $F_N = [x_1,...,x_N]$, the linear subspace consisting of all linear combinations[6] of (portfolios formed from) the assets $x_1,...,x_N$. Now let $F = \bigcup_{N=1}^{\infty} F_N$, the space of all portfolios. Unfortunately F is not large enough for our purposes. Some of the arguments in asset pricing theory involve diversification; they require that we be able to make sense out of things like the limit of the sequence of portfolios p_n where

17) $$p_n = \sum_{i=1}^{n} \frac{1}{n} x_i .$$

Each p_n costs a dollar; if asset returns are uncorrelated then in the limit p_n converges to a riskless asset. We need a mathematical structure which gives meaning to $\lim_{n \to \infty} p_n$. For this two things are needed: a space of investments which includes things which are not in F and a way of measuring distance between portfolios which allows us to give meaning to the convergence of portfolios like those defined in (17). Then we can frame a theory which deals with portfolios and the limits of portfolios. One way to do this is to suppose that asset returns are defined on some underlying probability space. Let $L_2(P)$ be the set of all random variables defined on that space with finite variances. We measure the length of a portfolio (or its distance from another portfolio) as $||p||$ where

18) $$||p|| = E(p^2)^{1/2}.$$

Associated with the mean square norm is an inner product denoted (p,q) and defined as

19) $$(p,q) = E(pq).$$

Note that

20) $$||p||^2 = (p,p).$$

If $p = \sum_{i=1}^{N} \alpha_i x_i$ and $q = \sum_{i=1}^{N} \gamma_i x_i$ then

21) $$||p||^2 = \alpha' \Gamma_N \alpha \text{ and } (p,q) = \alpha' \Gamma_N \gamma.$$

Analysis now takes place in $\bar{F}$, the closure of F. Things in $\bar{F}$ but not in F are **limit portfolios**; in the sequel we use the term portfolio to refer to all objects in $\bar{F}$ whether or not they belong to some F_N . $\bar{F}$ is a Hilbert space.

5. This section is taken from Chamberlain and Rothschild (1983). Many details are omitted.
6. In this paper, linear combinations are always finite linear combinations.

If p and q are portfolios and if $(p,q) = 0$, then we say that p is orthogonal to q and write $p \perp q$. If L is a linear subspace, then $L^{\perp} = \{ p \mid p \perp q$ for all $q \in L \}$ is its orthogonal complement. Many of the arguments we use are based on the projection theorem which states that if L is a closed subspace in $\overline{F}$, then every $p \in \overline{F}$ has a unique decomposition as $p = p_1 + p_2$ where $p_1 \in L$ and $p_2 \in L^{\perp}$. Note that p_1 is the point in L which is closest to p ; p_1 is the **projection** of p onto L ; $\|p_2\|$ is the distance from p to L is $\|p_2\|$; in the sequel we will let $\|p - L\|$ denote the distance p to the subspace L. Thus, $\|p - [b_1, \ldots, b_N]\|$ is the distance from p to the space spanned by the vectors $b_1, \ldots, b_N$.

3.1 The mean and cost functionals

Our analysis of asset pricing models will use some linear functionals[7] defined on the space of portfolios. These are the mean functional and the cost functional. If $p = \sum_{i=1}^{N} \alpha_i x_i$, then its expected value is $E(p) = \sum_{i=1}^{N} \alpha_i \mu_i$. It is easy to calculate that there is a portfolio m_N such that

22) $E(p) = (m_N, p)$ for any $p \in F_N$,

where

23) $m_N = \Gamma_N^{-1} \mu'_N x_N$.

The cost of any portfolio, $C(p) = \Sigma_i \alpha_i$, is also a linear functional. Again it is easy to calculate that

24) $C(p) = (c_N, p)$ for *any* $p \in F_N$

where

25) $c_N = \Gamma_N^{-1} e'_N x_N$.

and $e_N = (i, \ldots, i)$ is a vector of N ones.

When an equation like (22) holds, we say that m_N **represents** the linear functional $E(\bullet)$. A linear functional $L(\bullet)$ is **continuous** if $\|p_n\| \rightarrow 0$ implies $L(p_n) \rightarrow 0$. A basic mathematical result, the Reisz representation theorem, states that if L is a continuous linear functional on a Hilbert space H, then there is an element q in H that represents L in the sense that $L(p) = (p, h)$ for all $p \in L$. It is easy to see from its definition that the mean functional is continuous on $\overline{F}$; since $\|p\| = E(p)^2 + V(p)$ if $\|p_n\| \rightarrow 0$ $E(p_n) \rightarrow 0$. Thus, there is an $m \in \overline{F}$ such that

26) $E(p) = (m, p)$ for *all* $p \in \overline{F}$.

Furthermore, if m_N represents $E(\bullet)$ on F_N then $m_N \rightarrow m$.

7. A linear functional is a real valued linear mapping.

3.1.1 Arbitrage and the continuity of the cost functional.

Without further assumptions, the cost functional is not continuous. However, if we assume that the set of portfolios does not permit arbitrage opportunities then it is easy to show that the cost functional must be continuous. Let $\{p_N\}$ be a sequence of finite portfolios. Then we say the market permits no arbitrage opportunities if the following two conditions hold:

A.i) If $V(p_N) \rightarrow 0$ then $E(p_N) \rightarrow 0$.

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

Condition (A.i) simply states that it is not possible to make an investment that is costless, riskless and yields a positive return. Ross (1976) has shown that if (A.i) fails many (but not all) risk averse investors will want to take infinitely large positions. This is, of course, incompatible with equilibrium. A similar argument justifies (A.ii). Suppose that (A.ii) does not hold; that is suppose that the market allows investors to trade a portfolio that, approximately, costs a dollar and has a riskless, nonpositive return. Then investors face no budget constraints; by selling this portfolio short they can generate arbitrarily large amounts of cash while incurring no future obligations. This too is incompatible with equilibrium.

It is straightforward to show (see Chamberlain and Rothschild 1983) that (A.ii) implies that the cost functional is continuous. Thus, there is a portfolio c which represents $C(\bullet)$ in the sense that

27) $C(p) = (c, p)$ for all $p \in \overline{F}$.

Furthermore if c_N represents $C(\bullet)$ on F_N then $c_N \rightarrow c$.

One important implication of these conditions concerns the trade-off between risk and return which the market permits. Define

28) $\delta = \sup |E(p)| / V^{1/2}(p)$ subject to $p \in \overline{F}$, $C(p) = 0$ and $p \neq 0$.

Chamberlain and Rothschild (1983) have shown that if (A.i) holds then δ is finite.

3.2 Riskless Assets

We assume the Σ_N is nonsingular. Thus if $p = \sum_{i=1}^{N} \alpha_i x_i = \alpha' x_N$ is a portfolio in F_N, $V(p) = \alpha' \Sigma_N \alpha > 0$ if $\alpha \neq 0$. If p is a portfolio in $\overline{F}$ this conclusion need not hold. If there is a sequence of portfolios $\{p_N\}$ such that

29) $p_N \rightarrow p, C(p_N) \rightarrow 0$,

then we say that p is a **riskless asset**. If there is a riskless asset, the market permits investors to diversify away all risk. Chamberlain and Rothschild give necessary and sufficient conditions on the sequence Σ_N for there to be a riskless asset. It is straightforward to show that their condition implies that there is a riskless asset if and only if the portfolio p_N defined in (17) converges to it; i.e. if $\dfrac{e'_N \Sigma_N e_N}{N^2} \rightarrow 0$.

4. The Capital Asset Pricing Model

The CAPM is an asset pricing model which captures the notion that an asset's risk premium is determined by its diversifiable risk. In the CAPM, an asset's diversifiable risk is measured by its covariance with the market and is called "beta". We will define beta precisely below. The CAPM rests on the premise that all investors choose mean-variance efficient portfolios. That is, faced with the problem of selecting portfolios from $\bar{F}$, all investor will choose portfolios which are for some parameters γ and δ solutions to the problem:

(Q) Choose p to minimize $V(p)$ subject to $C(p) = \gamma$ and $E(p) = \delta$.

It follows from the definition of the mean square norm $\| \ \|$ and the portfolios c and m which represent the cost and mean functionals that problem **(Q)** is equivalent to problem

(P) Min $\|p\|$ subject to $(c, p) = \gamma$ and $(m, p) = \mu$.

Since the assumption that investors choose mean-variance efficient portfolios drives the CAPM it is natural to ask what justifies it. The answer is: very strong assumptions. Together, two conditions are sufficient. While if either of these conditions is not true, an investor could still choose a mean-variance efficient portfolio, he is extremely unlikely to do so. The first condition is that assets in $\bar{F}$ are comprehensive, that they comprise the entire universe from which investors choose investments. We have already indicated that this is unlikely to be true if $\bar{F}$ is portfolios formed from assets for which returns data is available. The second condition is that investors evaluate risky returns from investments in terms of their first two moments. This requires either quadratic utility functions or very special assumptions about the process generating returns. As $N \rightarrow \infty$ a kind of conditional normality is required. For finite N, the weaker but still strong assumption of spherical symmetry if required (Chamberlain 1983a).

4.1 The Market Portfolio Belongs to $[m,c]$

We now derive the basic CAPM pricing equation. Let $[m, c]$ denote the space spanned by m and c. Note that if m and c are collinear ($m = \lambda c$), all portfolios of a given cost will have a mean entirely determined by that cost. Problem **(P)** will not have a solution for arbitrary γ and δ; the market does not exhibit a well defined trade off between mean and variance. For this reason we assume

30) dimension $[m, c] = 2$.

The following proposition is basic to the CAPM.

Proposition 1: *Any solution to (P) belongs to [m,c]*

Proof: Suppose a portfolio p solves **(P)** but that p does not belong to $[m, c]$. Let q be the projection of p onto $[m, c]$. Then $E(q) = E(p)$ and $C(q) = C(p)$. However, since $\|p\|^2 = \|p - q\|^2 + \|q\|^2$ and $p - q \neq 0$, $\|p\| > \|q\|$. Thus, $V(p) > V(q)$ and p cannot be a solution to **(P)**.

Define the market portfolio, denoted p_M, as the portfolio (normalized to have unit cost) formed by summing the investments of all investors. Since every investor chooses mean-variance efficient portfolios (portfolios which belong to $[m, c]$, p_M is linear combination of portfolios which belong to $[m, c]$. Thus, for some coefficients α and γ,

31) $p_M = \alpha m + \gamma c.$

4.2 The CAPM pricing equation

We will now show that the CAPM pricing equation is a consequence of equation (31). For p be any portfolio define $\beta_p = cov(p, p_M)$. The CAPM pricing equation states that there exist constants a and b such that for any portfolio q,

32) $E(q) = a + b\beta_q$.

Proposition 2: *If* $p_M \in [m, c]$ *and there exist two unit cost portfolios p and q such that* $\beta_p \neq \beta_q$ *then (32) holds .*

This is simple arithmetic; (31) implies

$$(p_M, q) = cov(p_M, q) + E(m)E(q) = \alpha(m, q) + \gamma C(q)$$

or

33) $\beta_q = cov(p_M, q) = E(q)(\alpha - E(p_M) + \gamma$.

Now if $\alpha = E(p_M)$, (33) implies $\beta_q = \gamma C(q)$ so all unit cost portfolios have the same beta; thus we can assume $\alpha \neq E(p_M)$ and divide to get (32) with $a = \gamma/(\alpha - E(p_M))$ and $b = (\alpha - E(p_M))^{-1}$.

We now prove a converse of Proposition 2.

Proposition 3: *Equation (32) implies* p_M *belongs to [m,c] .*

Proof: Let q be a portfolio which is orthogonal to $[m, c]$. It will suffice to show that q is orthogonal to p_M. Since q is orthogonal to $[m, c]$, $E(q) = (q, m) = 0$ and $C(q) = (c, q) = 0$. Thus (34) implies $\beta_q = 0$. However, $(q, p_M) = E(q)E(p_M) + \beta_q = 0$.

The coefficients a and b in the CAPM pricing equation (32) have interesting interpretations. Suppose there is a riskless asset. Let q be this riskless asset in (32) and conclude that $a = \rho$. Now let q be p_M and conclude that $b = \dfrac{E(p_M) - \rho}{V(p_M)}$.

Thus we may write (32) as

34) $E(q) = C(q) + \dfrac{E(p_M) - \rho}{V(p_M)} \beta_q$.

If there is no riskless asset (34) still holds when ρ is interpreted as the rate of return on a "zero beta" portfolio, that is a portfolio which is uncorrelated with the market.

4.3 The CAPM in a conditional setting

The CAPM does not easily accommodate the conditional model of returns described above. To see this suppose that there are a finite number, N, of assets and, for concreteness, consider the version of the model in which at time t all investors' expectations are determined by the state variable $\tilde{s}_t$. In this subsection we will omit the subscript N. To compensate for this notational laxity we will be especially pedantic about the distinction between portfolios (which are random variables) and their distributions. Portfolios are simply linear combinations of the assets, $x_1, \ldots, x_N$. As such they are independent of beliefs. The distribution of these random variables is determined by beliefs. For example the vector $e = (1,...,1)$ gives the cost of all assets under any set of beliefs. If beliefs about second moments of returns are represented by

Γ_s, then the portfolio c_s represents the cost functional under these beliefs where $c_s = \gamma'_s x$ and $\gamma_s = \Gamma_s^{-1} e$. Simillary, the vector μ_s is the vector of mean returns under beliefs s, the portfolio m_s represents the mean functional under these beliefs and $m_s = \alpha'_s x$ where the portfolio weights α_s are given by $\alpha_s = \Gamma_s^{-1} \mu_s$.

If the CAPM holds conditionally then for each s there exist constants a_s and b_s such that if p is a portfolio of unit cost

35) $\quad E(p \mid s) = a_s + b_s(cov(p_M, p) \mid s)$.

Does this imply that the CAPM equation (32) holds? Clearly not in general. Suppose that $b_s = b$ for all s, then taking expectations in (35) we see that

$$E(p) = a + b\, E(cov(p_M, p) \mid s).$$

But,

$$cov(p_M, p) = E(cov(p_M, p) \mid s) + cov\Big[E(p_M \mid s), E(p \mid s)\Big].$$

Thus, we see that (32) holds only if

36) $\quad cov\Big[(E(p_M \mid s), E(p \mid s))\Big] = 0.$

If, as will generally be the case, b_s is not a constant but a random variable which fluctuates with s, things are worse. Now we require also that

37) $\quad cov\Big[b_s, cov((p, p_M) \mid s)\Big] = 0.$

Neither (36) or (37) seems plausible. The CAPM requires that each equation hold for **every** p. Thus (36) requires that the expected market return be uncorrelated with the expected return on every portfolio. This seems unlikely. We conclude that even if the CAPM holds for conditional data, the unconditional data available to the analyst will not be consistent with the CAPM pricing equation.

The preceeding discussion of the relationship between the conditional CAPM and unconditional CAPM focused on the validity of the CAPM pricing equation (32). A slightly different argument, with the same negative conclusion focuses on the question of whether or not the fact that the market portfolio is mean variance efficient in every state implies that the unconditional market portfolio is mean variance efficient. Intuition suggests that if the market portfolio is mean variance efficient in every state than it must be unconditionally mean variance efficient. Unfortunately, it is simply not true that a portfolio which is mean variance efficient in every state is unconditionally mean variance efficient. For a simple counterexample consider the two hedge portfolios with the following state dependent means and variances:

$$E(p \mid 1) = V(p \mid 1) = 1;\ E(p \mid 2) = V(p \mid 2) = 2$$

and

$$E(q \mid 1) = E(q \mid 2) = 1;\ V(q \mid 1) = V(q \mid 2) = 1.1.$$

For hedge portfolios mean variance efficiency is determined by the ratio of mean variance. Thus in each state p is mean variance efficient relative to q. However, $E(p) = 1.5$ and $V(p) = 1.5 + .25$ so that $E(p)/V(p) = 1.5/1.75 < E(q)/V(q) = 1/1.1$.

Unconditionally q is mean variance efficient relative to p. If $E(p)$ did not vary with s, then conditional mean variance efficiency would imply unconditional mean variance efficiency. Similarly inspection of (36) reveals that the CAPM will hold and the market portfolio will be mean variance efficient if the conditional mean of the market portfolio is the same in every state. This is a stringent requirement.

Up to this point I have ignored the problem that the market portfolio may change as the state s changes. As the proportion of different assets in the market portfolio is determined by supply and demand conditions which vary from period to period, the composition of the market portfolio will change as the state changes. Here it is not entirely clear what the statement that the market portfolio is mean variance efficient means. One possible interpretation is that the average or expected market portfolio is (unconditionally) mean variance efficient. This is unlikely to be the case. We can write the market portfolio as $p_{Ms} = \delta'_s x$ where the portfolio weights δ_s are given by

38) $\quad \delta_s = f_s \alpha_s + g_s \gamma_s$.

In (38) f_s and g_s are numbers and α_s and γ_s are vectors of portfolio weights which represent the conditional mean and cost functionals. These weights are random variables with long run average or ergodic values given by

39) $\quad \delta = \Sigma_s \pi_s \delta_s = \Sigma_s \pi_s (f_s \alpha_s + g_s \gamma_s)$.

We can then identify the long run or ergodic market portfolio, $\overline{p}_m$ as the portfolio with these weights. One possible version of an unconditional CAPM would be the requirement that this portfolio be mean-variance efficient, that is that

40) $\quad \overline{p}_m \in [\overline{m}, \overline{c}]$.

We now show that the conditional CAPM (equation (38)) does not imply this version of the unconditional CAPM.

Note that (40) defines δ as a linear function of $f = (f_1, \ldots, f_S)$ and $g = (g_1, \ldots, g_S)$. Thus we may write (39) as

41) $\quad \delta = L\begin{pmatrix} f \\ g \end{pmatrix}$

where L is an N by $2S$ matrix with typical entry

$$L_{ns} = \begin{cases} \alpha_{ns} \pi_s & \text{for } s \leqslant S \\ \\ \gamma_{ns-S} \ \pi_{s-S} & \text{for } s > S. \end{cases}$$

Theory in no way restricts the vectors f and g in (41). Thus, (38) can hold only if rank L is 2 or less. But theory also provides no reason to expect that L is so restricted. Suppose for example that all states are equally likely so that $\pi_s = S^{-1}$ for all s. Then for L to be of rank two it is necessary that dimension $[\Gamma_1^{-1} e, \ldots, \Gamma_s^{-1} e] \leqslant 2$. I see no reason to expect this to happen.

A similar argument holds if s is interpreted as the information of different investors. Under the assumptions of the CAPM each investor chooses a portfolio which is, according to his beliefs, mean-variance efficient. That is, his portfolio is in $[m_s, c_s]$ if his beliefs are indexed by s. The market portfolio is a linear combination of these portfolios with arbitrary weights. (The weights depend on the relationship between wealth and information.) The (unconditional) CAPM will only hold if the market portfolio lies in a two dimensional subspace. Without further restrictions, arbitrary linear combinations of vectors from several different subspaces will not belong to a single two dimensional subspace.

4.4 Test of the CAPM

Tests of the CAPM exploit the fact that the second moments of the expected return for all assets and for the market portfolio should explain risk premia completely. Many tests of the CAPM have been made. Whether one assesses the results of these tests as evidence for or against the CAPM depends to a great extent on how one views the relationship between theory and data in economics. Those with a strict or pure theoretical point of view tend to argue that the available evidence rejects or at least does not confirm the theory. The argument runs roughly as follows: The CAPM makes strong assumptions and produces strong predictions. The theory is either (i) not true because some of its implications are obviously false or (ii) has not been seriously tested because the assumptions on which the CAPM rests have not been satisfied (and probably cannot be satisfied) in any empirical test.

The first point is based on the observation that the CAPM is a complete theory. It does not simply predict that the coefficients in regression tests of (32) should be significant and have the right sign. The CAPM provides a complete explanation of risk premia. Except for measurement error, R^2 should be one. Variable other than beta should not help to explain risk premia unless they are in some way plausibly related to measurement error. This is true whether the data are external or internal but the finding that other internal data explain risk premia is particularly damming evidence against the CAPM. Since there is a lot of evidence of this kind, (we will discuss some below) the CAPM should be rejected.

Roll (1977) has stressed point (ii). Since the CAPM is logically equivalent to the statement that the market portfolio is mean-variance efficient, it cannot be tested unless the market portfolio is specified and observed. The typical test of the CAPM use some sort of index of returns on common stocks (the Standard and Poor 500 or a value weighted portfolio of the stocks traded on the New York Stock Exchange) as the market portfolio. These indices are not the market portfolio. All one can conclude from such a test is that the particular proxy for the market portfolio is, or is not, mean-variance efficient. One cannot conclude that the CAPM is, or is not, valid. This criticism must be somewhat muted by Stambaugh's (1982) work. Stambaugh constructed monthly series of values and returns on real estate, consumer durables, and various kinds of bonds -- the assets which together with equities comprise most of non-human wealth. From this data he constructed a number of different "market" portfolios. Then he conducted sophisticated tests of the CAPM in order to see whether the inferences one would draw about the validity of the CAPM depended crucially on the composition of the market portfolio. He found that they did not. (He also found that the question: "Does the data support the CAPM?" did not seem to have a clear cut yes or no answer; however, the ambiguity had little to do with the composition of the market portfolio).

A more casual and pragmatic view of the role of theory leads to the assertion that test of the CAPM support the theory. The argument is, roughly, that theory is never literally true. All asset pricing theories can hope to do is to suggest variables that might appear in equations explaining risk premia. The CAPM suggests that beta should appear in such regression. From this point of view a successful test is a finding that beta is significant and has the right sign. That beta should not explain risk premia completely is only to be expected. The world is more complicated than the assumptions of the CAPM make it out to be. Since the assumptions of the CAPM are not literally true, we cannot expect the CAPM equation (32) to provide a complete explanation of risk premia. Finding that other variable are significant is just an indication that the CAPM is not completely right. No one would expect it to be, but it is an important first step in the explanation of the structure of asset prices. Unless and until someone comes along with a theory which both has sharper empirical implications and better incorporates the intuitive notion that the market rewards only nondiversifiable risk, we should accept the CAPM.

5. The Arbitrage Pricing Theory

The arbitrage pricing theory rests on three assumptions. The first is that the market should not permit arbitrage opportunities; in our terms, (A) holds. The second assumption is that the market has a factor structure; the third is that there are a large number of assets. The basic result of the APT, which is known as Ross's Theorem, is that under these assumptions many risks can be almost completely diversified away. The implications of the APT for asset pricing are familiar from the CAPM: in equilibrium the market rewards only undiversifiable risk. The two theories defined undiversifiable risk differently. In the APT undiversifiable risk is factor risk; in the CAPM it is correlation with the market. The APT is a useful theory if, and only if, factors and factor risk can be easily identified.

5.1 Factor Structure and Ross' Theorem

Ross originally developed the APT for an asset markets with an **exact K factor structure**. There is an exact K factor structure if asset returns may be described as

$$42)\qquad x_i = \mu_i + \sum_{k=1}^{K} \beta_{ik} f_k + \theta_i$$

where the f_k and the θ_i are all uncorrelated and $V(f_k) = 1$. This specification implies that

$$43)\qquad \Sigma_N = B_N B'_N + D_N$$

where B_N is an N by K matrix whose i^{th} row is $(\beta_{il}, \ldots, \beta_{iK})$ and D_N is a diagonal matrix. The β_{ik} are called factor loadings. Equation (42) states that the divergence of asset i from its expected value can be partitioned into factor risk, $\Sigma_i \beta_k f_k$, and idiosyncratic risk, θ_i. The key insight of the APT is that on a large asset market it ought to be possible to diversify away all idiosyncratic risk. The market will not reward people for holding the risk represented by θ_i. If this were strictly true then there would exist numbers τ_k (which could be described as factor risk premia) such that

$$44)\qquad \mu_i - \rho = \sum_{k=1}^{K} \beta_{ik} \tau_k, \text{ for all } i.$$

If (44) holds then the mean vector, μ_N, is linear combination of a vector of ones and the factor loadings, the columns of B_N. That is, (44) implies

$$45)\qquad \mu_N \in [e_N, B_N].$$

5.1.1 Approximate Factor Structure

Except for some special cases, (45) is not exactly correct. However for large N, a weaker result holds. We will prove this in the context of an **approximate factor structure** -- a concept which we now define. The requirement (of an exact factor structure) that the idiosyncratic disturbances be uncorrelated is unnecessarily restrictive both for theoretical and empirical purposes. Given a symmetric matrix C, let $g_i(C)$ denote the i^{th} largest eigenvalue of C. The nested sequence of variance covariance matrices $\{\Sigma_N\}$ has an approximate K factor structure if there is a sequence $\{(\beta_{il}, \ldots, \beta_{iK})\}_{i=l}^{\infty}$ such that for all N

46) $\Sigma_N = B_N B'_N + R_N$

where $(\beta_{il}, \ldots, \beta_{iK})$ is i^{th} row of the N by K matrix B_N and R_N is a sequence of positive semidefinite matrices with

47) $\bar{\lambda} \equiv sup_N \; g_1(R_N) < \infty.$

Note that Σ_N can have an approximate K factor structure if (42) holds even if idiosyncratic disturbances are correlated. What is required is that the θ_i should be uncorrelated with the factors and that the eigenvalues of the variance covariance matrix of the $\theta_1, \ldots, \theta_N$ should be uniformly bounded.

Chamberlain and Rothschild (1983) characterized approximate factor structures. Given a nested sequence of positive definite matrices $\{\Sigma_N\}$, define

48) $\lambda_{jN} = g_j(\Sigma_N)$

and

49) $\lambda_k = sup_N \; \lambda_{KN}$.

Theorem 1: *Suppose* $\lambda_\infty = inf_N \; \lambda_{NN} > 0$. *Then* $\{\Sigma_N\}$ *has an approximate K factor structure if and only if* $\lambda_{K+1} < \infty$.

Proof: One implication is easy. Suppose there is an approximate factor structure, then

50) $g_{K+1}(\Sigma_N) \leqslant g_{K+1}(B_N B'_N) + g_1(R_N) = g_1(R_N) \leqslant \bar{\lambda} < \infty.$

The proof that if λ_{K+1} is finite then there is an approximate factor structure is not given here.

Chamberlain and Rothschild also showed that if there is an approximate factor structure that it is unique in the following sense: If there is a nested sequence of N by K matrices C_N such that

$$\Sigma_N = C_N C'_N + S_N$$

where $g_1(S_N)$ is uniformly bounded, then $C_N C'_N = B_N B'_N$.

If there is an approximate K factor structure then it is a reasonable hypothesis that for $k = 1,\ldots,K$ the λ_{kN} should grow linearly. Since

51) $g_k(B_N B'_N) + g_1(R_N) \geqslant g_k(\Sigma_N) \geqslant g_k(B_N B'_N),$

λ_{kN} grows at the same rate as $g_k(B_N B'_N) = g_k(B'_N B_N)$. Suppose that factor loadings for different assets are i.i.d. random vectors with a positive definite second moment matrix Ω. Then

$\frac{B'_N B_N}{N}$ converges almost surely to Ω ; the eigenvalues of $B'_N B_N$ will converge to N times the eigenvalues of Ω ; they will grow linearly.

5.1.2 Ross' Theorem

We now prove that if the market does not permit arbitrage opportunities then (45) is approximately true.

Ross' Theorem: *If (A) holds and if there is an approximate K factor structure then there exist numbers $\tau_0, \tau_1, \ldots, \tau_K$ such that*

$$52) \qquad \sum_{i=1}^{\infty} (\mu_i - \tau_0 - \tau_1 \beta_{i1} - \cdots - \tau_K \beta_{iK})^2 < \infty.$$

This is an approximate result. Risk premia are not exactly linear functions of factor loadings. The are, however, close; (52) cannot hold unless the average pricing error is small. Ross' Theorem is an immediate consequence of Theorem 2 below which is somewhat more useful for empirical work. Suppose there are N assets with mean vector μ and variance-covariance matrix Σ. Let p be a portfolio and define δ as in (28). Suppose that Σ can be decomposed as

$$53) \qquad \Sigma = BB' + R$$

where B is N by K and R is nonnegative definite. For vectors p in R^N let $\|p\|_2$ denote the standard Euclidean norm. Let γ be the (Euclidean) distance of the mean vector m from the space spanned by the columns of B and a vector of ones; then, in an obvious notation $\gamma = \|\mu - [e, B]\|_2$; then let γ be the (Euclidean) distance of the mean vector M from the space spanned by the columns of B and a vector of ones. If the entries of B are interpreted as factor loadings, γ^2 is a measure of the amount by which pricing differs from exact arbitrage pricing.

Theorem 2: $\gamma^2 \leqslant g_1(R)\, \delta^2$

The proof given here is essentially due to Huberman (1982).

Proof: For this proof we use the symbol $\perp$ to denote orthogonality in the Euclidean norm. Let $\mu = y + z$ where $y \in [e, B]$ and $z \perp [e, B]$. Then $\gamma^2 = z'z$. Consider $p = z'x$. Since $z \perp e, C(p) = 0$. Furthermore

$$54) \qquad E(p) = z'\mu = z'z = \gamma^2,$$

and

$$55) \qquad V(p)^{1/2} = (z'\Sigma z)^{1/2} = (z'BB'z + z'Rz)^{1/2} = (z'Rz)^{1/2}.$$

Recall that if C is nonnegative definite

$$56) \qquad g_1(C) = sup_{\alpha \neq 0} \frac{\alpha' C \alpha}{\alpha' \alpha} = sup_{\alpha \neq 0} \frac{\alpha' C \alpha}{\|\alpha\|_2^2}.$$

Thus $z'Rz \leqslant z'z\; g_1(R)$ and (55) implies

$$57) \qquad V(p)^{1/2} \leqslant \gamma (g_1(R))^{1/2}.$$

Together (54) and (57) imply that

$$\frac{E(p)}{V(p)^{1/2}} \geqslant \frac{\gamma}{[g_1(R)]^{1/2}} .$$

Since $\alpha \geqslant \frac{E(p)}{V(p)^{1/2}}$ this completes the proof.

Note that in extending this theorem to prove Ross' Theorem what is required is that γ_N remain bounded as the number of assets increases. As we observed above this is a consequence of assumptions (A).

5.2 Diversifiable risk

It remains to show that risk premia are rewards for holding undiversifiable risk.

5.2.1 Well diversified portfolios

We begin by noting that the norm $\|\bullet\|_2$ introduced in the last section can be applied to portfolios and is a measure of diversification. If $p \in F_N$ and $p = \Sigma_i \gamma_i x_i$ then $\|p\|_2 = \left[\Sigma_i \gamma_i^2\right]^{1/2}$. Chamberlain (1983b) shows that $\|p\|_2$ can be extended from F to the limit portfolios in $\overline{F}$ in such a way that if $\{p_N\} \to p$ then $\|p_N\|_2 \to \|p\|_2$. If $\|p\|_2 = 0$ then we will say that p is a **well diversified** portfolio. This terminology is justified by the fact that if p is well diversified then it contains only factor risk and no idiosyncratic risk. If p is well diversified there is a sequence of finite portfolios $p_N = \gamma'_N x_N$ such that $P_N \to p$ and $\|p_N\|_2 = \gamma'_N \gamma_N \to 0$.

Then

$$V(p_N) = \alpha'_N \Sigma_N \alpha_N = \alpha'_N B'_N B_N \alpha_N + \alpha'_N R_N \alpha_N .$$

Therefore,

$$V(p) = \lim_{N \to \infty} V(P_N) = \lim_{N \to \infty} \alpha'_N B'_N B_N \alpha_N$$

It is also easy to show that if there is a riskless asset and if $\lambda_\infty \to 0$ then the riskless asset must be well diversified; (56) implies that for a finite portfolio p,

$$V(p) = \alpha' \Sigma_N \alpha \geqslant \|p\|_2 \lambda_N .$$

Chamberlain shows that the space $\overline{F}$ of all portfolios can be partitioned into the space of well diversified portfolios, D and a space of portfolios, J, which are uncorrelated with the well diversified portfolios, in the sense that every portfolio can be decomposed uniquely into a portfolio in D and a portfolio in J; $p = p_D + p_j$.

If there is an approximate K factor structure and if $\lambda_{KN} \to \infty$, then D is a finite dimensional vector space. If there is a riskless asset then D is of dimension $K+1$ and consists of the riskless asset and portfolios which are related to the eigenvalues corresponding to the K largest eigenvectors of the variance covariance matrix Σ_N in the following sense. Let t_{jN} be the eigenvector of Σ_N corresponding to λ_{jN} normalized so that $\|t_{jN}\|_2 = 1$. Let $q_{jN} = \lambda_{jN}^{-1/2} t_{jN}$. Now consider the portfolios $p_{jN} = q'_{jN} x_{jN}$. It is easy to see that

$$cov(p_{jN}, p_{kN}) = \begin{cases} 1 \text{ if } k = j \\ 0 \text{ if } k \neq j. \end{cases}$$

Furthermore, $\|p_{jN}\|_2^2 = 1/\lambda_{jN}$. Thus if for $j \leqslant K$, p_{jN} converges, it must converge to a well diversified portfolio; Chamberlain and Rothschild (1983) showed that the p_{jN} converge. These limit portfolios form a set of K uncorrelated portfolios which, together with the riskless asset, span the set D. If there is no riskless asset, then D consists of the K limit portfolios $p_j = \lim_{N \to \infty} p_{jN}$ for $j = l, ..., K$.

If there is an approximate K factor structure so that (46) holds, then a way to find it is to use the eigenvectors corresponding to the K, largest eigenvalues of Σ_N. Consider the matrix B_{NM} whose j^{th} column consists of the first N elements of q_{jM}. Then $\lim_{M \to \infty} B_{NM} B'_{NM} = B_N B'_N$.

5.2.2 Risk premia as rewards for holding factor risk

The sharpest result relating risk premia to factor loadings is due to Chamberlain (1983b). We will give his result under the assumption that there is a riskless asset with rate of return ρ; an analogous result holds if there is not a riskless asset. Suppose there is an approximate K factor structure. Then D, the space of well diversified portfolios, is of dimension $K + 1$ and we may choose as a basis for D the riskless asset s and K portfolios $f_1, \ldots, f_K$ which are uncorrelated and have unit variance. These portfolios are factors. For each asset i define $\beta_{ik} = cov(x_i, \beta_k)$. Now let

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

where $\tau = (\tau_0, \tau_1, \ldots, \tau_K)$. Chamberlain showed that $\gamma(\tau) < \infty$ if and only if

58) $\tau_0 = \rho,\ \tau_k = E(f_k))$ for $k = l, ..., k$

the coefficients τ_k are factor risk premia[8].

5.3 Bounds

If $\gamma(\tau) = 0$ then factor pricing is exact; an asset's price is determined entirely by its factor loadings. Clearly it is important to know when factor pricing is exact and, more generally to get an estimate of the amount by which it fails to be exact. We discuss this topic now. Suppose there is an approximate K factor structure. Let

59) $\gamma_N = \|\mu_N - [e_N, B_N]\|_2^2$.

Luedecke's dissertation (1984) contains a version of Theorem 2 which is useful for empirical work. Let T_{NK} be N by K matrix whose columns are the eigenvectors corresponding to the K largest eigenvalues of Σ_N. Define

$$\delta_N = sup \mid E(p) \mid / V^{1/2}(p) \text{ subject to } p \in F_N,\ C(p) = 0.$$

Then,

60) $\|\mu_N - [e, T_{NK}]\|_2^2 \leqslant \lambda_{K+1,N} \delta_N^2$.

8. The result that the τ_k are factor risk premia does not depend on factor pricing being exact. See Admati and Pfleiderer (1984) for a simpler proof.

Chamberlain (1983b) obtained bounds of a different sort. He showed that if there is a riskless asset with rate of return ρ then

61) $$\lambda_\infty^2 \, \rho^2 \|c\|_2^2 \leqslant \gamma(\tau) \leqslant \lambda_{K+1} \, \rho^2 \|c\|_2^2$$

where c is the portfolio which represents the cost functional and τ is given by (58). It is an implication of (61) that factor pricing is exact if and only if c is well diversified. An analogous result holds if there is no risky asset. These bounds are especially interesting because they provide sharp upper and lower bounds on the accuracy of the APT.

Connors (1984) has developed an interesting set of conditions for factor prices to be exact. The setting is an exchange economy in which the participants exchange assets (random variables with given distributions). Let x be these assets. Connors' model is a general equilibrium model so the x's are specified in physical units. Suppose the x's have an approximate K factor structure and that idiosynchratic risk is mean independent of factor risk. That is, when disturbances are written as in (41), $E(\theta \mid f) = 0$. Let y be the vector of percapita endowments of assets in this economy. Connors shows that if y is well diversified, then in competitive equilibrium asset prices for this economy are consistent with exact factor pricing; a set of asset prices in which factor pricing is not exact, cannot be a competitive equilibrium[9]. We may phrase this result somewhat loosely as stating that if the economy is large, in the sense that there are many investors and many assets, factor pricing should be close to exact. In two similar papers, Dybvig (1983) Grinblatt and Titman (1983) have shown how close to exact factor pricing must be. Both papers assume a strict factor structure in which factors and idiosyncratic noise are independent. The argument can be easily extended to deal with an approximate factor structure, but the independence assumption seems crucial. Dybvig's bounds are so sharp (his rough estimates suggest that no asset's risk premium should deviate from its factor price -- $\Sigma_k \tau_k \beta_{ik}$ -- by no more than .04%) that his work turns the APT into a theory like the CAPM in the sense that it provides a complete explanation of risk premia. However, this does not imply that R^2 in regression tests of the APT must be high. Measurement error (in estimating expected returns and factors) is a serious problem in empirical work.

5.4 The APT in a conditional setting

Unlike the CAPM, the arbitrage pricing theory works well in a conditional setting. Our conditional model assumes that market participants have information s and make investment decisions on the basis of conditional distributions of asset returns. The analyst can not observe s and can only estimate aspects of the unconditional or ergodic distributions of asset returns. Suppose that the unconditional returns have an approximate K factor structure and that none of the conditional distributions (as described by μ_s and Σ_s) permit arbitrage. About the best result one could hope for is the following result due to Stambaugh (1983):

Theorem: *Suppose unconditional returns have an approximate K factor structure and that conditional distributions do not permit arbitrage, then*

62) $\|\overline{\mu} - [e_N, \overline{B}_N]\|_2$ *is uniformly bounded.*

Proof: It will suffice to show that assumption (A.i) holds for the unconditional data. Suppose

9. Connors' definition of well diversified is different from Chamberlain's. However in the context of an approximate K factor structure, they are equivalent. Connors does not note, but it is easy to prove, that an approximate K factor structure implies that his condition (5), necessary and sufficient for exact factor pricing, holds.

that (A.i) does not hold for the conditional data. Then there is a sequence of portfolios $\{p_N\}$ such that $V(p_N) \to 0, C(p_N) \to 0$ while $E(p_N)$ remains, at least on a subsequence, bounded away from 0. Let $q_N = p_N / E(p_N)$. Then $V(q_N) \to 0$, $C(q_N) \to 0$ and $E(q_N) = 1$. The variance decomposition formula, $V(x) = E[V(x) \mid \tilde{s}] + V[E(x \mid \tilde{s})]$, implies that $V(q_{Ns}) \to 0$ and $E(q_{Ns}) \to 0$ for all s. Note that the cost of the portfolio q_N is independent of s. Thus the cost of q_N coverages to zero while for those whose beliefs are given by s, $V(q_N \mid s) \to 0$ and $E(q_N \mid s) \to 1$. Thus (A.i) is violated conditionally.

While this is a satisfying result, one would like to know more about the relationship between conditional and unconditional inference in the context of the APT. There should be a sense in which unconditional arbitrage pricing is less exact than conditional arbitrage pricing. In particular it would be interesting to compare how the bounds discussed above behave in conditional and unconditional settings.

5.5 Empirical Tests of the APT

Tests of the arbitrage pricing theory have tended to focus on three questions: The first is can factors be identified and used to explain risk premia? The second is does the theory provide evidence against the CAPM? The third is are the explanations of risk premia sharp and are there a usefully small number of factors? I read the evidence as answering these questions: "Yes, yes, and no." Several investigators have found factors and used them to explain risk premia. A common finding is that factors explain risk premia better than beta. This is convincing evidence against the CAPM. Representative studies are Roll and Ross (1980), Chen (1982), and Luedecke (1984). Chen, Roll and Ross (1983) offer an intriguing explanation for the fact that beta seems to explain asset returns. Adopting an external approach, they attempt to explain asset returns in terms of their correlations with macroeconomic variables. One variable which does very well is a measure of future industrial production; stock market indices like the Standard and Poor 500 (which are used in empirical tests of the CAPM) predict this variable well.

While tests of the APT provide damning evidence against the CAPM, they do not lead to the conclusion that the APT itself provides a parsimonious explanation of risk premia. The APT would provide such an explanation if there were evidence that a small number of factors explain most risk premia. We can explore this question using the bounds given above.

Luedecke's bounds, (60) depend on δ. Attempts to estimate δ from CRSP data produced very large estimates. Similarly, exploratory estimates of $\|c\|_2$, the crucial parameter in Chamberlain's bound, produced disappointingly large results.

The evidence on the number of factors is also somewhat disheartening. Chamberlain and Rothschild's results suggest that the number of exploding eigenvalues of Σ_N is equal to the number of factors. While one cannot in a finite sample determine whether an increasing sequence is bounded, it is possible to simply look at the behavior of λ_{kN} for large N. Both Luedecke and Trzcinka (1984) have done this. They found that the first eigenvalue was much larger than the others; it seemed to grow linearly. However, other eigenvalues are also growing. It is impossible to tell from their data whether there are 2 factors or 10 or 15. The hypothesis that there is just one factor does not seem attractive because in cross sectional regressions the factors corresponding to eigenvalues other than the largest eigenvalue are often significant. Luedecke's results are most convincing when he uses the entire sample to estimate expected returns. This procedure is legitimate because the APT makes sense in a conditional context.

6. Conclusions

The APT and the CAPM are sophisticated theories which attempt to explain asset prices. The APT is framed in a way which makes it more compatible with the kind of data which is available for testing theories of asset pricing. The APT is more consistent with the evidence than the CAPM. However, evidence from tests of the APT suggests that it is unlikely that a simple equation with a few parameters will explain asset prices well.

REFERENCES

Admati, Anat R. and Paul Pfleiderer, 1984 "Interpreting the Factor Risk Premia in the Arbitrage Pricing Theory," *Journal of Economic Theory*, forthcoming.

Anderson, T.W., 1958 *An Introduction to Multivariate Statistical Analysis*, Wiley, New York.

Black Fischer, 1972 "Capital Market Equilibrium with Restricted Borrowing," *Journal of Business*, (45:444-454).

Chamberlain, Gary, 1983a "A Characterization of the Distributions that Imply Mean-Variance Utility Functions," *Journal of Economic Theory*, (28:183-192).

Chamberlain, Gary, 1983b "Funds, Factors and Diversification in Arbitrage Pricing Models," *Econometrica*, (51:1305-1324).

Chamberlain, Gary and Michael Rothschild, 1983 "Arbitrage, Factor Structure and Mean-Variance Analysis on Large Asset Markets," *Econometrica*, (51:1281-1304).

Chen, Nai-Fu, 1983 "Some Empirical Tests of the Arbitrage Pricing Theory." *Journal of Finance*, (38:1393-1414)

Chen, Nai-Fu, Richard Roll and Stephen A. Ross, 1983 "Economic Forces and the Stock Market: Testing the APT and Alternative Asset Pricing Theories," Working Paper No. 119 Center for Research in Security Prices, Graduate School of Business, University of Chicago.

Connor, Gregory, 1984 "A Unified Beta Pricing Theory," *Journal of Economic Theory*, (34:13-31).

Cragg, John and Burton Malkiel, 1982 *Expectations and the Structure of Share Prices*, University of Chicago Press.

Dybvig, Phillip H., 1983 "An Explicit Bound on an Individual Asset's Deviations from APT Pricing in a Finite Economy," *Journal of Financial Economics*, (12:483-96).

Dybvig, Phillips H and Stephen A. Ross, 1984 "Yes the APT is Testable," Unpublished paper, Yale School of Management.

Fama, Eugene F. and James D. MacBeth, 1973 "Risk Return and Equilibrium: Empirical Tests," *Journal of Political Economy*, (81:607-636).

Fogel, Robert William and Stanley L. Engerman, 1974 *Time on the Cross*, Little Brown, Boston.

Gibbons, Michael R., 1982 "Multivariate Tests of Financial Models: A New Approach," *Journal of Financial Economics*, (10:3-28).

Grinblatt, Mark and Sheridan Titman, 1983 "Factor Pricing in a Finite Economy," *Journal of Financial Economics*, (12:497-507).

Huberman, Gur, 1982 "A Simple Approach to Arbitrage Pricing Theory," *Journal of Economic Theory*, (28:183-92).

Jensen, Michael C., "Capital Markets: Theory and Evidence," *Bell Journal of Economics and Management Science*, (3:357-398).

Luedecke, Bernd Paul, 1983 "An Empirical Investigation into Arbitrage and Approximate K-Factor Structure on Large Asset Markets," Unpublished doctoral dissertation, University of Wisconsin.

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

Roll, Richard 1977, "A Critique of the Asset Pricing Theory's Tests," *Journal of Financial Economics*, (4:129-76).

Roll, Richard and Stephen A. Ross, 1980 "An Empirical Investigation of the Arbitrage Pricing Theory," *Journal of Finance*, (35:1073-1103).

Ross, Stephen A., "The Arbitrage Theory of Capital Asset Pricing," *Journal of Economic Theory*, (13:341-60).

Shanken, Jay, 1982 "The Arbitrage Pricing Theory: Is it Testable?" *Journal of Finance*, (37:1129-1140).

Sharpe, William F., 1964 "Capital Asset Prices: A Theory of Market Equilibrium Under Conditions of Risk," *Journal of Finance*, (19:425-442).

Stambaugh, Robert F., 1982 "On the Exclusion of Asset from Tests of the Two-parameter model: A Sensitivity Analysis," *Journal of Financial Economics*, (10:237-268).

Stambaugh, Robert F., 1982 "Arbitrage Pricing with Information," *Journal of Financial Economics*, (12-357-369).

Treynor, J. L., 1961 "Towards a Theory of Market Value of Risky Assets," Unpublished paper.

Trzcinka, Charles, 1984 "On the Number of Factors in the Arbitrage Pricing Model," University of Buffalo.