GOVERNMENT BOND RETURNS, MEAUREMENT OF INTEREST RATE RISK, AND THE ARBITRAGE PRICING THEORY

by

N. Bulent Gultekin and Richard J. Rogalski

Working Paper No. 13-84
(Revision of 21-83)

THE WHARTON SCHOOL

University of Pennsylvania

Philadelphia, PA 19104

The Pricing of Treasury Bond Futures:
The Quality Variation Option

By

Simon Benninga
and
Michael Smirlock

Working Paper No. 12-84

RODNEY L. WHITE CENTER
FOR FINANCIAL RESEARCH

University of Pennsylvania
The Wharton School
Philadelphia, PA 19104

GOVERNMENT BOND RETURNS, MEASUREMENT OF INTEREST RATE RISK, AND THE ARBITRAGE PRICING THEORY

by

N. Bulent Gultekin*

Richard J. Rogalski**

December, 1979

Seventh Revision: May 1984

Forthcoming in The Journal of Finance

*The Wharton School, University of Pennsylvania. **The Amos Tuck School of Business Administration, Dartmouth College. This paper was initially presented at the American Finance Association annual meeting in Atlanta, Georgia, December 28-30, 1979.

Subsequent comments by Michael J. Brennan, Edwin J. Elton, and the last of three anonymous referees have been especially helpful. We have also benefitted from comments by Marshall E. Blume, Stephen Brown, Eugene F. Fama, Wayne Ferson, Dennis E. Logue, Richard Roll, Seha Tinic and Randolph Westerfield. Dorothy Bower and Mustafa N. Gultekin provided valuable research assistance and computer programming skills far beyond the call of duty. Partial funding for this study came from the Rodney L. White Center for Financial Research and the Tuck Associates Research Program.

ABSTRACT

Empirical tests are reported for Ross' arbitrage pricing theory using monthly data for U.S. Treasury securities during the 1960-1979 period. We find that mean returns on bond portfolios are linearly related to at least two factor loadings. Multivariate test results, however, are not consistent with the APT. Our sample data in the U.S. Treasury securities market are also not consistent with either version of the CAPM. One-month-ahead forecasts of excess returns using factor-generating models are compared with corresponding naive predictions or predictions using the "market model" with various market portfolios.

GOVERNMENT BOND RETURNS, MEASUREMENT OF INTEREST RATE RISK, AND THE ARBITRAGE PRICING THEORY

I. INTRODUCTION

Bond risk is typically determined by considering bonds and bond portfolios much as one would stocks and stock portfolios and by assessing risk with the capital asset pricing model (CAPM). As CAPM should apply to all kinds of traded securities, one would expect application of the CAPM to bond markets to be straightforward.

Empirical application of CAPM to bonds, however, has for various reasons not been as effective as the theory would lead us to expect. One problem, pointed out by Roll [15], may lie in choice of market portfolio. Another important problem neglected in previous empirical work is peculiar to the very nature of a bond. Boquist, Racette, and Schlarbaum [2] show that the beta of a default-free bond is nonstationary because the beta of a default free bond is proportional to its duration and is a function of the covariance of interest rate movements with the market return. Duration itself is a function of the term to maturity of a bond, all else being equal.

This paper examines the factor structure of U.S. Treasury security returns and tests the Arbitrage Pricing Theory (APT) developed by Ross [17, 18] in the U.S. Treasury security market. We want to compare the empirical performance of the APT with that of the CAPM in the U.S. Treasury security market during the twenty year sample period 1960-1979. Section II briefly discusses the Arbitrage Pricing Theory. In Section III we describe our data and estimate the factor structure of Treasury security returns. Section IV details our test procedures, and empirical results are presented. A final section summarizes our findings.

II. THE ARBITRAGE PRICING THEORY

The arbitrage pricing theory (APT) as developed by Ross [17, 18] assumes that financial markets are perfectly competitive and frictionless and the returns on risky assets are linearly related to a limited number of common factors. More specifically, the return R_i on any asset i may be written

$$\tilde{R}_i = E(\tilde{R}_i) + \sum_{j=1}^{k} b_{ij} \tilde{\delta}_j + \tilde{\varepsilon}_i , \tag{2.1}$$

where $E(\tilde{R}_i)$ is the expected return on asset i with $i = 1,\ldots,M$, $\tilde{\delta}_j$ is a mean zero factor common to the returns of all risky assets p, b_{ij} is the sensitivity of the return of asset i to the movements of the j^{th} factor, and $\tilde{\varepsilon}_i$ is a mean zero noise term assumed to be sufficiently independent that the law of large numbers works. In more traditional terminology the b's represent systematic risk factors, and $\tilde{\varepsilon}_i$ is the unsystematic component of risk specific to the i^{th} asset.

Given the return-generating process in 2.1, Ross [17] develops the APT by setting up an "arbitrage" portfolio where the wealth invested in assets is exactly balanced by the amount borrowed from short sales. Thus, for such a portfolio there is zero investment and no systematic risk; that is, the b's for all factors are zero. The only risk in such a portfolio is the noise term, $\tilde{\varepsilon}_i$. If the arbitrage portfolio is sufficiently large, however, such unsystematic risks become negligible. As a result the return on such a portfolio in equilibrium should be approximately zero.[1]

Ross formalizes such arguments to show that if the number of assets, m, is sufficiently large, there exist k+1 weights, $\lambda_0, \lambda_1, \ldots, \lambda_k$, such that an equilibrium pricing relationship for all i is

$$E(\tilde{R}_i) \approx \lambda_0 + \sum_{j=1}^{k} b_{ij}\lambda_j . \tag{2.2}$$

λ_0 is the riskless rate if such an asset exists, or under certain conditions, it is the return on a "zero-factor" asset.[2]

The relation in 2.2 can also be rewritten in terms of expected excess returns by forming portfolios with unit systematic risk on each factor and no risk on other factors. Then each λ_j can be interpreted as the excess return or risk premium on portfolios with only systematic j risk.

An allegedly important advantage of the APT over the CAPM for empirical work is that the APT does not rely directly on a market portfolio. According to Roll and Ross [16], for example, problems associated with misspecification of the market portfolio in testing the CAPM (see Roll [15]) may be circumvented in testing the APT. Another property of the APT is that it will hold for subsets of assets conforming to its assumptions even if some groups of assets do not. This is particularly appealing in our case since we focus solely on the government securities market, a subset of the universe of risky assets.[3]

III. DATA AND FACTOR STRUCTURE

3.1 Treasury Security Data

All of our data are from the CRSP U.S. Government Bond File. We use return data on all securities in the CRSP file except securities with special tax or call provisions for the twenty-year period 1960-1979.

For our tests of the APT it is necessary to construct bond portfolios in which the maturity of the portfolio is approximately constant each month. Constant maturity portfolios help to alleviate the nonstationarity of beta encountered in previous tests of capital market theory using bond data.

We constructed two samples of constant maturity portfolios, each consisting of all securities within a given maturity range. The return over any maturity range is then measured as a simple average of all the individual returns that fall within the range.

Sample 1 consists of twelve portfolios. Maturity ranges were selected purposely to avoid any missing observations in the CRSP Bond File. Column 2 of Table 1 lists the ranges in days to maturity for the twelve portfolios. Thirty-day increments are taken from 60 days up to a half year, 90-day increments up to about a year, and 360-day increments thereafter to about six years. The final portfolio includes all securities with maturities of more than six years. Finer increments are used for small maturities because the government has more short-term than long-term debt outstanding and the yield curve is considerably more volatile for short-term maturities while it is usually flat for maturities five years and above. For the same reason the six-year cutoff ensures that each maturity range has at least one security in it.

The second sample is exactly the same out to 360 days, but beyond that time maturity ranges fall in 180-day increments. Sample 2 consists of sixteen portfolios at the expense of some missing observations for portfolios in the long maturity ranges.[4] While we were successful in constructing portfolios with very few or no missing observations, we paid an unavoidable price in terms of limiting the number of portfolios that can be constructed from the available data.

Summary statistics for the portfolios in Sample 1 are presented in Table 1. All returns are expressed as continuously compounded <u>excess</u> returns with monthly holding periods. Columns 1 and 2 identify the portfolio by maturity range. Columns 3 to 6 give the mean, variance, skewness, and kurtosis, in that order. There are 240 monthly observations for each portfolio. Column 7 gives the value of the studentized range, an overall test of the normality of the return distribution. In columns 8 to 12 of Table 1 sample autocorrelations for monthly returns have been computed for lags of from one to four months and also for twelve months.

Panel A is for Treasury portfolio returns. The monthly return variances increase with maturity, but the mean returns do not increase correspondingly. Portfolios 2, 3, 4, and 12 are skewed; the rest are not. All the portfolios have more central area than a normal distribution. All but one of the autocorrelations in columns 8 to 12 are insignificantly different from zero.

Panel B gives corresponding summary statistics for typical index portfolios. SP500, NYSE, EW, and (NYSE+EW)/2 denote the Standard & Poor 500 Index, the CRSP New York Stock Exchange equally weighted index, an equally weighted index of all Treasury securities on the CRSP Government Bond Tape, and a simple average of NYSE and EW. Skewness and autocorrelation do not seem to be a problem, but the index return distributions with the exception of the SP500 are not normally distributed.

Table 2 contains estimated variances, covariances, and correlations between the excess returns of the various portfolios in Panel A of Table 1. Variances are on the diagonal, covariances are above the diagonal, and correlation coefficients are below the diagonal. Treasury portfolios with similar maturity ranges are highly correlated, especially adjacent portfolios, which suggests that Treasury portfolios of about the same maturity range may be substitutes. Table 2 also indicates that shorter maturity portfolios have a relatively lower association with the longer-term portfolios.

A fundamental question that is logically prior to all the tests in this paper is whether the mean excess returns on the bond portfolios in panel A of Table 1 are not equal across the portfolios. If it turns out that the relevant multivariate statistic cannot reject equality of means, then we know there is nothing for asset pricing models to explain.

To test the equality of mean excess returns across all bond portfolios we use Hotelling's T^2 statistic. The null hypothesis is that the means in column (3) (in panel A of Table 1) for bond portfolios 2 to 13 are equal. Hotelling's T^2 value for the 240 sample observations is 201.20 and the corresponding F statistic with 12 and 228 degrees of freedom is 15.99. Thus, it is concluded that the average excess returns are different for the twelve bond portfolios in Sample 1. The same (unreported) conclusion is reached for the seventeen bond portfolios in Sample 2.[5]

3.2 Factor Structure of Treasury Security Returns

The first stage of the tests of the APT requires the determination of the number of factors and the estimation of factor loadings. We factor-analyze excess returns separately for one factor through the number needed to accept the null hypothesis that there are k-factors using Barlett's chi-square test.[6] Up to five factors are obtained for Sample 1 and up to seven for Sample 2. We will refer to the five factors from Sample 1 as Factor Set 1 and the seven factors from Sample 2 as Factor Set 2. The associated values of the chi-square test statistics are shown in Table 3 for both samples. Five and seven factors explain about 99% of the total variation in both samples. Since the APT model does not specify the number of factors, we use the five- and seven-factor models so that we may ascertain in our tests the appropriate number of factors.

Table 3 contains estimates of the factor loadings using excess returns for a five-factor model for Samples 1 and 2 in Panels A and B, respectively. The estimates of the first factor loadings for both groups have the same sign and increase in absolute magnitude as portfolio maturity increases. The estimates for the remaining factors display scattered signs and magnitudes. These factor loadings in Table 3 are estimates of the b_{ij} in 2.2 using excess returns. Gibbons [7] reports similar findings for a slightly longer time period.

IV. TESTS OF THE ARBITRAGE PRICING THEORY AND CAPM

4.1 Two-Stage Methodology

The APT model in 2.1 and 2.2 is essentially a one period model. In order to give empirical content to the model, we shall assume that the return generating model in 2.2 holds for excess returns each time period t. Thus we can rewrite 2.1 in matrix notation as

$$(4.1) \qquad \tilde{\mathbf{r}}_t = \mathbf{E}_t + \mathbf{B}\tilde{\boldsymbol{\delta}}_t + \tilde{\boldsymbol{\varepsilon}}_t \ ,$$

where tildes denote random variables, $\tilde{\mathbf{r}}$ is an M-element column vector of observed excess returns, and $\mathbf{E}$ is also an M-element column vector containing expected returns. $\mathbf{B}$ is an (M×k) matrix containing factor loadings b_{ij}, and $\boldsymbol{\delta}_t$ is a k-element column vector of unobserved common factors. Finally, $\tilde{\boldsymbol{\varepsilon}}_t$ is a sequence of independent identically distributed (i.i.d.) random vectors such that

$$(4.2) \qquad E(\tilde{\boldsymbol{\varepsilon}}_t) = \mathbf{0} \quad , \quad \mathrm{Cov}(\tilde{\boldsymbol{\varepsilon}}_t) = \mathbf{D} \ ,$$

covariance matrix D being diagonal with finite elements. The APT requires that

$$(4.3) \qquad \mathbf{E}_t \simeq \lambda_{ot}\boldsymbol{\iota}_M + \mathbf{B}\boldsymbol{\lambda}_t \ ,$$

where $\boldsymbol{\iota}_M$ is an M-element column of ones. If the APT is true, we can substitute 4.3 in 4.1 and obtain (ignoring the approximation):

$$(4.4) \qquad \tilde{\mathbf{r}}_t = \lambda_{ot}\iota_M + \mathbf{B}\boldsymbol{\lambda}_t + \mathbf{B}\tilde{\boldsymbol{\delta}}_t + \tilde{\boldsymbol{\varepsilon}}_t .$$

Since neither $\mathbf{B}$ nor $\tilde{\boldsymbol{\delta}}_t$ is directly observable, we specify $\tilde{\boldsymbol{\delta}}_t$ to be i.i.d. through time with zero mean, unit variances and zero contemporaneous correlation. This implies that $(\tilde{\mathbf{r}}_t - \mathbf{E}_t)$ is a sequence of random vectors with

$$(4.5) \qquad E(\tilde{\mathbf{r}}_t - \mathbf{E}_t) = 0 \quad \text{and} \quad \mathrm{Cov}\left[(\tilde{\mathbf{r}}_t - \mathbf{E}_t)\right] = \mathbf{BB}' + \mathbf{D} = \mathbf{V} .$$

The test methodology is based on a two-stage-factor analytic approach as implied by the formulation of the APT in 4.4 and 4.5. In the first step the number of factors, k, is determined, and the elements of the factor loadings matrix, B, are estimated. In the second stage we estimate risk premia, $\boldsymbol{\lambda}_t$, using the estimated matrix B as independent variables. This is an adaptation of the Fama-MacBeth [5] methodology to a factor-analytic framework.

If we assume that T, the time dimension of our sample, is large enough, we can estimate the covariance matrix, $\mathbf{V}$, without sampling error. Using $\hat{\mathbf{V}}$, we estimate factor loadings by $\hat{\mathbf{B}}$ and specific variances by $\hat{\mathbf{D}}$. That is, the following tests are conditional on $\mathbf{V} = \hat{\mathbf{V}}$ and $\mathbf{B} = \hat{\mathbf{B}}$. In the first stage, therefore, we estimate

$$(4.6) \qquad \hat{\mathbf{V}} = \hat{\mathbf{B}}\hat{\mathbf{B}}' + \hat{\mathbf{D}} .$$

In the second stage, for <u>each</u> time period t, we obtain a GLS estimate of regression coefficients in 4.4:

$$(4.7) \qquad \hat{\boldsymbol{\lambda}}^*_t = (\hat{\mathbf{B}}^{*\prime}\hat{\mathbf{V}}^{-1}\hat{\mathbf{B}}^*)^{-1}\,\hat{\mathbf{B}}^{*\prime}\hat{\mathbf{V}}^{-1}\tilde{\mathbf{r}}_t , \qquad t = 1,2,\ldots,T ,$$

where $\hat{\boldsymbol{\lambda}}^*_t = (\hat{\lambda}_{ot}, \hat{\boldsymbol{\lambda}}_t)$, λ_{ot} is the estimate of the intercept term and $\hat{\boldsymbol{\lambda}}_t$ is the estimate of risk premia vector. $\mathbf{B}^*$ is $[\iota_M{:}\mathbf{B}]$, so that $\boldsymbol{\lambda}^*_t$ contains an estimate for λ_{ot} as its first element. If the underlying error process admits of a central limit theorem,

$$(4.8) \qquad \sqrt{T}\,(\hat{\boldsymbol\lambda}^*_t - \boldsymbol\lambda^*_t) \sim N[0,\ (B^{*\prime}V^{-1}B^*)^{-1}]\ .$$

There are two testable implications of the APT using the two-step procedure. First, mean returns are linearly (approximately) related to factor loadings, that is, the risk premia are significant or "priced". Second, the intercept term is the risk free (or "zero beta") return.

Dhrymes, Friend, and Gultekin [4] suggest a test of joint significance of risk premia since factor loadings can be identified only up to left multiplication by an orthogonal matrix. A test statistic for the joint significance of risk premia hypothesis, provided that the number of factors, k, is determined, is

$$(4.10) \qquad T\bar{\boldsymbol\lambda}'\boldsymbol\Phi^{-1}\bar{\boldsymbol\lambda} \sim \chi^2_k\ ,$$

where $\bar{\boldsymbol\lambda} = \frac{1}{T}\sum_{t=1}^{T}\hat{\boldsymbol\lambda}_t$ and $\boldsymbol\Phi = \frac{1}{T}\sum_{t=1}^{T}(\hat{\boldsymbol\lambda}_t - \bar{\boldsymbol\lambda})(\hat{\boldsymbol\lambda}_t - \bar{\boldsymbol\lambda})'$. The test statistic is asymptotically chi-square with k degrees of freedom. (See Dhrymes [3], chapter 4, and Dhrymes, Friend and Gultekin [4]).

4.2 Two Stage Test Results

Significance tests of the risk premia for various factor models are presented in Table 4. Sample 1 and Sample 2 results are presented in Panels A and B, respectively. The first column gives the number of factors estimated, column 2 gives the mean of the cross sectional regression estimates of the intercept term, and columns 3 to 9 give the means of the estimates of risk premia in a sequential order from the one-factor model up to the five- or seven-factor model. That is, columns 2 to 9 represent the monthly risk premia estimates in 4.7 averaged over all 240 months in the sample period. Column 10 is the t-value for testing the significance of the intercept term under the null hypothesis that it equals zero. The chi-square values in column 11 are calculated using 4.10 and allow us to test the joint significance of risk premia.

A review of Table 4 reveals that the null hypothesis of no linear relation cannot be rejected for a one-factor model based on the small and insignificant chi-square values of .054 and .153 for Samples 1 and 2 in column 11. The opposite conclusion is reached for two or more factors. That is, once the number of factors is increased beyond one, the chi-square values become significant at the 5% level and do not comply with the null hypothesis. These findings for multiple factors are consistent with one of the implications of the APT and suggest two or more "priced" factors in the U.S. Treasury securities market.

A second testable implication of the APT evolves from the intercept terms in column (2) of Table 4. The intercept terms should be zero for a risk-free version of APT or positive for a zero beta APT model. For all factor models, the intercept terms are significantly greater than zero at the 1% level of significance. These positive intercept terms using excess returns are consistent with a zero beta APT model but not a risk-free rate version.[7]

In Table 5 we present similar two-stage test statistics for the CAPM as a comparison to the APT model.[8] To our knowledge, these are the first careful tests of the CAPM in the Treasury securities market in which portfolio maturities are fixed constant each time period in order to avoid nonstationarity of bond betas. The results in Table 5 are reported for excess returns in panel A and real returns in panel B. Tests based on real returns are included to avoid any potential bias in estimating betas.[9]

One testable implication of the Sharpe-Lintner version of CAPM is that the intercept term should be zero using excess returns or the average real 30-

day T-bill return using real returns. In Black's version of CAPM the expected zero beta return is greater than the risk-free rate. Positive intercept terms are consistent with Black's version. The intercept term t-values in Panel A of Table 5 are all positive and significant at the 5% level under a null hypothesis of zero. (See Column 5). In Panel B using real returns we entertain the null hypothesis that the intercept term is the average (real) risk-free rate. The relevant t-values given in column 7 are all significantly positive. These intercept tests are not supportive of the Sharpe-Lintner version of the CAPM but are consistent with Black's version.

Another testable implication of both the Sharpe-Lintner and Black versions is that the risk premium (in this case an estimate of the excess or real return on the market portfolio) should be positive. Estimates of the risk premia on the market portfolio in column 6 of Table 5 are all negative but not significantly different from zero at the 5% level. The evidence is the same for excess returns (panel A) and real returns (panel B) regardless of which market portfolio proxy is used. This means that there is no statistically observable positive relationship between Treasury portfolio returns and beta risk which is inconsistent with both versions of the CAPM.

Some information about the factor structure under consideration can be obtained by correlating the factor loadings with the betas relative to each market index. The resulting correlations turn out to be similar for excess returns and real returns regardless of the market index used. To illustrate, the estimated correlation coefficients for Sample 1 using excess returns are .9908, -.9325, .0262, .5347, .7456 between factor loadings 1 to 5 and the EW market portfolio beta estimates.[10] Thus the loadings of a one-factor model are highly positively correlated with the market betas. This means that the chi-square test in Table 4 that led us to conclude a single factor is not priced is consistent with the evidence in Table 5 about the pricing of the market over the same sample period.

In summary, mean returns on bond portfolios are linearly related to at least two factor loadings. The empirical evidence from the two-stage methodology is not, however, consistent with either version of the CAPM in the U.S. Treasury securities market.

4.3 Multivariate Methodology[11]

Another test of the APT model versus the CAPM is performed using "stacked" time series regressions. The relevant econometric analysis for the linear case is detailed in Theil [22]. For testing the Sharpe-Lintner [20, 13] version of the CAPM the appropriate statistical model is

$$(4.11) \qquad \tilde{\mathbf{r}}_i = \alpha_i^{(1)} \boldsymbol{\iota}_T + \beta_i \tilde{\mathbf{r}}_m + \tilde{\boldsymbol{\eta}}_i^{(1)}, \qquad i = 2, \ldots, M .$$

Tildes denote random variables. $\mathbf{r}_i$ is a T-element column vector of excess returns, $\boldsymbol{\iota}_T$ is a T-element column vector of ones, $\mathbf{r}_m$ is a T-element column vector of returns on a proxy market portfolio, and $\boldsymbol{\eta}_i^{(1)}$ is a T-element vector of residuals assumed to follow a multivariate normal distribution.

An F-test can be used to test the relevant null hypothesis.[12] The uniformly most powerful unbiased F-test is

$$(4.12) \qquad H_o: \ \alpha_i^{(1)} = 0, \ \ i = 2, \ldots, M$$

(i.e., the Sharpe-Lintner version of the CAPM is consistent with the data)

$$H_1: \ \alpha_i^{(1)} \neq 0, \ \ i = 2, \ldots, M$$

(i.e., the Sharpe-Lintner version of the CAPM is not consistent with the data)

The tests of Black's [1] version of the CAPM are more complicated because of the nonlinearity of the constraints and the zero beta rate is not observable. Gibbons [6] and Stambaugh [21] introduced applications of multivariate methods to nonlinear financial models. Jobson and Korki (11), Shanken (19) and MacKinley (14) provide further improvements for the multivariate tests. These authors introduce and compare tests that potentially have better finite sample properties than the asymptotic tests of Gibbons and Stambaugh.[13]

To test Black's [1] version of the CAPM we use the time series regression[14]

$$(4.13) \qquad \tilde{\mathbf{r}}_i - \hat{\lambda}_o^{(2)} \boldsymbol{\iota}_T = \alpha_i^{(2)} \boldsymbol{\iota}_T + \beta_i (\tilde{\mathbf{r}}_m - \hat{\lambda}_o \boldsymbol{\iota}_T) + \tilde{\boldsymbol{\eta}}_i^{(2)} , \qquad i = 2, \ldots, M .$$

$\hat{\lambda}_o^{(2)}$ is the estimate of the zero-beta rate obtained from the two-step procedure for CAPM outlined in section 4.1. The dimensions and notation of the other variables are the same as in (4.11). The null hypothesis to be tested with an F-test is

$$(4.14) \qquad H_o: \ \alpha_i^{(2)} = 0, \ \ i = 2, \ldots, M$$

(i.e., the Black version of CAPM is consistent with the data)

$$H_o: \ \alpha_i^{(2)} \neq 0, \ \ i = 2, \ldots, M$$

(i.e., the Black version of CAPM is not consistent with the data).

The APT can also be tested by generalizing this multivariate procedure. The time series regression for the excess return APT model in equation (4.4) can be rewritten in vector notation as

$$(4.15) \qquad \tilde{\mathbf{r}}_i - \hat{\lambda}_o^{(3)} \boldsymbol{\iota}_T = \alpha_i^{(3)} \boldsymbol{\iota}_T + \mathbf{B}(\hat{\boldsymbol{\Lambda}} - \hat{\lambda}_o^{(3)} \boldsymbol{\iota}_T) + \tilde{\boldsymbol{\eta}}_i^{(3)} , \quad i = 2, \ldots, M ,$$

where $\mathbf{r}_i$, $\mathbf{1}_T$ are defined as before. $\mathbf{B}$ is on $(M \times k)$ matrix of systematic risks (factor loadings) b_{ij} and $\mathbf{\Lambda}$ is a $(T \times k)$ matrix of risk premia. $\tilde{\boldsymbol{\eta}}_i^{(3)}$ is assumed to be multivariate normal. $\hat{\lambda}_o^{(3)}$ the estimate of the zero-beta rate obtained from the two-step procedure for APT outlined in section 4.1.

The application of the multivariate tests to the APT model is not as straightforward as it is to the CAPM. In the tests of the CAPM, $\mathbf{r}_m$, the excess return on the market portfolio proxy, is an observable variable, while $\mathbf{\Lambda}$, the matrix of risk premia in the APT model, is not and needs to be estimated.[15] In addition, the APT model does not a priori specify the number k of common factors. Ignoring the estimation error of $\mathbf{\Lambda}$ and $\lambda_o^{(3)}$, an F-test can be used to test the relevant null hypothesis

$$(4.16) \qquad H_o: \alpha_i^{(3)} = 0, \quad i = 2, \ldots, M$$

(i.e., the APT is consistent with the data)

$$H_1: \alpha_i^{(3)} \neq 0, \quad i = 2, \ldots, M$$

(i.e., the APT is not consistent with the data).

Equations (4.11), (4.13), and (4.15) are estimated as a "stacked" system of M regression equations by generalized least squares (GLS). We test in (4.12), (4.14), and (4.16) whether the intercept terms of the time series regressions are equal to zero for all i. An F-test is performed to compare the statistical fit of the unrestricted model with that of the restricted model in each of the three cases.

4.4 Multivariate Test Results

F-values for the multivariate intercept tests are tabulated in Table 6. The F-values are computed using excess returns in columns (2) and (3) and with real returns in the last two columns. Panels A, B, and C test the null hypotheses stated in (4.12), (4.14), and (4.16), respectively.[16]

The APT tests in panel C require further explanation. The seemingly unrelated regression method inverts the covariance matrix of the time series residuals $\tilde{\eta}_i^{(3)}$ in (4.15). This is not possible without reducing the number of Treasury portfolios by the number of factors because the risk premia estimates are a linear-weighted combination of the Treasury portfolios resulting in a singular residual covariance matrix. Therefore, we use estimates of the risk premia from Sample 1 as independent variables for Sample 2 and vice versa.[17]

As easily seen in Table 6 all F-values are significant at any conventional level, implying the rejection of all three null hypotheses.[18] This means the multivariate tests are not consistent with one to seven factor APT models or both versions of the CAPM as descriptive models of the U.S. Treasury securities market.[19]

4.5 One-Month-Ahead Predictions of Excess Returns

The two-stage tests suggest that there are at least two "priced" factors that generate Treasury security returns. The intercept tests using multivariate models, however, do not allow us to accept the APT model.

Unfortunately, complex estimation procedures like those entailed by the APT model are not straightforward and thus make it difficult to draw definitive conclusions regarding the APT. This is particularly true in our case because of the limitations of the data in the U.S. Treasury securities market. Recall from Section II that the derivation of the APT requires a relatively large number of assets so that the law of large numbers works. Given the small number of portfolios that we were able to construct our APT test results are at best inconclusive. (This is not the case, however, for the tests of the CAPM as an alternative model using various market indices.)

Nevertheless, an interesting yet basic question still remains to be answered. How well does a k-factor generating model predict excess returns? We will compare the "goodness of fit" of factor model forecasts with forecasts using standard market indices and a naive forecast. We have chosen the measure "goodness of fit" by the correlation coefficient between predicted and actual rates of return within the sample period.

All forecasting models use monthly returns over a five-year period beginning in January, 1960. After five years, when a forecast is made (for example, in January, 1965), the first month (January 1960) is dropped from the series and another (January, 1965) is added to get a forecast for the next month (February, 1965). The procedure is repeated until forecasts are obtained for January, 1965 through December, 1979, a total of 180 months.

The naive forecast is a simple average of the 60 monthly excess returns prior to each forecast month. More complicated autoregressive-integrated-moving average processes (ARIMA) for excess returns are not indicated from an examination of the autocorrelation function out to lag 24. (The autocorrelations for lags one to four months and also twelve months are presented in columns 8 to 12 in Panel A of Table 1.)

One-month-ahead forecasts based on the market indexes SP500, NYSE, EW, and (NYSE+EW)/2 are obtained by fitting the regression

$$(4.17) \qquad \tilde{r}_{it} = \alpha_i + \beta_i \tilde{r}_{mt} + \tilde{\varepsilon}_{it}, \quad i = 2, \ldots, M, \quad t = 1, \ldots, 60 .$$

In particular, $\hat{\alpha}_i$, $\hat{\beta}_i$ estimates are obtained over successive five-year periods. The $\hat{\beta}_i$ times the actual excess return of the market index next month plus $\hat{\alpha}_i$ is used as a one-month-ahead forecast of the excess return for portfolio i. The procedure is repeated for every five-year period prior to January, 1965, through December, 1979, yielding 180 forecasts.

Factor model forecasts are more complicated. We factor analyze excess returns each five-year period for one factor up to the maximum number k extractable with the maximum likelihood procedure. This gives an estimate of the loadings $\hat{b}_{ij}$. The $\hat{b}_{ij}$ are used to estimate the time series of factor scores $\hat{\delta}_{jt}$. This procedure can be represented as a multiple regression equation

$$(4.18) \qquad \tilde{r}_{it} = a_i^* + \sum_{j=1}^{k} \hat{b}_{ij}\hat{\delta}_{jt} + \tilde{\varepsilon}_{it}, \quad i = 2, \ldots, M, \quad t = 1, \ldots, 60 .$$

Thus, a regression between $\tilde{r}_{it}$ and $\hat{\delta}_{jt}$ gives an estimate of $\hat{a}_i^*$. Assuming the $\hat{b}_{ij}$ are the same next month, one-month-ahead factor scores can be calculated with next month's actual excess returns. The $\hat{a}_i^*$ plus the sum of the product of the estimated $\hat{b}_{ij}$'s and the one-month-ahead factor scores is used as a forecast of next month's excess return. The procedure is repeated for each five-year period prior to January, 1965, through December, 1979, yielding 180 forecasts.

A measure of the degree of association between two variables is the coefficient of correlation. For each of the forecasting models we calculated a correlation coefficient between the actual excess return and the predicted excess return. Table 7 presents the estimated correlation coefficients for Sample 1. Similar (unreported) results were obtained for Sample 2. Column 1 identifies the Treasury portfolio. Columns 2 to 6 contain correlation coefficients for the naive, SP500, NYSE, EW, and (NYSE + EW)/2 models, respectively. The last three columns are for one- to three-factor generating models. Averages and standard deviations of the correlation coefficient in each column are given in the last two rows.

Several observations are in order. First, the EW index model has an average correlation that is relatively larger than the other market index models. Second, the mean correlation of the factor models is higher than that of the EW index model with standard deviations about half as large.[20] A comparison of the correlations by portfolio reveals that factor models have higher correlations for short maturity portfolios whereas the EW index model has higher correlations for longer maturity portfolios. Third, 1-, 2-, and 3-factor models have about the same mean correlation so it is not obvious that forecasts using 2- or 3-factors explain returns better than forecasts based on only 1-factor.

V. SUMMARY AND CONCLUSIONS

This paper documents that mean U.S. Treasury bond portfolio returns segmented by maturity are not equal during the period 1960-1979. This finding implies that there is something for asset pricing models to explain. Therefore, we applied the APT and the CAPM to the sample data. We find that at least two factors are linearly related to mean bond portfolio returns. We did not, however, uncover a linear relation between mean bond returns and various market portfolio proxies. Furthermore, multivariate test results are not supportive of the APT or the Sharpe-Lintner and Black versions of the CAPM.

The tests here should be viewed simply as the first empirical attempt to properly measure interest-rate risk for bonds using factor generating models. Our results in terms of the existence of priced risk premia are more favorable to multifactor models that to single-factor models or the CAPM. Also, one-month ahead forecasts using factor generating models are somewhat better than corresponding naive predictions or predictions using a typical index portfolio. Multifactor models, however, do not seem to give a complete explanation of the risk-return relation in the U.S. Treasury security market.

FOOTNOTES

[1]See Huberman [9] for more precise definitions of arbitrage and conditions under which the APT relationship holds. Huberman shows that arbitrage portfolios need not be "sufficiently" diversified for Ross's original intuition to hold.

[2]See Ross [17] or Roll and Ross [16] or Ingersoll [10] for the derivations.

[3]Dhrymes, Friend, and Gultekin [4] have shown that the definition of the "universe" or "subset" of assets is as important to tests of the APT model as the misspecification of the market portfolio is to tests of the CAPM.

[4]There are only 6 missing observations out of 240 in Sample 2. There are none in sample 1.

[5]We also tested but could not accept the null hypotheses that mean total returns or mean real returns (total returns less the CPI inflation rate) are equal.

[6]Barlett's chi-square test is quite sensitive to departures of the data from normality. If the data are not normal, the actual level of significance may differ substantially from the specified one.

[7]The APT does not require that the "zero beta" rate equal the observed return on 30-day T-Bills. We further pursued the intercept tests using raw and real bond returns. When we use raw returns, the intercept terms are significantly smaller than the observed nominal 30-day T-Bill rate. Similarly, when we use real returns, the intercept term is significantly smaller than the observed real 30-day T-Bill rate.

[8]Fama and MacBeth [5] use a grouping procedure based on ranked betas to minimize measurement error problems in the cross-sectional regressions. It is not possible to exactly duplicate the Fama and MacBeth grouping procedure with Treasury bond data. Instead, we grouped Treasury securities into two samples of portfolios based on maturity (see Section 3.1). The bond portfolios will contain different numbers of bonds through time. It turns out that the rank correlation between maturity rankings and either variance or beta rankings of the bond portfolios in Samples 1 and 2 is one, regardless of the various market portfolio proxies used to estimate beta.

[9]Statistical properties of raw and excess returns are very different. This raises the question of whether the beta estimates using excess returns in place of raw (or real) returns are appropriate risk measures. Our results are not sensitive to the choice of returns to estimate beta. It might be of further interest to note that the rank correlation among betas estimated using raw, real and excess returns are exactly one for both Sample 1 and 2.

[10]The corresponding correlations using real returns are .9909, -.9314, .0252, .5305, .7410.

[11]We thank Michael J. Brennan and a referee for suggesting this methodology.

[12]To conserve space the computational form of the F-tests to follow are not detailed in the body of the paper. The interested reader can find these specifications in MacKinlay [14].

[13]As Shanken [19] points out, the likelihood ratio test of Gibbons rejects the null hypothesis too often when the number of assets approaches the number of observations. The Lagrange multiplier test of Stambaugh will never reject the null hypothesis as the number of assets increases.

[14]This is what MacKinlay [14] refers to as the Black model F-test (1). MacKinlay finds that the effect of "fat-tailed" distributions is to accept the null hypothesis too often with this F-test.

[15]One important advantage of the multivariate tests over the Fama-MacBeth type of two-stage test is the avoidance of the measurement problems encountered in the estimation of betas and the increase in the precision of parameter estimates for the risk premia. In the case of the APT the use of estimates of risk premia would invariably introduce measurement errors. We are therefore trading off one measurement problem for another. Since very little is known about the measurement errors introduced during the estimation of factor loadings, we should point out that it is not clear whether such tradeoff provides any advantage over the two stage tests in section 4.2.

[16]Recall from panel A in Table 1 that the Treasury portfolio excess return distributions are leptokurtic. The simulation results of MacKinlay [14] support the robustness of the tests of CAPM in panels A and B of Table 6 for leptokurtic distributions.

[17]We repeated the tests using the risk premia estimated within the same sample group. In this case we reduce the number of portfolios (dependent variables) as we increase the number of factors in order to ensure that the covariance matrix of residuals is nonsingular. We obtained similar results.

[18]The F-test of the Sharpe-Lintner model is the uniformly most powerful unbiased test. MacKinlay [14] finds, however, that the F-test of the Black model tends to <u>accept</u> the null hypothesis too often. This bias is in the opposite direction of our results which <u>reject</u> the null hypothesis.

[19]We correlated the estimated factor scores $\hat{\Lambda}$ with the various index returns using excess returns and real returns. At a 1% level of significance only factor one scores are correlated with the returns of the market portfolios (i.e. SP500, NYSE, EW, (NYSE+EW)/2).

[20]A one-tailed t-test can be used to test whether the mean correlations of the factor models are greater than .839915. The appropriate t-values of 2.16, 2.23, 2.45 for 1-, 2-, and 3-factor models are all significant at the 1% level. This suggests that the correlation of the factor models is typically higher than that of the EW index model.

REFERENCES

1. Fisher Black, "Capital Market Equilibrium with Restricted Borrowing," Journal of Business, 45 (July, 1972), 444-454.

2. John A. Boquist, George A. Racette, and Gary G. Schlarbaum, "Duration and Risk Assessment for Bonds and Common Stocks," Journal of Finance, 30 (December, 1975), 1360-1365.

3. Phoebus J. Dhrymes, Econometrics (New York: Harper & Row, Inc., 1974).

4. Phoebus J. Dhrymes, Irwin Friend and N. Bulent Gultekin, "A Critical Reexamination of the Empirical Evidence on Arbitrage Pricing Theory," Journal of Finance (June, 1984), 323-346.

5. Eugene F. Fama and James D. MacBeth, "Risk, Return, and Equilibrium: Empirical Tests," Journal of Political Economy, 38 (May, 1973), 607-636.

6. Michael Gibbons, "Multivariate Tests of Financial Models: A New Approach," Journal of Financial Economics, 10 (June, 1982), 3-29.

7. __________, "An Empirical Examination of the Return Generating Process of the APT," Working paper, Stanford University (1983).

8. N. Bulent Gultekin and Richard J. Rogalski, "Alternative Duration Specifications and the Measurement of Basis Risk: Empirical Tests," Journal of Business (April, 1984).

9. Gur Huberman, "Arbitrage Pricing Theory: A Simple Approach," Journal of Economic Theory, 28 (1982), 183-191.

10. Jonathan E. Ingersoll, Jr., "Some Results in the Theory of Arbitrage Pricing," Journal of Finance, 39 (September, 1984), 1021-1039.

11. J. O. Jobson and Bob Korkie, "Potential Performance and Tests of Portfolio Efficiency," Journal of Financial Economics, 10 (December, 1982), 433-466.

12. D. N. Lawley and A. E. Maxwell, Factor Analysis as a Statistical Method (London: Butterworth, 1963).

13. John Lintner, "The Valuation of Risk Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets," The Review of Economics and Statistics, 67 (February, 1965), 13-37.

14. A. Craig MacKinlay, "An Analysis of Multivariate Financial Tests," unpublished manuscript, University of Chicago (January, 1984).

15. Richard Roll, "A Critique of the Asset Pricing Theory's Tests," Journal of Financial Economics, 4 (May, 1977), 129-176.

16. Richard Roll and Stephen A. Ross, "An Empirical Investigation of the Arbitrage Pricing Theory," Journal of Finance, 35 (December, 1980), 1073-1104.

17. Stephen A. Ross, "The Arbitrage Theory of Capital Asset Pricing," Journal of Economic Theory, 13 (December, 1976), 341-360.

18. __________, "Return, Risk, and Arbitrage," in Irwin Friend and James L. Bicksler, eds., Risk and Return in Finance, Vol. 1 (Cambridge, Mass.: Ballinger, 1977), 189-218.

19. Jay Shanken, "Multivariate Tests of the Zero-beta CAPM," Working paper, University of California at Berkeley, 1984.

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

21. Robert Stambaugh, "On the Exclusion of Assets from Tests of the Two-Parameter Model: A Sensitivity Analysis," Journal of Financial Economics, 10 (November, 1982), 237-268.

22. Henri Theil, Principles of Economics (New York: John Wiley & Sons, Inc., 1971).

Table 1

SUMMARY STATISTICS FOR MONTHLY EXCESS RETURNS ON TREASURY PORTFOLIOS AND MARKET PORTFOLIOS: 1/1960-12/1979

Portfolio (1)	Maturity Range (days) (2)	Mean (3)	Variance (4)	Skew-ness (5)	Kur-tosis (6)	Student-ized Range (7)	$\hat{\rho}_1$ (8)	$\hat{\rho}_2$ (9)	$\hat{\rho}_3$ (10)	$\hat{\rho}_4$ (11)	$\hat{\rho}_{12}$ (12)
Panel A: Treasury Portfolios											
2	60-89	3.31	.0030	1.58*	9.31*	8.77*	.1870*	.1093	.0798	.0727	.1207
3	90-119	3.18	.0083	1.55*	13.41*	10.75*	.0596	.1101	.0216	.0967	.0995
4	120-149	4.90	.0158	.94*	10.41*	10.27*	.0695	.1582	.0105	.0971	.0623
5	150-179	6.06	.0248	.40	9.48*	10.50*	.0260	.0253	.0478	.0747	.0994
6	180-269	4.12	.0489	-.20	6.35*	7.72*	.0499	.0208	.1075	.0668	.1217
7	270-359	4.08	.1037	-.08	6.56*	8.38*	.0314	.0181	.0245	.0662	.1318
8	360-719	3.13	.2362	-.12	6.03*	8.22*	.0404	-.0053	.0673	.0535	.0839
9	720-1079	2.17	.5522	-.13	6.08*	8.48*	-.0060	-.0201	.0373	.0599	.0672
10	1080-1439	1.20	.8501	-.06	6.79*	8.86*	.0018	-.0175	.0410	.0817	.0261
11	1440-1799	-.69	1.1646	-.04	6.85*	8.97*	-.0387	.0074	.0188	.0679	.0233
12	1800-2159	.01	1.6950	-.45*	6.90*	8.45*	-.0431	.0581	.0340	-.0033	.0746
13	2160-	-4.86	3.0159	-.17	5.67*	7.76*	-.0472	.1078	-.1034	-.0065	.0113
Panel B: Market Portfolios											
SP500		13.35	16.7064	-.18	3.93*	6.80	.0504	.0833	.0746	-.0363	.0881
NYSE		58.24	30.8927	.39	6.10*	8.45*	.0212	.0391	.0253	-.0438	.1671
EW		2.21	.2673	-.18	6.66*	8.86*	-.0026	.0373	.0112	.0384	.0422
(NYSE+EW)/2		30.18	8.1620	.33	5.98*	8.40*	.0262	.0558	.0190	-.0442	.1663

Means and variances are multiplied by 10,000. SP500, NYSE, EW, (NYSE+EW)/2 denote the Standard & Poor 500 Index, New York Stock Exchange equally weighted index, an equally weighted index of all Treasury securities on the CRSP Government Bond Tape, and a simple average of NYSE and EW.

Skewness is measured by the third moment from the mean divided by the second moment to the 3/2 power. Kurtosis is computed by dividing the fourth moment about the mean by the square of the second moment about the mean. A normal distribution has zero skewness and kurtosis of three. The studentized range is the range of observations in the sample divided by the square root of the sample variance. $\hat{\rho}_1$, $\hat{\rho}_2$, $\hat{\rho}_3$, $\hat{\rho}_4$, $\hat{\rho}_{12}$ are sample autocorrelations for lags one to four and twelve, respectively. Asterisks represent values that are significant at the 1% level.

Table 2

ESTIMATED VARIANCE-COVARIANCE MATRIX AND CORRELATION MATRIX FOR TREASURY PORTFOLIO EXCESS RETURNS

	Treasury Portfolio											
Treasury Portfolio	2	3	4	5	6	7	8	9	10	11	12	13
2	.0339	.0387	.0315	.0347	.0365	.0461	.0613	.0828	.0942	.1023	.1304	.1443
3	.8567	.0380	.0387	.0417	.0427	.0557	.0708	.0942	.1079	.1169	.1516	.1696
4	.8549	.9424	.0443	.0470	.0503	.0665	.0849	.1128	.1292	.1401	.1797	.1979
5	.8131	.9234	.9644	.0536	.0569	.0758	.0982	.1324	.1532	.1681	.2158	.2445
6	.7175	.7931	.8653	.8888	.0764	.0910	.1225	.1672	.1954	.2145	.2640	.2927
7	.7024	.8020	.8861	.9188	.9236	.1271	.1690	.2320	.2729	.3031	.3799	.4323
8	.6622	.7220	.8019	.8434	.8816	.9424	.2529	.3564	.4182	.4726	.5707	.6805
9	.6016	.6463	.7165	.7651	.8092	.8704	.9479	.5591	.6650	.7547	.8730	1.0595
10	.5546	.6001	.6654	.7170	.7661	.8295	.9012	.9638	.8514	.9650	1.1014	1.3421
11	.5158	.5568	.6176	.6738	.7204	.7892	.8722	.9368	.9707	1.1608	1.2789	1.5958
12	.5453	.5986	.6572	.7175	.7353	.8202	.8736	.8987	.9188	.9137	1.6878	1.9808
13	.4518	.5016	.5420	.6089	.6106	.6992	.7802	.8170	.8386	.8540	.8791	3.0082

Variances are on the diagonal. Covariances are above the diagonal and correlation coefficients are below the diagonal. Variances and covariances are multipled by 10,000.

Table 3

CHI-SQUARE TEST OF THE HYPOTHESIS THAT k-FACTORS GENERATE TREASURY PORTFOLIO EXCESS RETURNS AND ESTIMATED FACTOR LOADINGS FOR THE 5-FACTOR MODEL

	Estimated Factor Loadings $\mathbf{B} = [\hat{b}_{ij}]$ b/					Chi-square Tests c/		
Treasury Portfolio a/	$\hat{b}_{i1}$	$\hat{b}_{i2}$	$\hat{b}_{i3}$	$\hat{b}_{i4}$	$\hat{b}_{i5}$	Number of Factors (k)	χ^2_v	Probability d/
Panel A: Sample 1								
2	3.688	2.034	2.132	1.451	-.205	1	1956.95	.0001
3	6.654	3.908	4.076	1.368	.139	2	440.83	.0001
4	10.121	5.420	4.757	.527	-.215	3	182.82	.0001
5	13.358	5.930	5.141	-.214	.566	4	96.89	.0001
6	19.645	5.880	2.991	-2.701	-1.581	5	10.65	.8303
7	30.356	7.552	3.189	-5.943	.571			
8	48.239	4.573	-4.505	.532	-.101			
9	72.254	-11.024	-.216	2.700	-2.195			
10	87.502	-27.944	8.404	-.431	-1.606			
11	99.139	-35.261	6.105	1.984	7.002			
12	118.239	-25.804	6.756	-2.845	31.179			
13	141.338	-43.894	.855	12.042	68.922			
Panel B: Sample 2								
2	3.833	-2.336	1.407	-.701	.519	1	2262.14	.0001
3	7.074	-4.534	2.071	-.365	.437	2	718.47	.0001
4	10.724	-5.772	1.502	-.229	-.084	3	350.25	.0001
5	14.024	-6.002	1.136	.554	-.005	4	207.73	.0001
6	19.875	-4.581	-2.478	-.332	-1.924	5	88.66	.0001
7	30.741	-4.898	-5.101	2.413	1.787	6	55.27	.0438
8	40.967	-.110	-6.789	-1.372	2.163	7	29.09	.4605
9	55.972	7.085	-4.848	-4.587	5.513			
10	66.208	14.651	-1.912	-7.599	3.750			
11	71.975	19.579	2.214	-12.846	-1.650			
12	81.221	25.286	7.062	-3.784	-8.761			
13	86.088	34.549	12.449	1.985	-9.461			
14	91.558	40.742	14.661	2.818	-8.302			
15	96.963	40.032	13.117	7.397	-.470			
16	114.634	33.086	9.339	27.421	11.822			
17	132.449	53.758	18.700	41.148	34.384			

a/ See Table 1 for maturity ranges for the portfolios. b_{ij}'s are multiplied by 10^4.

b/ The factor loadings are estimated using the maximum likelihood procedure by solving the equations (a) $\text{diag}(\mathbf{S}) = \text{diag}(\hat{\mathbf{B}}\hat{\mathbf{B}}' + \tilde{\mathbf{D}})$ and (b) $[\hat{\mathbf{D}}^{-1/2}(\mathbf{S}-\hat{\mathbf{D}})\hat{\mathbf{D}}^{-1/2}]\hat{\mathbf{D}}^{-1/2}\hat{\mathbf{B}} = \tilde{\mathbf{D}}^{-1/2}\hat{\mathbf{B}}(\hat{\mathbf{B}}'\hat{\mathbf{D}}^{-1}\hat{\mathbf{B}})$. $\mathbf{S}$ is the sample covariance matrix of excess returns. $\hat{\mathbf{B}}$ is the estimate of factor loadings and $\hat{\mathbf{D}}$ is the estimated diagonal matrix of residual error variance of excess returns.

c/ The test statistic is $\chi^2_v = [T - 1 - (1/6)(2m + 5) - (2/3)k]\ln[\,|\tilde{\mathbf{B}}\tilde{\mathbf{B}}' + \tilde{\mathbf{D}}| \,/\, |\mathbf{S}|\,]$ with degrees of freedom $v = \frac{1}{2}[(m-k)^2 - m - k]$, where T is the number of observations, m is the number of securities, and k is a prespecified number of factors.

d/ The p-value associated with the statistic, i.e., the probability that the test statistic (under the null hypothesis) will assume a value at least as large as the statistic obtained in this particular test.

Table 4

TWO STAGE TESTS OF THE APT
1/1960 - 12/1979

Number of Factors a/	$\bar{\lambda}_o$ b/	$\bar{\lambda}_1$	$\bar{\lambda}_2$	$\bar{\lambda}_3$	$\bar{\lambda}_4$	$\bar{\lambda}_5$	$\bar{\lambda}_6$	$\bar{\lambda}_7$	$t(\bar{\lambda}_o)$ d/	χ^2_k d/
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Panel A: Sample 1										
1	.3996	-.0154							9.55*	.054
2	.2009	.0988	.3114						5.90*	19.034*
3	.1081	.1273	.3837	.2619					3.06*	32.066*
4	.0981	.1238	.3945	.2789	-.0621				2.82*	34.545*
5	.1129	.0511	.2650	.3706	.2013	-.0570			2.89*	35.801*
Panel B: Sample 2										
1	.4043	-.0264							9.33*	.153
2	.2148	.0827	-.3068						6.08*	17.410*
3	.1358	.1035	-.2744	.2251					3.98*	30.121*
4	.1188	.1044	-.3939	.2375	-.0929				3.48*	32.935*
5	.1008	.0987	-.3790	.2898	-.0757	-.0506			3.35*	36.043*
6	.1112	.0713	-.3678	.2684	-.0758	.1112	.1094		3.16*	34.825*
7	.1267	-.0178	.2881	-.3066	.2018	-.0740	-.0247	.0142	3.30*	34.865*

a/ Cross sectional regressions are estimated for one- through five- or seven-factor models.

b/ $\bar{\lambda}_o$ through $\bar{\lambda}_7$ are the arithmetic means of the monthly cross sectional regression estimates using the GLS model in 4.7. λ_o is the intercept term and these are multiplied by 1000. λ_1 through λ_7 are the regression coefficients for the factor loadings.

c/ $t(\bar{\lambda}_o)$ is the t-statistic for the intercept term under a null hypothesis of zero. Asterisks indicate significant t-values using a 5% level.

d/ χ^2_k is the test statistic in 4.10 to test the null hypothesis that none of the risk premium are priced. The test statistic is distributed as chi-square with k degrees of freedom (k is the number of factors). Asterisks indicate chi-square values significant at a 5% level.

Table 5

TWO STAGE TESTS OF THE CAPM
1/1960 - 12/1979

Market Portfolio Proxies (1)	$\bar{\lambda}_o$ a/ (2)	Statistic $\bar{\lambda}_1$ a/ (3)	$\overline{\lambda_o - r_f}$ (4)	$t(\bar{\lambda}_o)$ b/ (5)	$t(\bar{\lambda}_1)$ b/ (6)	$\overline{t(\lambda_o - r_f)}$ b/ (7)
Panel A: Excess Returns						
Sample 1						
SP500	.00050971	-.00749493		5.20	-.78	
NYSE	.00051126	-.00026566		5.08	-.77	
EW	.00052953	-.01179253		4.97	-.78	
(NYSE+EW)/2	.00052568	-.00453573		5.00	-.77	
Sample 2						
SP500	.00051752	-.00763635		5.13	-.85	
NYSE	.00052821	-.00029276		4.99	-.86	
EW	.00054089	-.01204096		4.99	-.84	
(NYSE+EW)/2	.00053927	-.00478130		5.00	-.85	
Panel B: Real Returns						
Sample 1						
SP500	.00080879	-.00767680	.00068525	2.20	-.78	2.61
NYSE	.00083519	-.01196246	.00071165	2.13	-.77	2.31
EW	.00080059	-.00043090	.00067705	2.18	-.77	2.58
(NYSE+EW)/2	.00082724	-.00473069	.00070370	2.14	-.77	2.43
Sample 2						
SP500	.00080879	-.00076713	.00068525	2.20	-.78	2.61
NYSE	.00083551	-.01169754	.00071154	2.19	-.79	2.50
EW	.00080398	-.00043254	.00068044	2.20	-.78	2.62
(NYSE+EW)/2	.00082880	-.00046634	.00070556	2.19	-.79	2.51

a/ $\bar{\lambda}_o$ and $\bar{\lambda}_1$ are the arithmetic means of the monthly cross sectional regression estimates using the OLS model $\hat{\lambda}^*_t = (\hat{B}^{*\prime}\hat{B}^*)^{-1}\hat{B}^{*\prime}\tilde{r}_t$ where $\tilde{r}_t$ denotes excess returns in panel A and real returns in panel B. B^* is a (M×2) matrix with the first column containing ones and the second column containing beta estimates for the portfolios. $\hat{\lambda}_{ot}$ is the intercept term; $\hat{\lambda}_{1t}$ is the estimate of the risk premium on the market portfolio. $\overline{\lambda_o - r_f} = \frac{1}{T}\sum_{t=1}^{T}(\hat{\lambda}_{ot} - r_{ft})$.

b/ t-ratios for the regression coefficients are calculated using T = 240 observations as follows: $\sqrt{T}\,(\bar{\lambda}_i/s_{ii})$, where $\bar{\lambda}_i = \frac{1}{T}\sum_{t=1}^{T}\hat{\lambda}_{it}$ and $s_{ii} = [\frac{1}{T}\sum_{t=1}^{T}(\hat{\lambda}_{it} - \bar{\lambda}_i)^2]^{1/2}$, for i = 0 and 1.

Table 6

MULTIVARIATE TESTS OF THE CAPM AND APT:
F-VALUES FOR INTERCEPT TESTS
1/1960-12/1979

Model Tested (1)	Using Excess Returns Sample 1 (2)	Using Excess Returns Sample 2 (3)	Using Real Returns Sample 1 (4)	Using Real Returns Sample 2 (5)
Panel A: Sharpe-Lintner CAPM H_o: $\alpha_i^{(1)} = 0$				
SP500	13.353	8.138	13.353	8.138
NYSE	13.579	8.290	13.579	8.290
EW	16.545	12.017	16.545	12.017
(NYSE+EW)/2	13.006	8.317	13.006	8.317
Panel B: Black CAPM H_o: $\alpha_i^{(2)} = 0$				
SP500	18.900	15.732	10.253	8.128
NYSE	20.237	17.155	10.613	8.426
EW	22.449	17.454	11.041	8.387
(NYSE+EW)/2	20.027	17.106	10.654	8.465
Panel C: APT H_o: $\alpha_i^{(3)} = 0$				
1	11.082	8.135	11.050	7.934
2	8.703	6.680	7.520	6.883
3	8.087	5.698	8.351	6.566
4	8.026	5.546	8.059	6.431
5	7.373	5.423	9.102	5.538
6	7.394		6.415	
7	7.841		6.312	

Excess returns are total returns less the monthly risk-free rate. Real returns are total returns less the monthly CPI inflation rate. All F-values are significant at any conventional level and hence none of the null hypotheses can be accepted. Note that results are the same for excess returns and real returns in tests of Sharpe-Lintner CAPM because subtracting similar rates on both sides of 4.11 does not change the computed F-value.

Table 7

CORRELATION COEFFICIENTS BETWEEN MONTHLY REALIZED RETURNS AND
SEVERAL EXCESS RETURN GENERATING MODELS

Treasury Portfolio (1)	Naive (2)	Market Index Models				Factor Models		
		SP500 (3)	NYSE (4)	EW (5)	(NYSE + EW)/2 (6)	1 Factor (7)	2 Factors (8)	3 Factors (9)
2	.010840	.003411	.090426	.582474	.159051	.737883	.736219	.739868
3	.014368	.100577	.053905	.638496	.136274	.820258	.819596	.821403
4	.049902	.011387	.170946	.717503	.254380	.889292	.889256	.888576
5	.017502	.048106	.163835	.770865	.248211	.918135	.918811	.924607
6	.074253	.243227	.278237	.806359	.356133	.905381	.905988	.926317
7	.076680	.236192	.273660	.876609	.356055	.957012	.955472	.956531
8	.119520	.273451	.276723	.934224	.364802	.957792	.966245	.966026
9	.083702	.281520	.252280	.958024	.338484	.930803	.932672	.933949
10	.085547	.258330	.202169	.964504	.289294	.909117	.912739	.914534
11	.081840	.257291	.187032	.956000	.273631	.884578	.897024	.895315
12	.112951	.156434	.119829	.960171	.215991	.885926	.879814	.874289
13	.082811	.204991	.165120	.913746	.250730	.776949	.781137	.795338
Average	.067493	.172910	.186180	.839915	.270253	.881094	.882914	.886396
Standard Deviation	.035087	.100912	.071411	.129366	.072506	.065843	.066790	.065564