THE JOURNAL OF FINANCE • VOL. XXXVIII, NO. 5 • DECEMBER 1983

Some Empirical Tests of the Theory of Arbitrage Pricing

NAI-FU CHEN*

ABSTRACT

We estimate the parameters of Ross's Arbitrage Pricing Theory (APT). Using daily return data during the 1963–78 period, we compare the evidence on the APT and the Capital Asset Pricing Model (CAPM) as implemented by market indices and find that the APT performs well. The theory is further supported in that estimated expected returns depend on estimated factor loadings, and variables such as own variance and firm size do not contribute additional explanatory power to that of the factor loadings.

THE ARBITRAGE PRICING THEORY (APT), originally formulated by Ross [35, 36] and extended by Huberman [23] and Connor [13], is an asset pricing model that explains the cross-sectional variation in asset returns. Like the Capital Asset Pricing Model (CAPM) of Sharpe [39], Lintner [26], and Black [2], the APT begins with an assumption on the return generating process: each asset return is linearly related to several, say k, common "global" factors plus its own idiosyncratic disturbance. Then in a well-diversified, frictionless, and perfectly competitive economy, the no arbitrage condition requires that the expected return vector must lie (asymptotically) in the $k + 1$ dimensional vector space spanned by a vector of all one's and the k vectors of asset response amplitudes (to the k common global factors).

The initial empirical evidence on the model has been rather encouraging (see Gehr [19] and Roll and Ross [34]). In this paper, we shall compare the empirical performance of the APT with that of the CAPM. We shall also test whether the APT can explain some of the empirical "anomalies" related to the CAPM in recent years. The paper has six sections. In Sections I and II, some basic results related to the APT are given so that testable implications of the model can be clearly identified and the parameters of the model can be estimated. Section III contains the cross-sectional results of the APT. In Sections IV and V, we attempt to reject the APT by looking at variables that are known to be highly correlated with returns to see if they have any additional explanatory power after the APT parameters are included. We summarize our findings in Section VI. Some

* Graduate School of Business, University of Chicago. I thank my dissertation committee chairman, Richard Roll, for his guidance and encouragement; Eugene Fama for his suggestions; and Glenn Graves for his assistance in the mathematical programmings. I have also benefited from the comments of Armen Alchian, George Constantinides, Tom Copeland, Robert Geske, Jack Hirshleifer, Robert Hamada, Herb Johnson, Edward Leamer, Robert Litzenberger, David Mayers, Merton Miller, Marc Reinganum, Fred Weston, Arnold Zellner, and especially Stephen Brown and Michael Gibbons. This research was partially supported by the Research Program in Competition and Business Policy at the University of California, Los Angeles, and a University of California, Santa Barbara Academic Senate grant 8-581557-07427-7.

mathematical derivations and the estimation procedure for the factors are described in the Appendix.

I. The Arbitrage Pricing Theory and Its Implications

A. A Brief Review of the APT[1]

Assume that asset markets are perfectly competitive and frictionless and that individuals believe that returns on assets are generated by a k-factor model, so that the return on the i^{th} asset can be written as:

$$\tilde{r}_i = E_i + b_{i1}\tilde{\delta}_1 + \cdots + b_{ik}\tilde{\delta}_k + \tilde{\epsilon}_i \tag{1}$$

where E_i is the expected return; $\tilde{\delta}_j$, $j = 1, \cdots, k$, 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 $\tilde{\epsilon}_i$ is the "nonsystematic" risk component idiosyncratic to the i^{th} asset with $E\{\tilde{\epsilon}_i | \tilde{\delta}_j\} = 0$ for all j. In a well diversified economy with no arbitrage opportunity, the equilibrium expected return on the i^{th} asset is given by:

$$E_i = \lambda_0 + \lambda_1 b_{i1} + \lambda_2 b_{i2} + \cdots + \lambda_k b_{ik}. \tag{2}$$

If there exists a riskless (or a "zero beta") asset, its return will be λ_0. The other parameters, $\lambda_1, \ldots, \lambda_k$, can be interpreted as risk premiums corresponding to risk factors $\tilde{\delta}_1, \ldots, \tilde{\delta}_k$. In other words, λ_1 is the expected return per unit of long investment of a portfolio with zero net investment and $b_{p1} = 1$ and $b_{p2} = \cdots = b_{pk} = 0$.

B. Factor Analysis and the Estimation of the Factor Loadings

The procedure to estimate factor loadings (i.e., the b_{ij}'s) for all assets corresponding to the same set of common factors is quite involved and expensive. We first do a factor analysis on an initial subset of assets, and then we extend the factor structure of the subset to the entire sample. This is accomplished via a large scale mathematical programming exercise. Section II contains a brief outline.

It is clear that the development of the theory of arbitrage pricing is quite separate from the factor analysis. We use factor analysis here only as a statistical tool to uncover the pervasive forces (factors) in the economy by examining how asset returns covary together. As with any statistical method, its result is meaningful only when the method is applied to a representative sample. In the present context, the initial subset to which the factor analysis is applied should consist of a large random sample of securities of net positive supply in the economy; thus the sample would be closely representative of the risks borne by investors. In a recent article, Shanken [37] points out some of the potential pitfalls of testing the APT when the factored covariance matrix is unrepresentative of the covariation of assets in the economy. By forming portfolios from

[1] See Ross [35, 36], Huberman [23], and Connor [13] for the formal development. See also Ingersoll [24], Chen and Ingersoll [10], Grinblatt and Titman [22], and Dybvig [16].

any given set of assets, Shanken demonstrates that factor analysis can produce many different factor structures from the manipulated portfolios. In the extreme case where the constructed portfolios are mutually uncorrelated, factor analysis produces no common factor. Of course, forming uncorrelated portfolios by longing and shorting securities merely repackages the risk bearing and potential reward associated with the original securities and does not alter the fundamental forces and characteristics inherent in the economy. However, as a statistical tool, factor analysis can no longer detect those pervasive forces from such manipulated portfolios.[2] This, of course, should not be construed as a criticism of the theory or of the testability of the APT, but rather should serve as a reminder of the potential problems involved in doing statistical analysis on unrepresentative samples. In this paper, we select 180 securities for each of the initial factor analyses. If we miss an important factor because of unrepresentativeness, all the tests that follow will be biased against the APT.

C. Testable Hypothesis of the APT

We regard Equation (2) as the main result of the APT that explains the cross-sectional differences in asset returns, and it is (2) that will be tested in the following sections.

A logical first step in testing (2) would be to look for priced factors. However, the task of finding priced factors turns out to be not particularly straightforward. If we have determined that k factors exist (in the sense of Connor [13]) in the generating process of asset returns in the economy, then the number of priced factors—as long as there is at least one—is not well defined. Intuitively, this can be most easily seen by noting that a k factor pricing equation can always be collapsed into a single beta equation via mean-variance efficient set mathematics (and with an additional orthogonal transformation similar to that described below, the number of priced factors can be arbitrary). Mathematically, let v be the risk premium vector for a particular set of risk factors and let u be *any* vector of the same dimension with $u'u = v'v$. Then there exists an orthogonal transformation that will transform the original set of factors to a new set of orthogonal factors whose associated risk premium vector is u. In other words, if it has been established that k factors are present, the number of priced factors can be any number between 1 and k. It should be emphasized here, however, that those factors that are not priced are just as important as those that are priced in an individual's investment decision (see Breeden [4], Constantinides [12] and Roll [32] for related issues). They are irrelevant only in predicting expected return. This should be borne in mind when interpreting the cross-sectional results in Section III.

A question that naturally arises in this investigation is how the APT fares against other asset pricing models. It is immediately apparent from Equations

[2] The term "pervasive forces" was made popular by Connor [13]. See Shanken [37] for his interpretation of Connor's result and its implications to the APT. Contrary to some beliefs, Shanken's results were not driven by the idiosyncratic terms in a finite sample or the approximate nature of Ross' original formulation. The no common factor result can be obtained in an economy with exactly k factors and *no* idiosyncratic risks.

(1) and (2) that any model that predicts a linear relation between "risk" and return is potentially consistent with the APT.[3] For example, if we let δ_1 be the return on the market and δ_2 be changes in interest rate, we obtain Merton's [27] two factor pricing equation. Therefore, unless the "market" and/or the "factors" can be properly identified, accepting the APT (i.e., not being able to reject the APT) does not necessarily reject the other pricing equations. However the results of APT can be compared with other models as popularly implemented. In Section III, we examine the cross-sectional results of the APT and the CAPM.

It has been suggested that the APT is so general that it is not rejectable. Fortunately, this is not true. The most important result of the APT is that only those risks that are reflected in the covariance matrix are priced, *nothing else.* So if we are able to find a variable that is priced even after the factor loadings (FL) are accounted for, the APT would be rejected. For example, if the daily return to every small firm were increased by 1% per day, the returns to small firms would be statistically significantly higher than the returns to large firms, no matter how many factors were extracted from the covariance matrix to account for risk ex ante, as long as the number of factors is small[4] relative to the number of securities.

In Sections IV and V, we attempt to reject the APT using variables such as own variance and size of firm equity. Those variables are chosen because of the well documented high correlation between them and the average returns.

II. Estimation of the Factor Loadings

A. The Data

The data are described in Table I. All the parameters in Sections III, IV, and V are computed using only data within each subperiod so that we may have four independent tests of each hypothesis.

The computation of the b_{ij}'s, the factor loadings (FL), uses only data from odd days within each subperiod. Even day returns are reserved for testing purposes.

B. Methodology

The details of the estimation procedure are described in the Appendix; the following is a brief outline:

(i) The first 180 stocks in the sample (alphabetically) are selected and their sample covariance matrix computed. The choice 180 is the upper limit

[3] It is often asserted that the CAPM is a special case of the APT. This is true if and only if there exists a rotation of the factors such that one of the factors is the "market." Ex ante, there is no reason to assume this is the case in our finite economy. Ex post, we discover in this study as well as in subsequent studies [7, 11] that the first factor is highly correlated with a measure of the "market" such as the equally or value weighted NYSE stock index. However, throughout this study, we maintain the ex ante position that the CAPM and the APT are nonnested for hypothesis testing.

[4] "Small" here refers to the number of pervasive factors that influence the returns of most assets, and this number is no more than a few and certainly less than, say, 20. We insert "small" here to rule out the pathological case where the number of factors is equal to the number of assets and therefore any pattern of returns can be explained. In the present study, the maximum number of factors is limited to 180, the size of the initial factor analysis, while the number of securities in each period is more than 1,000.

Table I

Data Description

Source:	Center for Research in Security Prices Graduate School of Business University of Chicago Daily Returns File
Sample period:	1963–78 inclusive. The entire period is divided into four subperiods: I. 1963–66, II. 1967–70, III. 1971–74, and IV. 1975–78.
Selection criterion:	All the securities that do not have missing data during each subperiod.[a]
Basic data unit:	Return adjusted for all capital changes and including dividends.

Number of selected securities:	Subperiod	Total sample
	I	1,064
	II	1,522
	III	1,580
	IV	1,378

[a] Their average daily return in absolute value is less than 0.01 to eliminate outliers. Only Burma Mining Inc. was excluded with this criterion during the first subperiod.

placed by the processing capacity of the IBM 3033 used (in 1980) for the computation of the factor loadings.

(ii) The first *ten* factor loadings for each stock are obtained with the computer software package EFAP II.

(iii) Five portfolios are formed using linear programming[5] so that the resultant portfolios will balance estimation errors with other desirable properties. The time series of the five portfolios will contain linear combinations of the $\tilde{\delta}_1, \cdots, \tilde{\delta}_k$ (i.e., the factor scores).

(iv) The first *five* factor loadings are produced for every stock in the sample by solving a matrix equation (Equation (A1) in the Appendix).

One of the difficulties in empirically testing the APT is that it does not tell us what the number of common factors should be. Since *ex post* data are being used to test for an *ex ante* relation, the number of factors to be included must be independently determined and prespecified in order to avoid potential data mining and to give the alternative hypothesis a fair chance *ex post.* Five were selected based on previous empirical studies (see Roll and Ross [34] and Reinganum [30]).[6] A study by Brown and Weinstein [5] also confirms that the number of pervasive factors is probably no greater than five.

III. Cross-Sectional Results

A. The APT and the CAPM

To see how well the data support the models, we examine the result of cross-sectional regression of assets' returns on the hypothesized parameters in each of the subperiods. The independent variables will be the FL for the APT and the

[5] This is the GUB programming within the elastic programming in the *XS* mathematical programming system developed by Glenn Graves, UCLA.

[6] Based on the analysis and the plot of eigenvalues, for each period five factors also look sufficient.

betas for the CAPM (both computed on the odd days):

$$r_i = \lambda_0 + \lambda_1 \hat{b}_{i1} + \cdots + \lambda_k \hat{b}_{ik} + \epsilon_i, \qquad \text{(APT)} \tag{3}$$

$$r_i = \lambda_0 + \lambda_1 \hat{\beta}_i + \eta_i. \qquad \text{(CAPM)} \tag{4}$$

The returns are computed on the even days of each subperiod. The betas are computed with market proxies: (1) the S&P 500 index, (2) the value weighted stock index, and (3) the equally weighted stock index. The returns on the indices are taken from the CRSP tape index file. The result of the regression is given in Table II, Parts A and B.

The adjusted R^2 comes from the cross-sectional regression of assets' *average* (even day) returns over the entire period on the independent variables. The estimated coefficients and their t statistics come from the time series of cross-sectional regression coefficients as in the studies of Black, Jensen, and Scholes (BJS) [3] and Fama and Macbeth (FM) [17]. In our case, we first compute the average of every five (even) day returns and perform a cross-sectional regression on each of them, thus generating a time series for each estimated coefficient. The mean and the t statistic are then derived from the time series.[7] Almost all the serial correlation coefficients are insignificant; therefore, the time series sample may be treated as essentially independent. A nonparametric test is also performed on the time series of each coefficient to hedge against nonnormality of the population. Here the "sign test" is used to test whether the median is zero. Since the power of the nonparametric test is in general lower than parametric tests, both significance at the 0.1 level and at the 0.05 level are reported next to the estimated coefficient. The Hotelling T^2 (see Morrison [28]) in Table II is computed from the time series of $\hat{\lambda}_1, \cdots, \hat{\lambda}_5$. The T^2 statistics are reported alongside the F's because it is easier to add up T^2 (which asymptotically approaches χ^2). The interpretation of these statistics follows.

In looking at the results in Table II, Part A, recall the rotation indeterminancy associated with factor analysis (Section I. C above). Thus, comparison of $\hat{\lambda}_i$ across time periods is not meaningful. The only exceptions are the $\hat{\lambda}_0$'s, which are the estimated expected return of a zero-FL asset. It also happens that the first FL of each asset is highly correlated with the β of CAPM. The simple correlation between the $\hat{b}_{i1}$ and the $\hat{\beta}_i$ (for each market proxy) is in the neighborhood of 0.95. Almost all the $\hat{b}_{i1}$'s are negative; therefore, one would expect the risk premium (i.e., the estimated λ_1) to be negative. Indeed, all the estimated λ_1's are negative (even for the period 1967–70 when the estimated market premium is negative; see Table II, Part B); however, only the first and the fourth periods' estimated λ_1's are significantly different from zero. This is consistent with the result in Table II, Part B, where the CAPM β is priced only in those two periods. As for the other estimated λ's, there is no a priori information on their signs; therefore, one can judge only by their significance level whether that factor is priced. From

[7] The error matrix of the estimated premia is $(X'X)^{-1}X' \sum X(X'X)^{-1}$, where $\sum$ is the cross-sectional covariance matrix of the idiosyncratic terms under the null hypothesis $\lambda_1 = \lambda_2 = \cdots \lambda_5 = 0$. If the b_{ij} are estimated with error, the error matrix for the risk premia remains the same under the null hypothesis in some special cases. See Gibbons [21] or Shanken [38].

Table II

Cross-sectional Regression of Returns

A. On the Factor Loadings[a]

$$r_i = \lambda_0 + \lambda_1 \hat{b}_{i1} + \lambda_2 \hat{b}_{i2} + \lambda_3 \hat{b}_{i3} + \lambda_4 \hat{b}_{i4} + \lambda_5 \hat{b}_{i5} + \epsilon_i$$

	Statistics[b]						Hotelling T^2			Adjusted
Period	$\hat{\lambda}_0$	$\hat{\lambda}_1$	$\hat{\lambda}_2$	$\hat{\lambda}_3$	$\hat{\lambda}_4$	$\hat{\lambda}_5$	T^{2c}	F^d	F_4^e	R^2
1963–66	0.0028	−0.5959**	−0.2689	0.3013**	0.1640**	−0.2942*				
	(0.18)	(−3.88)	(−1.48)	(2.16)	(1.21)	(−2.47)	27.49	5.28	4.23	.2784
1967–70	0.0144	−0.0582	0.2879*	0.0360	0.0119	−0.2042**				
	(0.61)	(−0.33)	(2.25)	(0.38)	(0.11)	(−2.28)	11.45	2.20	2.77	.0281
1971–74	−0.0118	−0.0142	0.1318**	−0.2286	0.0740	−0.0255				
	(−0.44)	(−0.08)	(1.41)	(−2.34)	(1.14)	(−0.27)	11.66	2.24	2.80	.0419
1975–78	0.0342	−0.4255**	−0.2339*	0.0727	0.0472	0.0256				
	(1.62)	(−2.18)	(−1.98)	(0.75)	(0.44)	(0.28)	10.36	1.99	1.36	.1411
							$\sum T^2 = 60.96$		$\sum T^2(16) = 45.98$	

B. On Beta[a]

$$r_i = \lambda_0 + \lambda_1 \hat{\beta}_i + \eta_i$$

	Market[f]	Statistics[b]				Adjusted
Period	Proxy	$\hat{\lambda}_0$		$\hat{\lambda}_1$		R^2
1963–66	I	−0.0071	(−0.42)	0.0879**	(3.77)	0.1963
	II	−0.0082	(−0.48)	0.0848**	(3.80)	0.2019
	III	−0.0107	(−0.67)	0.0891**	(4.01)	0.2167
1967–70	I	0.0401	(1.67)	−0.0077	(−0.27)	0.0011
	II	0.0391	(1.64)	−0.0066	(−0.24)	0.0007
	III	0.0343	(1.46)	−0.0032	(−0.09)	−0.0004
1971–74	I	−0.0343	(−1.05)	0.0254	(0.75)	0.0088
	II	−0.0329	(−1.03)	0.0225	(0.67)	0.0072
	III	−0.0124	(−0.48)	0.0006	(0.02)	−0.0006
1975–78	I	0.0579	(2.36)	0.0629*	(1.84)	0.0698
	II	0.0528	(2.23)	0.0629**	(1.93)	0.0785
	III	0.0284	(1.42)	0.0816**	(2.45)	0.1357

[a] The average daily returns are multiplied by 100 before regressing.

[b] The t statistics are in parentheses. There are 100 d.f. except for the period 1967–70, which has 97 d.f. The significance level of the sign test is indicated by * (0.1 level) and ** (0.05) level.

[c] Critical values of χ^2 with 5 d.f. are 9.24 (0.1), 11.1 (0.05), 15.1 (0.01), and χ^2 with 20 d.f. are 28.4 (0.1), 31.4 (0.05), 37.6 (0.01), 45.31 (0.001).

[d] The F statistics have degree of freedom (5, 96) for periods 1, 3, and 4 and (5, 93) for period 2. The upper percentage points of the F distribution are 1.95 (0.1 level), 2.37 (0.05 level) for (5, 60) d.f., and 1.90 (0.1 level), 2.29 (0.05 level) for (5, 120) d.f.

[e] The Hotelling T^2 statistics on the last four risk premium series, the upper percentage points of the F distribution are 2.04 (0.1 level) and 2.53 (0.05 level) for (4, 60) d.f. The critical values of χ^2 with 16 d.f. are 23.54 (0.1), 26.29 (0.05), 32.00 (0.01), 39.25 (0.001).

[f] The betas are computed using market proxies: (I) S&P 500 index; (II) value weighted stock index; and (III) equally weighted stock index.

Table II, Part A, it is interesting to note that there is at least one significant factor beside $\hat{\lambda}_1$ for every period.

To see whether APT has any explanatory power in cross-section we test the null hypothesis that $\lambda_1 = \lambda_2 = \lambda_3 = \lambda_4 = \lambda_5 = 0$. This is equivalent to the hypothesis that the expected return of all assets are the same and equal to λ_0. The Hotelling T^2 is a multivariate test (see Morrison [28]) appropriate for this purpose. However, it is tricky to apply the test to the APT because, as mentioned previously, the APT only requires at least one of the factors to be priced. If we then choose only the first few most significant factors to be included in the Hotelling T^2, obviously the Type I error is underestimated and there is a bias against the null hypothesis. If we always take all five, the test will be weak and the Type II error will be large.[8] Despite the relative low power of the test when all five are included, the F statistic is still significant at least at the 0.1 level for every period (see Table II, Part A), so the overall significance level would be very high. If we assume independence across the four test periods, we can sum up the $T^2(\sum T^2 = 60.96)$ which asymptotically approaches χ^2 with 20 degrees of freedom and with a critical value of 45.3 for the 0.001 level.[9] Therefore, with the FL, we can confidently reject the null hypothesis of constant expected return across assets.

B. A Comparison between the APT and the CAPM[10]

The next question that comes up naturally is: which of the two competing models, CAPM or APT, do the data favor? To answer this, one is tempted to do a regression with both CAPM beta and the FL as independent variables. However, this is not done because: (i) this specification is not justified on theoretical grounds and the two models are nonnested (see also footnote 3), and (ii) the betas and the FL are intended to measure the same thing—risk. Thus, to put them together in a regression would mean including the same variable twice on the right-hand side. The high degree of multicollinearity (intertwined with measurement error) tends to produce regression coefficients that make no sense.

One possible way to discriminate among nonnested alternative models was suggested in Davidson and Mackinnon [14]. Let $\hat{r}_{\text{APT}}$ and $\hat{r}_{\text{CAPM}}$ be the expected return generated by the APT and the CAPM, i.e., they are obtained from the cross-sectional regression of (3) and (4) (on even days) without the error terms; then if we estimate α in

$$r_i = \alpha \hat{r}_{i,\text{APT}} + (1 - \alpha)\hat{r}_{i,\text{CAPM}} + e_i \tag{5}$$

(cf. [14], Equation (4)), we would expect α to be close to 1 if the APT is the correct model relative to the CAPM. We analyze (5) rather than many of its

[8] Gibbons [20] showed that a much more efficient estimate of the risk premium is possible by pooling cross-sectional time series data in a seemingly unrelated regression. The standard error in his study was roughly one third of the BJS-FM standard error. Unfortunately, there are over a thousand securities in our cross-sectional sample. Thus the SUR, which requires inversion of the sample covariance matrix, is not feasible. Furthermore, the probability limit of the Hotelling T^2 statistic is biased downward in the presence of nonsystematic measurement error.

[9] An approximate F statistic reconstructed from the overall T^2 statistic is 2.46 with (20, 78) d.f.

[10] I am indebted to Stephen Brown for a discussion on this topic.

Table III

Estimated Weights of the Expected Return from the APT and CAPM[a]

$r_i = (1 - \alpha)\hat{r}_{CAPM} + \alpha\hat{r}_{APT} + e_i$

Period	S&P 500	$\hat{\alpha}$ Value Weighted Stock Index	Equally Weighted Stock Index
1963–66	0.968	0.970	0.992
	(0.014)	(0.014)	(0.010)
1967–70	1.006	0.994	0.952
	(0.014)	(0.011)	(0.010)
1971–74	0.938	0.945	0.951
	(0.021)	(0.019)	(0.025)
1975–78	0.953	0.970	0.994
	(0.014)	(0.010)	(0.020)

[a] Standard errors in parentheses.

close alternatives (see [14]) because of the symmetric treatment of $\hat{r}_{APT}$ and $\hat{r}_{CAPM}$. The penalty for regressing (5) is that the asymptotic standard error of α is underestimated. A multiplicative adjustment is necessary. However, since there are so much data in the stock returns, the mean and standard error of α can be estimated directly from its time series, which is obtained by performing (5) in subintervals within each period. Each subinterval contains five (even) days. The point estimate of α, which is contained in Table III, is persuasive of the APT, even though in many cases the estimated α is significantly different from 1.

A better criterion, based on theoretically sound foundations, is suggested by the Bayesians. Had the residuals of (3) and (4) satisfied the i.i.d. multivariate normal assumption, posterior odds ratios can be computed to provide a selection rule.[11] With diffuse prior and some assumptions,[12] the formula for posterior odds in favor of model 1 over model 0 is given by

$$R = [\mathrm{ESS}_0/\mathrm{ESS}_1]^{N/2}N^{(k_0-k_1)/2} \qquad (6)$$

(cf. Leamer [25, p. 114]), where ESS is the error sum of squares, N is the number of observations, and k is the dimension of the respective models.

The posterior odds thus computed are, with one exception, overwhelmingly in favor of the APT over the CAPM as implemented by the three market indices. The odds ratios in favor of the APT are never less than $3.64E + 3$ for all four periods and all three market indices and are as high as $5.17E + 19$, except for the equally weighted stock index in the period 1975–78 ($R = 2.94E - 4$).

Unfortunately, while we try to extract all the cross-sectional covariance through factor analysis in the case of APT, the same cannot be said about CAPM. In fact, it is well known that the residuals across firms tend to be positively

[11] For a survey of posterior odds methods, see Zellner [41].

[12] See Zellner [40], Chapter 10, pp. 306–312 for details. The side assumptions are those that reduce the posterior odds ratio to goodness of fit. See related issues in Gaver and Geisel [18] and Leamer [25]. I thank Edward Leamer and Arnold Zellner for a discussion on this topic.

correlated in CAPM, and the residual variances are underestimated.[13] Therefore, readers should keep this caveat in mind as they interpret the odds ratios.

Further comparison of the two models requires additional information; tests incorporating this information are examined in the following sections. The results of this section, together with additional evidence we gather later, will enable a more definitive comparison.

C. Performance Measurement and the APT

The residuals from the CAPM are of some interest. They are sometimes used for performance measurement. If the CAPM is not misspecified, the expected return of asset i would be captured by β_i, and the residual η_i in Equation (4) would behave like noise and have a zero mean (across time). However, if the CAPM is misspecified and the β_i does not capture all the information about expected return, the remaining part will be contained in the residual η_i. In this case, the η_i will no longer behave like noise across time, and, if there is another model that can price the remaining part of the expected return, the η_i will be priced by that model. Therefore, the logical tactic is to run a BJS-FM type regression with the $\hat{\eta}_i$ as the dependent variable and the set of FL as the independent variables. The results are reported in Table IV, Part A. The statistics are obtained in the same manner as in Table II, Part A. For completeness, we also regress the residual from APT (Equation (3)) on the CAPM betas and the results are reported in Table IV, Part B.

The following two points (which will be made more precisely later) should be kept in mind in interpretation of the results of Table IV. (i) The prior information contained in Table II can be used to determine which significant statistic is merely a statistical artifact; specifically, any factor that is not priced in Table II should not be priced in Table IV. A case in point is the estimated λ_1 in the 1967–70 period in Table IV, Part A, which carries significant t statistics with the right sign for each market proxy. However the estimated λ_1 in Table II, Part A, is not significant. Thus the t statistics of λ_1 in Table IV, Part A, should be regarded with suspicion, and the reason for this is related to the next point. (ii) The left-hand side variable in the regression is $\hat{\eta}_i = r_i - \hat{\lambda}_0 - \hat{\lambda}_1\hat{\beta}_i$ and $\hat{\beta}_i$ and $\hat{b}_{i1}$ are highly correlated (while $\hat{\beta}_i$ and $\hat{b}_{ij}$, $j \neq 1$ are not). In the regression of $\hat{\eta}_i$ on the $\hat{b}_i$'s, the estimated λ_1 may be spuriously induced by $\hat{\beta}_i$. Note that, if $\hat{\beta}_i$ and $\hat{b}_{i1}$ are perfectly correlated, this will not happen. But when they are only highly correlated, the effect would produce some (often not too large) spurious change in the estimated coefficient but rather nonsensical t statistics.

More precisely, we first posit that expectations in the market are rational, and we write the realized return r_i as

$$r_i = E_i + \nu_i, \tag{7}$$

where E_i is the market rational expected return (before the realization) and ν_i is the noise term.[14] Let $\hat{\eta}_i$ be the residual of a cross-sectional regression of r_i on $\hat{\beta}_i$

[13] The error terms may also be heteroscedastic.

[14] Each realization of η_i may not be zero and may be correlated cross-sectionally. Thus the vector of η_i cross-sectionally may be correlated with estimated factor loadings. However, since the time

(i.e., Equation (4)), then $\hat{\eta}_i$ can be written as

$$\hat{\eta}_i = [E_i - \hat{E}_i(\text{CAPM})] + \nu_i, \tag{8}$$

where $\hat{E}_i$(CAPM) is the estimated expected return from the CAPM with the market proxies. If the CAPM is correct with the market proxies, that is, $E_i = E_i$(CAPM), then $E_i - \hat{E}_i$(CAPM) and, consequently, $\hat{\eta}_i$ should not be priced by any other models. However, if the CAPM is misspecified and E_i contains priced information that is not captured by E_i(CAPM), then $\hat{\eta}_i$ should still contain such information, and, consequently it should be priced by any model that captures this information. When $\hat{\eta}_i$ is priced by, say, a factor from the APT, it must not be spuriously induced by $\hat{E}_i$(CAPM). The prior information in Table II, Part A, enables determination of this. If a factor is priced in Table IV but not II, it is probably spurious. If a factor is priced in Table II, Part A (i.e., (7) on the FL), and again in Table IV, Part A (i.e., (8) on the FL), with the same sign and roughly the same magnitude, this pricing is not spuriously induced by $\hat{E}_i$(CAPM). The t and the Hotelling T^2 statistics then indicate individually and collectively the probability that the pricing is due to chance. Therefore, it is important to remember Parts A and B of Table IV should never be considered in isolation; rather, each should be examined with its counterpart in Table II.

With this in mind, the results in Table IV, Part A, can be examined. For the CAPM to be true, the beta must be able to capture all the priced information and render all the priced factors in Table II, Part A, insignificant in Table IV, Part A. Indeed, this seems to be the case with market proxy III (the equally weighted stock index) in the 1975–78 period. There are two priced factors in Table II, Part A, but none in Table IV, Part A (last row). However, this is the only case where the beta of the CAPM did so well. With the other proxies in the same period and with all proxies in other periods, except for factor one, the priced factors in Table II, Part A, remain priced with the same sign and roughly the same magnitude.

The Hotelling T^2 statistics are once again computed to provide summary measures. They are listed under the heading F. Because of the irregularities in estimated λ_1, the Hotelling T^2 test using only the last four risk premium time series is also reported under the heading F_4. Based on the prior information in Table II, Part A, we know F_4 will be weak because known insignificant factors have been carried along. Despite the apparent lack of power, all the F_4's are significant, except for the last period. Therefore, it can be concluded that, in most cases, the CAPM is misspecified and the missing priced information is being picked up by the APT. This also serves as a warning to all those performance measurements based on the residuals from the CAPM.

Table IV, Part B, contains the regression results of the residual from the APT on the betas of the CAPM. All the t statistics are insignificant except those for the 1971–74 period. Table II, Part B, indicates that the remaining significant t

series expectation of each η_i is zero, the time series of regression coefficient of η_i on the factor loadings should have a mean that is statistically indistinguishable from zero. That is, if the residuals of (3) or (4) contain only noise terms, they will not be priced.

Table IV

Regression of Residuals

		A. Of the CAPM on the Factor Loadings $\hat{\eta}_i = \lambda_0 + \lambda_1\hat{b}_{i1} + \lambda_2\hat{b}_{i2} + \lambda_3\hat{b}_{i3} + \lambda_4\hat{b}_{i4} + \lambda_5\hat{b}_{i5} + \xi_i$								B. Of the APT on Beta $\hat{\epsilon}_i = \lambda_0 + \lambda_1\hat{\beta}_i + \omega_i$	
		Statistics[b]						Hotelling T^2		Statistics	
Period	Market Proxy[a]	$\hat{\lambda}_0$	$\hat{\lambda}_1$	$\hat{\lambda}_2$	$\hat{\lambda}_3$	$\hat{\lambda}_4$	$\hat{\lambda}_5$	F^b	F_4^c	$\hat{\lambda}_0$	$\hat{\lambda}_1$
1963–66	I	0.0056	0.0536	−0.2113	0.3976	0.1856	−0.2569	3.81	4.75	−0.0003	0.0003
		(0.66)	(0.989)	(−1.49)	(2.91)	(1.36)	(−2.20)				(0.19)
	II	0.0073	0.0652	−0.2108	0.3866	0.1763	−0.2611	3.68	4.64	−0.0005	0.0005
		(0.87)	(1.20)	(−1.49)	(2.81)	(1.28)	(−2.23)				(0.31)
	III	0.0127	0.0961	−0.2450	0.2783	0.1093	−0.3122	3.34	4.17	−0.0008	0.0008
		(1.64)	(1.87)	(−1.69)	(1.99)	(0.81)	(−2.60)				(0.58)
1967–70	I	−0.0256	−0.1028	0.2854	0.0305	0.0173	−0.2029	2.25	2.76	0.0012	−0.0010
		(−2.35)	(−2.42)	(2.22)	(0.34)	(0.14)	(−2.25)				(−1.05)
	II	−0.0247	−0.0979	0.2863	0.0318	0.0158	−0.2035	2.32	2.76	0.0006	−0.0005
		(−2.36)	(−2.41)	(2.23)	(0.35)	(0.13)	(−2.26)				(−0.774)
	III	−0.0199	−0.0741	0.2892	0.0360	0.0113	−0.2042	2.25	2.77	−0.0011	0.0011
		(−2.34)	(−2.17)	(2.29)	(0.38)	(0.10)	(−2.27)				(1.25)
1971–74	I	0.0232	0.1029	0.1101	−0.2317	0.0721	−0.0160	2.27	2.65	−0.0083	0.0094
		(1.57)	(1.66)	(1.21)	(−2.35)	(1.11)	(−0.17)				(2.20)
	II	0.0216	0.0944	0.1150	−0.2311	0.0720	−0.0185	2.31	2.67	−0.0072	0.0077
		(1.59)	(1.66)	(1.25)	(−2.34)	(1.10)	(−0.19)				(2.21)
	III	0.0006	−0.0113	0.1320	−0.2286	0.0740	−0.0256	2.34	2.82	0.0054	−0.0051
		(0.11)	(−0.46)	(1.41)	(−2.34)	(1.15)	(−0.26)				(−1.72)
1975–77	I	−0.0226	−0.0860	−0.2938	0.0679	0.0646	0.0475	2.13	1.89	0.0036	−0.0040
		(−2.45)	(−2.41)	(−2.42)	(0.68)	(0.59)	(0.49)				(−1.63)
	II	−0.0180	−0.0615	−0.2881	0.0690	0.0591	0.0448	1.85	1.82	0.0026	−0.0027
		(−2.29)	(−2.09)	(−2.35)	(0.71)	(0.54)	(0.47)				(−1.50)
	III	0.0033	0.0409	−0.1629	0.0896	0.0093	−0.0121	0.81	0.97	−0.0052	0.0052
		(0.64)	(0.89)	(−1.44)	(0.92)	(0.08)	(−0.13)				(1.58)

[a,b] See Table II.

[c] The Hotelling T^2 statistics on the last four risk premium series, the upper percentage points of the F distribution are 2.04 (0.1 level) and 2.53 (0.05 level) for (4, 60) d.f. The $\sum T^2$ for the last four series are 49.71, 49.06, and 44.20 for market proxies I, II, and III, respectively, and they approach χ^2 with 16 d.f. and a critical value of 39.25 (0.001 level).

statistics are spurious (note the point estimates). Therefore, if the APT is misspecified, whatever it misses is not being picked up by the CAPM betas.

IV. Test of the APT against the "Own Variance" Effect

In this section we test the APT against the specific alternative that own variance has explanatory power after accounting for the FL. This alternative hypothesis is chosen because of the well-documented correlation between average returns and own variance in Douglas [15], subsequently studied by Miller and Scholes [29], Fama and Macbeth [17], and Roll and Ross [34].

Again, we do not simply enter a measure of own variance as an additional explanatory variable to Equation (3) for the same reason we did not do so with β in the previous section. Although own variance (or standard deviation) may not be as collinear with the FL as beta is, we have no way of knowing in which direction the bias will occur when multicollinearity is intertwined with measurement error in a multiple regression. In the methodology proposed below, linear programming is used to minimize the errors-in-variables problem in the FL. If it is unsuccessful, the bias is always against the null hypothesis (i.e., more likely to reject APT when it is true).[15]

To test for the own variance effect, we first compute the own variance of each asset within each subperiod using data from all even days divisible by six. All securities are separated into two groups, one consisting only of securities whose own variance is above the median variance and the other below. Then a programming technique is used to form two portfolios from the two groups so that: (1) each security in the portfolio has *nonnegative* weight, and the weight should not be too far away from $1/n$, where n is the number of securities in the portfolio; and (2) the resultant two portfolios have the same estimated FL. Ordinary linear programming was not used because it is prohibitively expensive. The elastic programming developed by Graves offers a rather inexpensive alternative (see footnote 5). A piecewise linear penalty is imposed whenever the weight of a security deviates from the target $1/n$, then the total penalty is minimized (subject to the constraints). In this way, the weights are computed and the two portfolios are formed with the same FL, but one consists only of assets with high variance and the other, low. The null hypothesis that APT is correct means that the portfolios should have insignificantly different returns. The alternative hypothesis, formulated by

$$E_i = \lambda_0 + \lambda_1 b_{i1} + \cdots + \lambda_k b_{ik} + f(\sigma_i^2)$$

or

$$E_i = g(\sigma_i^2)$$

[15] See Brown and Warner [6] for some discussions on testing hypotheses with control portfolios. Here we use a large scale linear programming technique to compute the weights of each security so that they are nonnegative *and* close to $1/n$ (n = the number of securities in the portfolio). This prevents the weight of any single security becoming so large that the idiosyncratic disturbance of that security reduces the power of our statistical tests.

where f and g are monotonic functions, would mean there should be a statistically significant difference in returns between the two portfolios. The returns of the two portfolios are computed using even days in the subperiods that are not divisible by six. The difference in returns is then computed for each of those days,[16] and the paired t statistics are computed.

The result is presented in Table V, Part A. The t statistic for each subperiod is insignificant (third column).[17] Since all the securities have a return on every trading day and there is at least one trading day separating the returns used in the portfolios, the autocorrelation of the returns of the portfolios should not be significant. To examine possible correlation between the samplings, we autoregress the difference on the first ten lags and the F statistics appear in the column headed F in Table V, Part A. If there is correlation between samplings, an adjusted t statistic can be approximated by (9) (see Roll [33]) if it is assumed that (i) the variance of the returns of the portfolio for the entire period is a good approximation to the true variance, (ii) lags beyond ten are insignificant, and (iii) the difference between the continuous and discrete compounding rate is small for one day:

$$\text{adjusted } t = t(1 + 2\rho_1 + \cdots + 2\rho_{10})^{-1/2}. \tag{9}$$

This adjusted t statistic is presented in the last column of Table V, Part A, and can serve as an approximate reference to the effect of spurious autocorrelation in the sampling. Judging from the results in Table V, Part A, we cannot reject the null hypothesis that the APT is correct, and the own variance has no explanatory power net of FL.

While the results in Table V, Part A, are sufficient for the purpose of testing the APT against the specific alternative of the own variance effect, it is interesting to determine which of the insignificant t statistics can be directly ascribed to the risk adjustment by the FL. Thus we form two equally weighted (rather than FL adjusted) portfolios, one from the high variance group and the other from the low variance group. Then we examine the time series of difference in returns and compute the paired t statistics. If the t statistic is significant for difference in the equally weighted portfolios but not in the FL adjusted portfolios, then difference in returns is accounted for by the FL.

The result of difference in return for the two equally weighted portfolios are presented in Table V, Part B. Of course, the APT would be most strongly supported if all the t statistics in Table V, Part B, were significant while all the t statistics in Table V, Part A, were insignificant. This is not the case. Only two out of four periods show significant paired t statistics for the equally weighted portfolio difference. However, for every t statistic that is significant in Table V, Part B, the corresponding t statistic in Part A is insignificant after adjusting for the FL.

[16] When return distributions are skewed, it is well known (see Miller and Scholes [29]) that spurious dependence might occur if we use the same data to compute sample mean and sample variance. This effect is avoided here by separating the data to estimate the returns and to estimate sample variance.

[17] The histogram of the sample looks symmetric and has a bell shape. The sums of t^2 are 2.54 and 1.88 for the unadjusted and adjusted t, respectively.

Table V

Difference in Mean Daily Return

A. Of Two Portfolios with Same Factor Loadings but Different Asset Own Variance

			Autocorrelation Coefficient[c]						Adjusted
Period	$\mu_d = \mu_H - \mu_L$[a]	$t(\mu_d)$[b]	1	2	3	4	5	F[d]	t[e]
1963–66	−0.000048	−0.230	−0.046	0.115	0.026	0.032	0.076	1.225	−0.188
1967–70	0.000383	1.146	0.022	0.063	−0.000	−0.020	−0.020	0.429	1.093
1971–74	0.000086	0.342	0.063	−0.031	0.058	−0.013	0.040	2.497*	0.261
1975–78	0.000236	1.045	0.044	0.070	0.030	0.107	0.035	0.655	0.757

B. Of Two Equally Weighted Portfolios Consisting of High and Low Own Variance Assets

			Autocorrelation Coefficient[c]						Adjusted
Period	$\mu_d = \mu_H - \mu_L$[a]	$t(\mu_d)$[b]	1	2	3	4	5	F[d]	t[e]
1963–66	0.000452	1.939	−0.032	−0.024	−0.011	0.090	0.015	0.634	2.142
1967–70	−0.000024	−0.070	0.087	−0.117	0.025	0.043	0.040	1.048	−0.066
1971–74	−0.000039	−0.155	0.217	0.046	0.074	0.125	0.115	2.284*	−0.095
1975–78	0.000790	2.804	0.073	0.174	−0.159	0.039	0.014	2.910*	2.235

[a] The mean daily difference between the high variance asset portfolio and the low variance asset portfolio.

[b] The paired t statistics.

[c] The autocorrelation coefficients of the difference return time series for lag 1 to 5.

[d] The F statistics for the autoregression with 10 lagged terms. Asterisk indicates significant at 0.05 level.

[e] Adjusted $t = t(\mu_d)(1 + 2\rho_1 + 2\rho_2 + \cdots + 2\rho_{10})^{-1/2}$. See text for explanation.

Table VI

Firm Size Statistics for the Portfolios[a]

	Large Firm			Small Firm		
Period	Max	Median	Min	Max	Median	Min
1963–66	34078.51	303.59	85.37	85.02	32.09	0.84
1967–70	35183.47	335.06	102.22	101.64	37.70	0.84
1971–74	24581.48	409.03	115.40	114.86	38.27	3.23
1975–78	28856.84	341.49	85.57	85.38	25.21	1.75
	First Decile			Last Decile		
Period	Max	Median	Min	Max	Median	Min
1963–66	34078.51	1297.17	750.16	12.12	7.11	0.84
1967–70	35183.47	1394.60	839.47	15.20	9.57	0.84
1971–74	24581.48	1605.44	938.54	16.45	10.81	3.23
1975–78	28856.84	1654.13	886.96	12.11	7.39	1.75

[a] Firm size in millions.

The same test is repeated for the two portfolios consisting of firms from the top and bottom decile ranked according to own variance. The results are essentially the same and therefore not reported here.[18] Overall, the tests performed in this section have provided some evidence for the usefulness of the factor loadings.

V. Test of the APT against the "Firm Size" Effect

The "small firm" effect has recently attracted wide attention in the academic literature. Empirical studies by Banz [1] and Reinganum [31] showed that small firms seem to have higher average returns than large firms after adjusting for "risk."

In this section, we determine whether the FL are sufficient to account for this effect for our sample firms. In each of the subperiods, the data in the first year were used to obtain the market value of the equity; then the even days return during the next three years was used to test for the firm size effect.

The market value data are provided by Mark Reinganum.[19] Consequently, the firms included are the intersection between his samples and mine. The number of firms for the four subperiods is 928, 1,270, 1,312, and 1,378, respectively. The number of firms in our original sample is 1,064, 1,522, 1,580, and 1,378, and in Reinganum's data set, 1,457, 1,638, 1,767, and 2,582, respectively. Some of the firm size statistics are reported in Table VI.

The firms are separated into large and small firms, and mathematical programming is used to form two portfolios with the same FL. Essentially, the procedure

[18] The results are available from the author upon request. Interested readers may also write directly to CRSP, Graduate School of Business, University of Chicago, Chicago, Illinois 60637 for Chen [9], October 1982 revision.

[19] I am indebted to Marc Reinganum for generously allowing use of his data set, which starts with firm size data at the end of 1963.

is similar to that in the preceding section. Paired t statistics are then computed using the even days during the last three years in each subperiod (even days refer to the even days of the entire subperiod). The result is presented in Table VII, Part A. In only one out of four subperiods is the difference significant at the 0.05 level, and that particular t statistic drops to 1.498 after correcting for autocorrelation. Hence, we cannot reject the null hypothesis that the firm size has no explanatory power after adjusting for risk by the FL.[20]

Again, we are curious about the behavior of equally weighted portfolios (of small firms and large firms). The results, analogous to Table V, Part B, are presented in Table VII, Part B. Perhaps not surprisingly, the t statistics are significant in only the first and the fourth period and are insignificant, as in Table V, Part B, in the middle two periods.

The same test for the top and bottom decile firms ranked according to equity value has essentially the same result and therefore is omitted here (see footnote 18). In this case, the paired t statistics (both unadjusted and adjusted for autocorrelation), which measure the return difference between the top and bottom decile firms after correcting for the FL, are all insignificant. Therefore, our conclusion is that firm size does not have additional explanatory power after risk is adjusted by the FL.

VI. Conclusion

Based on the empirical evidence gathered so far, the APT cannot be rejected in favor of any alternative hypothesis, and the APT performs very well against the CAPM as implemented by the S&P 500, value weighted, and equally weighted indices. Therefore, the APT is a reasonable model for explaining cross-sectional variation in asset returns.

In perspective, this study can be regarded as a moderate step toward solving the problem of what determines the expected return of assets. There are two, somewhat equivalent, ways to solve that problem: we can make assumptions and produce a theory that specifies which variables should enter the pricing equation and then test it; or, we can examine assets' realized return and determine empirically to which macro variables (suggested by theories) they correspond. The APT is more in the spirit of the second approach. The computation of the FL in this paper would enable construction of a portfolio corresponding to each of the common factors. Of course, some idiosyncratic term may still remain, but by constructing these large portfolios, each consisting of over a thousand securities, we can obtain time series (with noise) of the behavior of the common factors and can match these against time series behavior of global economic variables such as industrial production, interest rates, and so on. This is probably the most important direction for future research—an economic interpretation of the common factors.[21]

[20] The sums of t^2 are 6.45 and 3.85 for the unadjusted and adjusted t, respectively, in Table VII, Part A.

[21] Preliminary results have been obtained; see Chen et al. [11] and Chan et al. [7].

Table VII

Difference in Mean Daily Return

A. Of Two Portfolios with Same Factor Loadings but Different Firm Size

Period	$\mu_d = \mu_L - \mu_S$[a]	$t(\mu_d)$[b]	Autocorrelation Coefficient[c] 1	2	3	4	5	F[d]	Adjusted t[e]
1963–66	−0.000008	−0.076	0.001	0.046	0.135	−0.006	0.093	1.530	−0.061
1967–70	−0.000132	−0.809	0.084	0.080	0.006	−0.009	0.016	0.800	−0.610
1971–74	−0.000227	−1.121	−0.022	−0.051	0.014	−0.034	−0.022	1.222	−1.110
1975–78	−0.000407	−2.133	0.093	−0.013	0.140	0.004	−0.015	1.548	−1.498

B. Of Two Equally Weighted Portfolios Consisting of Large and Small Firm Size Securities

Period	$\mu_d = \mu_L - \mu_S$[a]	$t(\mu_d)$[b]	Autocorrelation Coefficient[c] 1	2	3	4	5	F[d]	Adusted t[e]
1963–66	−0.000341	−2.525	0.054	0.102	0.012	−0.060	0.080	1.833	−1.996
1967–70	−0.000037	−0.218	0.154	0.193	−0.006	0.114	−0.000	2.548*	−0.139
1971–74	0.000212	0.995	0.176	0.002	0.062	−0.099	0.072	2.078*	0.786
1975–78	−0.000652	−3.426	0.129	0.020	0.274	0.006	0.070	4.119*	−2.093

[a] The mean daily difference between the large firm portfolio and the small firm portfolio.
[b] The paired t statistics.
[c] The autocorrelation coefficients of the difference return time series for lag 1 to 5.
[d] The F statistics for the autoregression with 10 lagged terms. Asterisk indicates significant at 0.05 level.
[e] Adjusted $t = t(\mu_d)(1 + 2\rho_1 + 2\rho_2 + \cdots + 2\rho_{10})^{-1/2}$. See text for explanation.

Appendix

A. Extension of the Factor Structure from a Subset to the Entire Sample

Theorem: Let $\tilde{r}_i$, $i = 1, \cdots, k + 1$ be the returns of $k + 1$ linearly independent assets (portfolios) such that b_{ij}, the loadings corresponding to a set of factors $\tilde{\delta}_j$, $j = 1, \cdots, k$, are known. Define $\sigma_j^2 = \text{Var}(\tilde{\delta}_j)$, $j = 1, \cdots, k$. Let r_p be the return of the p^{th} asset, $p \neq i$, $i = 1, \cdots, k + 1$ with $\text{Cov}(\tilde{\epsilon}_p, \tilde{\epsilon}_i) = 0$.

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

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

Proof:

$$\text{(i) } \text{Cov}(\tilde{r}_p, \tilde{r}_i) = b_{p1}b_{i1}\text{Var}(\tilde{\delta}_1) + \cdots b_{pk}b_{ik}\text{Var}(\tilde{\delta}_k)$$
$$= b_{p1}b_{i1}\sigma_1^2 + \cdots + b_{pk}b_{ik}\sigma_k^2,$$
$$i = 1, \cdots, k.$$

The above k linear equations can be written as

$$\begin{bmatrix} b_{11}\sigma_1^2 & b_{12}\sigma_2^2 & \cdots & b_{1k}\sigma_k^2 \\ b_{21}\sigma_1^2 & & & \cdot \\ \cdot & & & \\ b_{k1}\sigma_1^2 & b_{k2}\sigma_2^2 & & b_{kk}\sigma_k^2 \end{bmatrix} \begin{bmatrix} b_{p1} \\ b_{p2} \\ \cdot \\ b_{pk} \end{bmatrix} = \begin{bmatrix} \text{Cov}(\tilde{r}_p, \tilde{r}_1) \\ \cdot \\ \cdot \\ \text{Cov}(\tilde{r}_p, \tilde{r}_k) \end{bmatrix}. \quad \text{(A1)}$$

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

(ii) See Chen [8, 9]. Q.E.D.

B. Estimation of the Factor Loadings

Initially, the first 180 stocks in the sample (alphabetically) were selected and their sample covariance matrix was computed. Factor loadings of the 180 stocks were computed with the Jöreskog's method, an asymptotic maximum likelihood method, in the computer software package EFAP II, which was also written by Jöreskog. The first *ten* factor loadings for each stock were obtained. Based on the computed factor loadings, *five* portfolios were formed. The time series of the five portfolios will contain linear combinations of the first five factor scores.

The portfolio formation was done using the mathematical programming GUB developed by Graves (see footnote 5). The goal of the programming is to form portfolios so that their associated factors loadings possess some desirable properties which would enable solution of Equation (A1) with numerical stability.[22] In this study, the first portfolio was formed using all 180 stocks with the constraints that $b_2 = \cdots = b_{10} = 0$ for the resultant portfolio.[23] As for the j^{th}

[22] See Chen [8, 9] for a discussion of the numerical stability problem.

[23] The extra constraints $b_6 = \cdots = b_{10}$ were included to ensure the unbiasedness of the estimated FL in case there exists more than five factors (up to ten).

portfolio, $j = 2, \cdots, 5$, we first rank the j^{th} factor loadings of the 180 stocks then select the 100 stocks with the largest j^{th} factor loadings[24] and then form the j^{th} portfolio using those 100 stocks, with the constraints $b_k = 0$, $k \neq 1$, and $k \neq j$ for the resultant portfolio.[25] The objective function of the programming is the sum of piecewise linear penalty functions that impose a positive penalty whenever the absolute deviation of an asset's weight (in the portfolio) from $1/n$ exceeds a certain bound (where n is the number of securities in the portfolio). By minimizing the objective function, the weight of each security would be nonnegative and close to $1/n$.[26] The penalty function $p(x_i)$, where x_i is the weight of the i^{th} asset, is defined by

$$p(x) = \begin{cases} \infty & x < 0 \\ 0.8/n - x & 0 \le x < 0.8/n \\ 0 & 0.8/n \le x \le 1.2/n \\ x - 1.2/n & 1.2/n < x \end{cases}$$

for portfolio one, with $n = 180$; and by

$$p(x) = \begin{cases} \infty & x < 0 \\ 0.75/n - x & 0 \le x \le 0.75/n \\ 0 & 0.75/n \le x \le 1.25/n \\ x - 1.25/n & 1.25/n < x \end{cases}$$

for portfolio j, $j = 2, \cdots, 5$, with $n = 100$.

After the five portfolios are formed, (A1) can be used to solve for consistent estimates of factor loadings of each security.[27]

[24] Almost all factor loadings for the first factor are negative. For the second to the tenth, they are roughly half positive and half negative.

[25] Here we ranked and computed the b_{ij}, $i = 1, \cdots, 180$ with the same data. This would aggravate the error-in-variable problem in the diagonal entries of the result coefficient matrix except the (1, 1) element. The consequence of such would be similar to the error in σ_j^2 and therefore unimportant. The resultant matrix has all zero entries except the main diagonal and the first column.

[26] The estimation errors for the factor loadings are not independent. The most efficient aggregation of coefficients would require inversion of a large matrix. Owing to the cost and precision of such an algorithm, we selected the simpler alternative of weighting the securities approximately equally to diversify the estimation error.

[27] It can be demonstrated that running OLS on portfolios constructed from the initial 180 securities to obtain factor loadings for all assets is similar to a special case of our algorithm. Interested readers may consult Chen [9].

REFERENCES

1. Rolf Banz. "The Relationship between Return and Market Value of Common Stocks." *Journal of Financial Economics* 9 (1981), 3–18.
2. F. Black. "Capital Market Equilibrium with Restricted Borrowing." *Journal of Business* 45 (1972), 444–54.
3. ———, M. Jensen, and M. Scholes. "The Capital Asset Pricing Model: Some Empirical Results." In Michael C. Jensen (ed.), *Studies in the Theory of Capital Market.* New York: Praeger, 1972, 79–121.
4. Douglas T. Breeden. "An Intertemporal Asset Pricing Model with Stochastic Consumption and Investment Opportunities." *Journal of Financial Economics* 7 (1979), 265–96.
5. S. Brown and M. Weinstein. "A New Approach to Testing Asset Pricing Models: The Bilinear Paradigm." *Journal of Finance*, 38 (1983), 711–43.

6. ——— and J. Warner. "Measuring Security Price Performance." *Journal of Financial Economics* 8 (1980), 205–58.
7. C. Chan, N. Chen, and D. Hsieh. "An Exploratory Investigation of the Firm Size Effect." unpublished manuscript, University of Chicago, November 1982.
8. N. Chen. "Arbitrage Asset Pricing: Theory and Evidence." unpublished Ph.D. Dissertation, UCLA, 1981.
9. ———. "Some Empirical Tests of the Theory of Arbitrage Pricing." Working Paper No. 69, October Revision, University of Chicago, 1982.
10. ——— and J. Ingersoll. "Exact Pricing in Linear Factor Models with Finitely Many Assets." *Journal of Finance*, 38 (1983), 985–88.
11. ———, R. Roll, and S. Ross. "Economic Forces and the Stock Market," unpublished manuscript, Yale University, 1983.
12. George Constantinides. "Admissible Uncertainty in the Intertemporal Asset Pricing Model." *Journal of Financial Economics* 8 (1980), 71–86.
13. Gregory Connor. "A Factor Pricing Theory of Capital Assets." Working Paper, Northwestern University, 1982.
14. R. Davidson and J. Mackinnon. "Several Tests for Model Specification in the Presence of Alternative Hypotheses." *Econometrica* 49 (1981), 781–93.
15. George W. Douglas. "Risk in the Equity Market: An Empirical Appraisal of Market Efficiency." *Yale Economic Essays* 9 (1969).
16. Philip H. Dybvig. "An Explicit Bound on Individual Assets' Deviation from APT Pricing." Working Paper, Yale University, 1983.
17. Eugene Fama and James Macbeth. "Risk, Return and Equilibrium: Empirical Tests." *Journal of Political Economy* 81 (1973), 607–36.
18. K. M. Gaver and M. S. Geisel. "Discriminating Among Alternative Models: Bayesian and Non-Bayesian Models." In P. Zarembka (ed.), *Frontier in Econometrics*. New York: Academic Press, 1973.
19. Adam Gehr, Jr. "Some Tests of the Arbitrage Pricing Theory." *Journal of the Midwest Finance Association* (1975), 91–105.
20. Michael Gibbons. "Multivariate Tests of Financial Models: A New Approach." *Journal of Financial Economics* 10 (1982), 3–27.
21. ———. "Estimating the Parameters of the Capital Asset Pricing Model: A Minimum Expected Loss Approach." unpublished manuscript, Graduate School of Business, Stanford University, 1980.
22. Mark Grinblatt and Sheridan Titman. "Factor Pricing in a Finite Economy." Working Paper, UCLA, 1983.
23. Gur Huberman. "Arbitrage Pricing Theory: A Simple Approach." *Journal of Economic Theory* 28 (1982), 183–91.
24. J. Ingersoll. "Some Results in the Theory of Arbitrage Pricing." Working Paper No. 67, University of Chicago, 1982.
25. Edward Leamer. *Specification Searches*. New York: John Wiley and Sons, 1978.
26. John Lintner. "The Valuation of Risk Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets." *Review of Economics and Statistics* 47 (1965), 13–37.
27. Robert C. Merton. "An Inter-Temporal Capital Asset Pricing Model." *Econometrica* 41 (1973), 867–87.
28. Donald Morrison. *Multivariate Statistical Methods*. New York: McGraw-Hill, 1976.
29. Merton H. Miller and Myron S. Scholes. "Rates of Return in Relation to Risk: A Re-examination of Some Recent Findings." In Michael C. Jensen (ed.), *Studies in the Theory of Capital Markets*, New York: Praeger, 1972.
30. Marc Reinganum. "The Arbitrage Pricing Theory: Some Empirical Results." *Journal of Finance* 36 (1981), 313–21.
31. ———. "Misspecification of Capital Asset Pricing: Empirical Anomalies Based on Earnings' Yields and Market Values." *Journal of Financial Economics* 9 (1981), 19–46.
32. Richard Roll. "A Critique of the Asset Pricing Theory's Tests." *Journal of Financial Economics* 4 (1977), 120–76.
33. ———. "A Possible Explanation of the Small Firm Effect." *Journal of Finance* 36 (1981), 879–88.

34. ——— and S. Ross. "An Empirical Investigation of the Arbitrage Pricing Theory." *Journal of Finance* 35 (1980), 1073–1103.
35. Stephen A. Ross. "Return, Risk and Arbitrage." In Irwin Friend and James Bickler, (eds.), *Risk and Return in Finance*, Vol. I, Cambridge, MA.: Ballinger, 1977.
36. ———. "The Arbitrage Theory of Capital Asset Pricing." *Journal of Economic Theory* 13 (1976), 341–60.
37. J. Shanken. "The Arbitrage Pricing Theory: Is It Testable." *Journal of Finance* 37 (1982), 1129–40.
38. ———. "An Asymptotic Analysis of the Traditional Risk-Return Model." unpublished Manuscript, University of California, Berkeley, 1982.
39. William F. Sharpe. "Capital Asset Prices: A Theory of Market Equilibrium Under Conditions of Risk." *Journal of Finance* 19 (1964), 425–42.
40. A. Zellner. *An Introduction to Bayesian Inference in Econometrics*. New York: John Wiley and Sons, 1971.
41. ———. "Posterior Odds Ratios for Regression Hypotheses: General Considerations and Some Specific Results." Working Paper, University of Chicago, 1979.