THE JOURNAL OF FINANCE • VOL. XLII, NO. 2 • JUNE 1987

Tests of Asset Pricing with Time-Varying Expected Risk Premiums and Market Betas

WAYNE E. FERSON, SHMUEL KANDEL, and ROBERT F. STAMBAUGH*

ABSTRACT

Tests of asset-pricing models are developed that allow expected risk premiums and market betas to vary over time. These tests exploit the relation between expected excess returns and current market values. Using weekly data for 1963 through 1982 on ten common stock portfolios formed according to equity capitalization, a single-risk-premium model is not rejected if the expected premium is time varying and is not constrained to correspond to a market factor. Conditional mean-variance efficiency of a value-weighted stock index is rejected, and the rejection is insensitive to how much variability of expected risk premiums is assumed.

FINANCIAL ECONOMISTS HAVE STUDIED valuation models in which securities' expected returns are related to the covariances of unanticipated returns with the return on the market portfolio or with other risk factors. Empirical tests of such models typically have employed unconditional sample moments of returns. Recent studies provide mounting evidence that expected returns change over time with information.[1] Such evidence presents an opportunity to develop new approaches for evaluating and implementing asset-pricing models that exploit this time variation. This paper presents new tests of financial valuation models in which expected returns are allowed to vary over time. Time series of relative market values are included as data, thereby allowing conditional market betas to vary as well.

We conduct tests using weekly returns in excess of the Treasury bill rate. The tests' assets are ten value-weighted portfolios of common stocks ranked by firm-equity capitalization over the period 1963 through 1982. The tests reject conditional mean-variance efficiency of a value-weighted stock index, but they do not reject a single-risk-premium model if the premium is not restricted to be a "market" factor and is allowed to vary over time. We investigate the properties of the tests using simulations. These experiments indicate that the rejection of conditional mean-variance efficiency is insensitive to variability in expected risk

* All authors from Graduate School of Business, University of Chicago. We are grateful to John Abowd, Nai-Fu Chen, George Constantinides, Kenneth Dunn, Eugene Fama, Michael Gibbons, Gur Huberman, Ravi Jagannathan, and participants in seminars at the American Finance Association meetings, Ohio State University, Southern Methodist University, the Wharton School, the Western Finance Association meetings, and Yale University for helpful discussions and comments. We thank Richard Rogalski for providing the Treasury bill data for 1963 through 1981 and Vikram Nanda for collecting additional data to extend the series through 1982.

[1] A few of the studies that provide evidence of predictable variation in returns on common stocks, the class of assets examined here, include Fama and Schwert [14], Hall [19], Rozeff [34], Campbell [5], Keim and Stambaugh [24], and Fama and French [13].

premiums and that the test for exact single-premium asset pricing has power against a multipremium model.

The paper is organized as follows. The next section describes the multibeta relation for assets' conditional expected excess returns that is the focus of the paper. The empirical formulation of this pricing equation is discussed in the contexts of mean-variance efficiency, multibeta equilibrium models, and exact arbitrage pricing. Section II presents the statistical methodology. Section III describes the data. Section IV presents the empirical results and discusses the simulations used to assess significance levels and power. Section V concludes the paper.

I. Formulations of Multibeta Pricing with Time-Varying Expectations

Let $E_t = \mathrm{E}(\tilde{r}_{t+1} \mid \phi_t)$ be the vector of expected excess returns (rates of return minus a risk-free rate) on N assets from time t to time $t + 1$. The expectation is conditioned on the information ϕ_t available in the market at time t. Let $V = \mathrm{cov}(\tilde{r}_{t+1} \mid \phi_t)$ be the conditional covariance matrix of returns, which is assumed to be constant.

The focus of the paper is the following multibeta relation for assets' conditional expected excess returns:

$$E_t = a_t V w_t + \Omega h_t, \tag{1}$$

where

- w_t is the N-vector of aggregate demands (market portfolio's weights),
- h_t is a K-vector of time-varying risk premiums,
- a_t is a scalar related to aggregate relative risk aversion, and
- Ω is the matrix of the conditional covariances of asset returns with the state variables, which is assumed constant.

Asset-pricing equations similar to (1) have been derived in many papers. A well-known example is the single-period Capital Asset Pricing Model (CAPM) of Sharpe [35] and Lintner [25], which is equivalent to the mean-variance efficiency of the market portfolio.[2] The CAPM is essentially a special case of equation (1) containing only the market-premium term $a_t V w_t$. (This term is proportional to the vector of the market "betas" as $V w_t$ is the vector of the covariances of the N asset returns with the return on the market portfolio.)

Although mean-variance efficiency is typically studied in the context of single-period models, conditional mean-variance efficiency of the market portfolio is derived by Merton [28] as a special case of his intertemporal asset-pricing model when investors are assumed to have logarithmic utility functions.

Multiple-premium models ($K > 0$) arise in a variety of contexts. Pricing equations similar to (1) arise in intertemporal models such as those of Merton [28], Long [26], and Cox, Ingersoll, and Ross [9] if optimal consumption responds to fluctuations in K state variables for a given level of wealth. Equation (1) with

[2] Fama [12], Roll [30], and Ross [33] discuss this equivalence.

$K > 0$ can also describe expected returns in a single-period model if not all assets are traded (e.g., Mayers [27]).

An interesting special case of equation (1) is the case in which risk measures need not correspond to covariances with the return on a market portfolio. Expected returns are consistent with "exact arbitrage pricing" with K factors if

$$\mathrm{E}(\tilde{r}_{t+1} \mid \phi_t) = \Omega h_t,$$

where h_t is a vector of K ($< N$) expected risk premiums and Ω is a matrix of factor loadings. Such models for expected excess returns are obtained as approximations in large economies by Ross [32], Chamberlain and Rothschild [6], Stambaugh [36], Ingersoll [22], and Connor [8]. Grinblatt and Titman [18] and Dybvig [10] study the magnitude of deviations from exact pricing in equilibrium models with a finite number of assets.

Assuming rational expectations, realized excess returns minus the forecast error $\tilde{\varepsilon}_{t+1}$ may be substituted for E_t in (1) to obtain an empirical formulation of equation (1):

$$\tilde{r}_{t+1} = a_t V w_t + \Omega h_t + \tilde{\varepsilon}_{t+1}; \quad \mathrm{E}(\tilde{\varepsilon}_{t+1} \mid \phi_t) = 0, \tag{2}$$

$$\mathrm{E}(\tilde{\varepsilon}_{t+1} \tilde{\varepsilon}'_{t+1} \mid \phi_t) = V. \tag{3}$$

In the tests that follow, we maintain the assumptions that V and Ω are constant for the test assets over the relevant sample period but that the other parameters in (2) may change over time. A sufficient condition for constant V and Ω is that the joint distribution of the returns, the hedged state variables, and the relevant information ϕ_t is a stationary normal distribution. In that case, conditional expectations of returns depend on the realization of the conditioning information, but the conditional covariances are constant. The assumption that V and Ω in equation (2) are constant allows expected excess returns E_t to change over time with market weights w_t and with $K + 1$ parameters (a_t and h_t). Market "betas" of the test assets (proportional to Vw_t) can also change as the composition of the market (w_t) changes. This approach complements the tests of Hansen and Hodrick [20] and Gibbons and Ferson [17], who assume that expected returns change over time but that conditional betas are constant.

Equations (2) and (3) require that the matrix of partial regression coefficients of returns on $a_t w_t$ equals the covariance matrix of unanticipated returns. Frankel and Dickens [15] observe that conditional mean-variance efficiency of a market portfolio (the case where $K = 0$ in equation (1)) implies such a covariance restriction. They test this restriction assuming a constant conditional covariance matrix and a constant ratio of expected excess return to variance of the market-portfolio proxy (a_t in (1) being constant). Rayner [29] studies Frankel and Dickens' formulation as well as a version of the model with a constant Sharpe measure (ratio of expected excess return to standard deviation) of the market portfolio. Bollerslev, Engle, and Wooldridge [2] extend Frankel and Dickens' formulation to allow time variation in the covariance matrix. Like Frankel and Dickens, we assume that the conditional covariance matrix is constant over time. Unlike any of these studies, we allow the ratio of expected excess return to conditional variance of the value-weighted stock index, as well as that portfolio's

Sharpe measure, to vary over time. Equations (2) and (3) are also used to extend the covariance restriction to the context of a multibeta asset-pricing model, which is not examined by the above authors.

We test three asset-pricing hypotheses as special cases of equations (2) and (3). In each test, the market portfolio is approximated by a value-weighted stock index, and the test assets are ten size-sorted portfolios of stocks from the New York and American Stock Exchanges. The first case is conditional mean-variance efficiency of the value-weighted stock index. This is tested as the hypothesis $K = 0$ in equation (2), subject to the covariance restriction in (3).

The second case is an exact K-risk-premium model in which market weights do not enter the pricing equation. This case corresponds to $a_t = 0$ for a given number of premiums, K. Note that, if $a_t = 0$, then the covariance restriction (equation (3)) is not imposed. Unlike many tests of exact arbitrage pricing, a test of $a_t = 0$ does not assume a strict factor structure.

The third case hypothesizes that there exist K risk premiums in equation (2) in addition to a "market" premium. If the model is accepted for a given K, then K is interpreted as an upper bound for the number of "nonmarket" risk premiums. We interpret the results as indicating an upper bound because of the effects of missing assets on such tests. If a subset of assets is sampled and a_t is not zero in equation (1), then the partition of (1) containing the sample assets can include spurious "premiums" if the conditional correlation between the test assets and the excluded assets is nonzero.

II. Statistical Methodology

Following Gibbons [16] and others, this study applies maximum-likelihood methods to a pooled time series and cross-section of returns. We extend such an approach from unconditional moments to a framework with changing expected returns. We develop estimates and tests using a normal likelihood function for the unexpected component of returns. Bootstrap simulations presented below indicate that the distributions of our test statistics are not sensitive to nonnormality of our data. If the unanticipated return $\tilde{\varepsilon}_{t+1}$ is normal $(0, V)$ and independent over time, then equations (2) and (3) imply that the log-likelihood function l for observations $t = 1, \cdots, T$ can be written as

$$l = -\frac{nT}{2}\ln(2\pi) - \frac{T}{2}\ln|V| - \frac{1}{2}\sum_t (r_{t+1} - a_t V w_t - \Omega h_t)' V^{-1}(r_{t+1} - a_t V w_t - \Omega h_t). \quad (4)$$

The alternative hypothesis is the "unrestricted" model in which K is set equal to the number of test assets. Note that (4) embeds the restriction (3) on the covariance matrix V, while, under the alternative, the covariance restriction (3) is not binding.

Although the likelihood function in (4) presents a natural framework for testing asset-pricing models, several issues must be addressed in developing such tests. First, the parameters a_t and h_t that allow expected returns to vary over

time are incidental. That is, the number of parameters increases with the sample size. As a result, standard asymptotic properties of the maximum-likelihood approach cannot be invoked. We examine below the properties of the likelihood-ratio test statistics using simulations, and we find that the distributions of the statistics are well approximated by normal distributions.

Because Ω and h_t enter the likelihood function only as the product Ωh_t, the individual parameters in Ω and h_t are not identified uniquely. This indeterminacy is analogous to that encountered in factor analysis. We adopt identifying restrictions on Ω that are similar to those employed in factor analysis: we set $\Omega' V^{-1} \Omega = I$ and $\Omega' \Omega = D$, where D is a diagonal matrix.

The structure provided by equation (1) for expected returns is not sufficient to bound the likelihood function. We adopt a simple assumption that bounds the likelihood function but still allows expected excess returns to follow a variety of sample paths through time. We assume that expected excess returns move smoothly enough through time to be approximated with a step function. For example, with weekly data, expected excess returns are allowed to change every two weeks. The intuition of how the test makes use of this assumption is seen in the following illustration.

For simplicity, let a_t and, thus, the first term on the right-hand side of equation (1) equal zero. Consider the vector of expected excess returns E_t in a two-asset economy. Under the general alternative hypothesis ($K = 2$), E_t can vary over time in a two-dimensional space (a plane). The null hypothesis ($K = 1$) restricts E_t to move along a line. We first examine the test when the null hypothesis is false. Figure 1a depicts a hypothetical situation where the alternative hypothesis ($K = 2$) is true: the solid curve represents the time path of the "true" expectations, and the numbered points "×" represent the sequence of weekly excess returns observed over ten weeks. The two-week sample mean excess returns, represented as the five "+'s" in Figure 1b, provide an estimate of the time path of the true expectations. The likelihood function under the alternative contains the weighted sum of squared residuals about these sample means. Under the null hypothesis ($K = 1$), the likelihood function constrains the two-week means to lie along a straight line, and these constrained means are represented as "◇'s" in Figure 1b. Comparing the "fit" of the data around the two-week means under the null (◇'s) versus the alternative (+'s), we obtain a "large" value of the test statistic. The likelihood-ratio test statistic (the negative of twice the difference between the log-likelihood functions under the null and alternative hypotheses) is 16.68 using the hypothetical data plotted in Figures 1a and 1b.

We next examine the test when the null hypothesis ($K = 1$) is true and expected returns move over time along a line, as shown in Figure 2a. Note that the time sequence of the actual returns (×'s) is important. These examples are constructed so that, except for a different numbering of the time sequence, the scatter of actual returns in Figure 2a is identical to that in Figure 1a. Thus, the unconditional sample means and covariance matrices are the same in both cases. The two-week sample means are again shown as the +'s in Figure 2b. In this case, the fit of the actual returns around the two-week means does not differ from the fit around the constrained two-week means (◇'s) as much as in the previous case. The likelihood-ratio test statistic equals 1.85 in this example. By comparing the

Figure 1a: Actual and Expected Returns

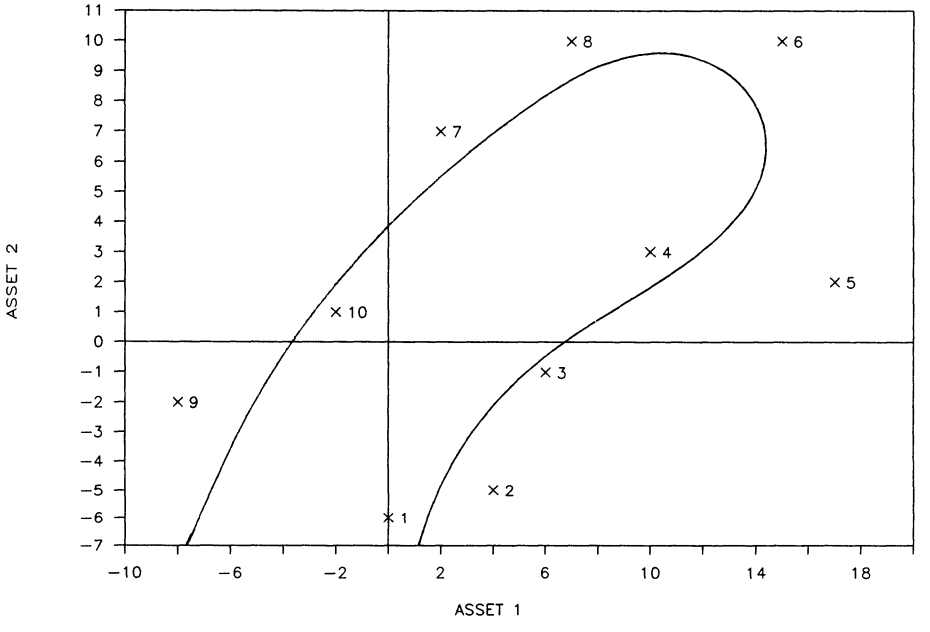

Figure 1b: Test of K = 1 (LRTS = 16.68)

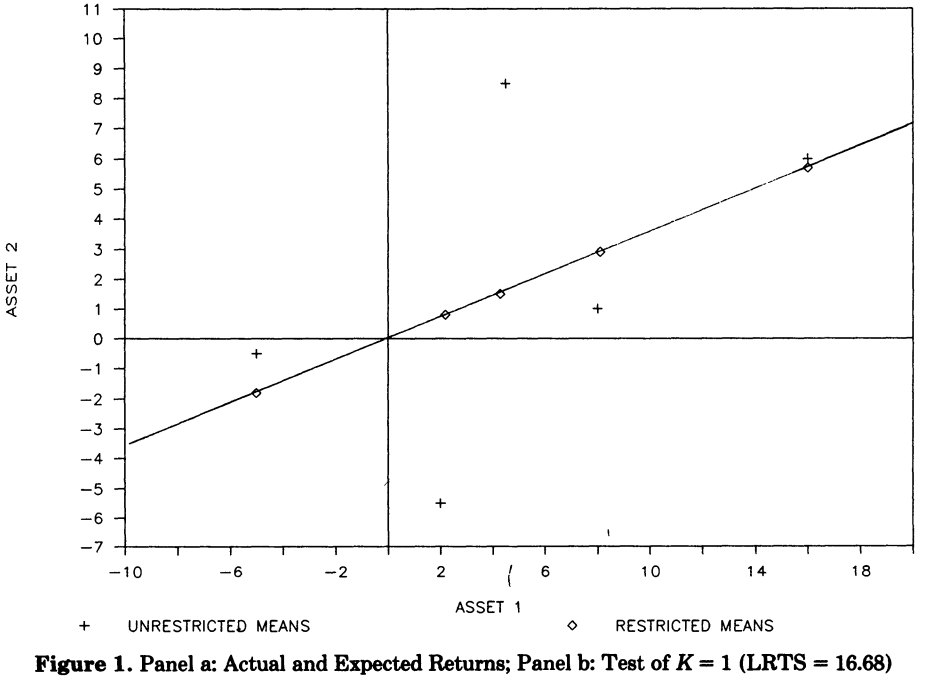

Figure 1. Panel a: Actual and Expected Returns; Panel b: Test of $K = 1$ (LRTS = 16.68)

Figure 2a: Actual and Expected Returns

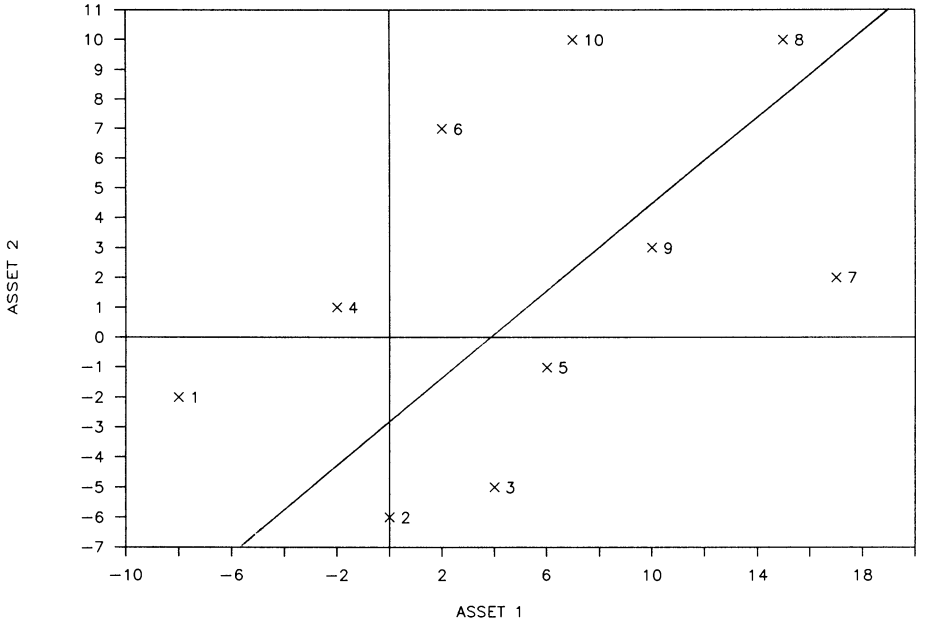

Figure 2b: Test of K = 1 (LRTS = 1.85)

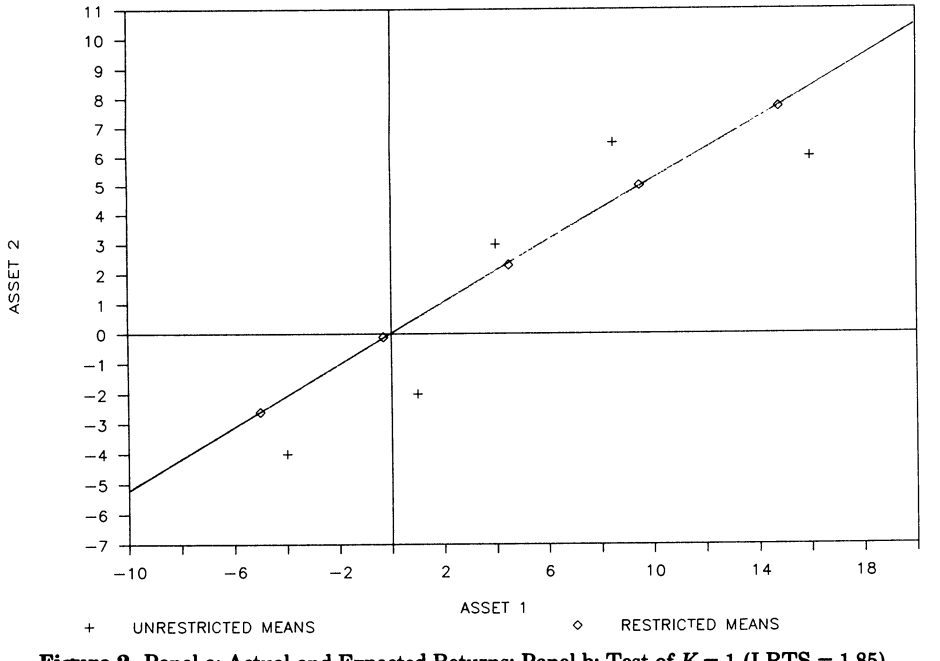

Figure 2. Panel a: Actual and Expected Returns; Panel b: Test of $K = 1$ (LRTS = 1.85)

dispersion around the unrestricted two-week means with the dispersion around the two-week means restricted to a subspace of lower dimension, we obtain a measure of how closely the data conform to a particular dimensionality hypothesis about expected returns.

Our assumption of a two-week step function is but one of many ways to approximate expected returns that move smoothly through time. One could, for example, model expected returns as an autoregressive process or as a function of specified instrumental variables. The costs of such alternative approaches include assuming a specific functional form for the relation of expected returns to past expectations or to the instruments. Furthermore, if equation (2) is examined and if the covariance restriction (3) is imposed, it is implicitly assumed that the market information set ϕ_t is completely captured by the instruments. In other words, if the restriction (3) is violated, the rejection could simply result from the market using more information to form expected returns than is captured by the instruments.[3] Of course, a rejection could also result if our approximation that expectations change only every two weeks is a poor one. Empirically, we observe that this approximation seems to allow ample variation of expected returns through time.[4]

An appendix available from the authors shows the first-order conditions that maximum-likelihood estimates of the parameters in (4) must satisfy, shows how the conditions are modified when the step-function approximation is employed, and outlines the computational procedure.

III. Data

We study weekly returns on common stocks for the twenty-year period 1963 through 1982. Data are obtained from the Center for Research in Security Prices (CRSP) at the University of Chicago. Weekly returns are examined because they represent a compromise between the many daily observations within a given calendar time interval and the less severe measurement errors in monthly returns. For example, problems arising from nonsynchronous trading, rounding errors, and bid-ask spreads should not be as important in weekly return data as in daily data. Weekly excess returns for each stock are computed by compounding the total daily returns reported by CRSP over a calendar week and subtracting the return on a one-week U.S. Treasury bill.

The test assets are ten portfolios of stocks from the New York and American Exchanges. Stocks are ranked by market value at the end of each year and assigned to deciles for the following year. The market value for a particular stock is the price per share multiplied by the most recent number of shares outstanding

[3] Alternative tests of intertemporal equilibrium models based on Euler equations stated in terms of marginal utility of consumption, such as Hansen and Singleton [21], can be robust to missing information, although Bergman [1] suggests that consumption-based models are less robust to the form of consumer preferences than are specifications such as (1).

[4] The estimates of expected excess returns for the models discussed below have sample time-series variances that range across models and subperiods from a low of three percent to a high of sixty-two percent of the total variance of weekly excess returns.

as reported by CRSP. Each portfolio is a value-weighted portfolio of the firms in its decile, where the weights change every week. Relative market values sum to unity across portfolios at each observation.

Table I presents descriptive statistics for the excess returns and relative market values of the ten portfolios. The relative values summarized in Table I are observed every two weeks, and they are used as the weights (w_t) in (2). Note that the decile of the largest firms accounts for almost seventy percent of the total market value in the sample and that the average fraction declines rapidly across deciles—to less than two tenths of one percent for the decile of smallest firms. The returns in Table I illustrate well-known "size effects." The returns of the smaller firms tend to have larger autocorrelations and standard deviations than the returns of larger firms. The small-firm portfolios tend to have higher average returns, but the relation is not monotonic in this sample. The smallest decile returns and values appear to be slightly more positively skewed and kurtotic than the other deciles.

Table I

Summary Statistics of Test Assets January 9, 1963 through December 30, 1982

Portfolio Weekly Excess Return[a]	Mean	S.D.	ρ_1	Skewness	Kurtosis	Studentized Range
Smallest	0.126	2.732	0.33	0.39	2.41	8.87
2	0.166	2.484	0.34	0.00	1.83	8.97
3	0.186	2.445	0.28	−0.01	1.48	9.13
4	0.183	2.384	0.27	−0.11	1.30	9.10
5	0.140	2.368	0.26	−0.23	1.15	8.64
6	0.155	2.258	0.24	−0.18	1.20	8.64
7	0.123	2.178	0.23	−0.17	0.95	7.82
8	0.100	2.124	0.20	−0.21	1.05	7.90
9	0.079	2.032	0.17	−0.09	0.91	7.60
Largest	0.031	1.960	0.04	0.01	1.25	8.08
Relative Market Value[b]	**Mean**	**S.D.**	ρ_1	**Skewness**	**Kurtosis**	**Studentized Range**
Smallest	0.162	0.096	0.96	2.87	9.43	5.80
2	0.340	0.115	0.99	1.05	1.01	4.98
3	0.614	0.188	0.99	0.49	−0.38	4.18
4	0.978	0.267	0.99	0.40	−0.39	4.23
5	1.537	0.410	0.99	0.59	0.29	4.70
6	2.409	0.439	0.99	−0.68	−0.78	3.91
7	3.842	0.636	0.99	−0.27	−1.15	3.90
8	6.786	0.842	0.99	−0.12	−1.08	4.07
9	14.048	0.958	0.98	−0.15	−0.92	4.14
Largest	69.284	3.463	0.99	0.30	−1.12	3.90

[a] Percent returns in excess of the one-week Treasury bill rate are shown for ten value-weighted portfolios of common stocks on the New York and American Stock Exchanges, formed by decile rankings of market value of equity outstanding at the end of the previous year. The number of observations is 1024. S.D. denotes the sample standard deviation, and ρ_1 is the first-order autocorrelation. Sample kurtosis is measured in excess of 3.0; thus, a value of zero indicates mesokurtosis, negative values indicate platykurtosis, and positive values indicate leptokurtosis.

[b] Relative market values are observed every two weeks (512 observations) and are normalized to sum to one hundred percent across portfolios.

IV. Empirical Results

A. Tests of Hypotheses

We follow the convention of many asset-pricing studies of conducting the analysis in nonoverlapping subperiods; thus, the constant parameters of the model are held fixed over four 256-week periods. For each subperiod, we test the three hypotheses discussed in Section I: (a) $K = 0$, $a_t \neq 0$, corresponding to conditional mean-variance efficiency of the market proxy; (b) $K = 1$, $a_t = 0$, corresponding to exact pricing with a single time-varying risk premium; and (c) $K \leq 1$, $a_t \neq 0$, corresponding to the hypothesis that there is a single, time-varying extra-market risk premium. (Models with larger numbers of risk premiums were also tested, but these results are not reported.)

We conduct simulation experiments to investigate the behavior of our test statistics. The first set of experiments examines behavior under the null hypotheses. For each hypothesis, three simulation trials are conducted. Each trial uses estimates of the model parameters to form conditional mean excess returns. The parameter estimates are chosen from the actual 256-week subperiods that capture the widest range of parameter values. (The first three subperiods are used; the estimates from the fourth subperiod are within the range of values provided by these.) Table II provides summary statistics describing the parameter values employed in the trials. The average a_t values range from -3.28 to 29.49, similar to the wide range of estimates of relative risk aversion that have appeared in recent studies.[5] Other summary measures of the parameters also indicate that the three trials include a wide range of parameter values. For example, the average "nonmarket" risk premiums range from -0.43 percent to $+0.37$ percent per week; the average variance of the value-weighted portfolio varies by more than a factor of five across the trials; and the average "market" beta of the small-firm portfolio ranges from 0.55 to 1.35.

For each simulation trial, we generate three hundred samples of artificial data by adding to the conditional mean vector a random error term independently drawn from a normal $(0, V)$ distribution, where V is the conditional covariance matrix specified for that trial. This procedure gives 256 weekly excess returns for each of three hundred samples. Each sample satisfies the null hypothesis at the chosen parameter values.

Summary statistics for the samples of the simulated test statistics are presented in Table III. The means, standard deviations, and most of the summary statistics are very similar across the three trials for each of the test statistics. Thus, the null distributions of the test statistics are not sensitive to the different values of the parameters chosen in these simulations. (The distributions are so similar that each of the three trials produces a virtually identical inference if the distribution for that trial is used to evaluate the sample test statistic.) Summary statistics for the pooled null distribution, combining the nine hundred samples from the three trials, are also shown in Table III.

[5] Merton [28] and Long [26] identify a_t as the coefficient of relative risk aversion for wealth, i.e., $a_t = -WJ_{WW}/J_W$, where $J(\cdot)$ is the indirect utility of wealth and subscripts on $J(\cdot)$ denote partial derivatives. Of course, a simple comparison of the estimates of a_t with estimates of other studies ignores many issues, such as differences between measured consumption and measured wealth.

Table II

Summary Statistics for Parameter Values Used in Simulations[a]

Null Hypothesis	Trial Number	Mean of Fitted a_t Values	Covariance Matrix ($\ln \mid V \mid$)	Value-Weighted Portfolio Conditional Variance ($w_t' V w_t$) Mean $\times 10^5$ (S.D. $\times 10^7$)	Fitted "Nonmarket" Premium (Ωh_t) Mean $\times 10^3$ (S.D. $\times 10^2$)	Average "Beta" ($\overline{Vw_t/w_t' Vw_t}$) of Size-Based Portfolio				
						Smallest	2	5	9	Largest
$K = 0$	1	29.49	−103.84	5.892	—[b]	0.98	1.10	1.17	1.01	0.98
		(154.74)		(2.356)		(0.02)	(0.02)	(0.01)	(0.01)	(0.01)
$a_t \neq 0$	2	0.73	−101.08	13.867	—	1.28	1.34	1.19	1.01	0.95
		(104.98)		(10.186)		(0.01)	(0.01)	(0.003)	(0.001)	(0.01)
	3	−3.28	−99.19	31.902	—	0.73	0.79	0.97	1.02	1.00
		(48.12)		(1.718)		(0.01)	(0.01)	(0.01)	(0.003)	(0.001)
$K = 1$	1	—	−104.40	—	3.68	—	—	—	—	—
					(1.29)					
$a_t = 0$	2	—	−101.55	—	0.57	—	—	—	—	—
					(1.17)					
	3	—	−100.15	—	−4.25	—	—	—	—	—
					(2.42)					
$K = 1$	1	26.72	−105.28	5.890	3.13	0.91	1.05	1.14	1.01	0.98
		(154.63)		(1.756)	(1.18)	(0.008)	(0.01)	(0.007)	(0.006)	(0.005)
$a_t \neq 0$	2	0.82	−102.45	13.866	0.76	1.35	1.37	1.21	1.01	0.95
		(104.60)		(10.981)	(1.56)	(0.005)	(0.004)	(0.001)	(0.001)	(0.006)
	3	−2.43	−100.66	31.456	−4.30	0.55	0.66	0.90	1.01	1.01
		(48.10)		(4.953)	(2.12)	(0.004)	(0.005)	(0.005)	(0.003)	(0.004)

[a] Standard deviations of time series of 128 parameter values are in parentheses.

[b] Indicates not applicable.

Table III

Simulation Evidence on the Behavior of the Likelihood-Ratio Test Statistics (LRT) Under the Null Hypotheses

Null Hypothesis: Normally Distributed Disturbances		Summary Statistics for Samples of Three Hundred LRTs and Pooled Samples of Nine Hundred LRTs					Mass of LRT distribution to right of percentile points of a normal distribution[a]		
		Mean	Standard Deviation	Skewness	Kurtosis[b]	Studentized Range[c]	Ten Percent	Five Percent	One Percent
$K = 0$	trial 1	1652.9	71.98	0.03	−0.02	5.83	0.110	0.050	0.010
$a \neq 0$	trial 2	1648.3	71.05	0.17	0.52	6.52*	0.110	0.057	0.010
	trial 3	1648.5	74.90	0.14	−0.20	5.84	0.103	0.043	0.017
	Pooled	1649.9	72.66	0.11	0.09	6.50	0.110	0.051	0.011
$K = 1$	trial 1	1638.9	73.72	−0.04	−0.30	5.50	0.077	0.047	0.013
$a = 0$	trial 2	1631.7	70.38	0.10	−0.46	5.26	0.117	0.057	0.010
	trial 3	1633.5	67.50	0.19	0.32	6.17	0.097	0.060	0.013
	Pooled	1634.7	70.57	0.08	−0.17	6.08	0.100	0.053	0.012
$K = 1$	trial 1	1461.3	63.51	0.10	−0.09	5.50	0.093	0.047	0.023
$a \neq 0$	trial 2	1449.1	68.12	0.16	−0.03	6.00	0.110	0.073	0.007
	trial 3	1457.1	66.78	−0.03	0.07	5.58	0.110	0.063	0.010
	Pooled	1455.8	66.29	0.07	−0.01	6.17	0.106	0.061	0.012

Table III—*Continued*

Null Hypothesis		Summary Statistics for Samples of Three Hundred LRTs and Pooled Samples of Nine Hundred LRTs					Mass of LRT distribution to right of percentile points of a normal distribution[a]		
Bootstrap Disturbances		Mean	Standard Deviation	Skewness	Kurtosis[b]	Studentized Range[c]	Ten Percent	Five Percent	One Percent
	trial 1	1656.4	73.84	0.07	−0.20	5.52	0.090	0.053	0.007
	trial 2	1654.6	66.77	−0.05	0.22	6.52*	0.097	0.063	0.007
$K = 0$	trial 3	1654.3	66.08	−0.03	−0.02	5.88	0.083	0.047	0.007
$a \neq 0$	Pooled	1655.1	68.92	0.01	0.02	6.45	0.100	0.052	0.009
	trial 1	1628.8	71.44	0.18	−0.39	5.12	0.120	0.043	0.010
	trial 2	1634.6	71.99	−0.19	−0.34	5.57	0.097	0.050	0.020
$K = 1$	trial 3	1632.9	68.86	0.22	−0.19	5.45	0.100	0.050	0.007
$a = 0$	Pooled	1632.1	70.82	0.66	−0.31	6.20	0.098	0.050	0.011
	trial 1	1451.8	65.03	0.09	−0.06	6.14	0.100	0.057	0.013
	trial 2	1454.3	64.85	0.08	−0.02	5.40	0.090	0.030	0.007
$K = 1$	trial 3	1457.0	72.56	0.27	0.44	6.67*	0.110	0.057	0.007
$a \neq 0$	Pooled	1454.4	67.65	0.17	0.24	7.44	0.100	0.047	0.010

[a] Mass of the simulated distribution to the right of the z-value implied by a normal distribution with mean and standard deviation equal to those of the simulated distribution.

[b] The sample kurtosis is measured in excess of 3.0, the value for a normal distribution; thus, a value of zero indicates mesokurtosis, negative values indicate platykurtosis, and positive values indicate leptokurtosis.

[c] An asterisk indicates that the sample value of the studentized range lies in the right five percent tail of the sampling distribution for the null hypothesis of normality.

The use of normally distributed disturbances in the simulations is a potential problem. An examination of the fitted residuals from equation (2) indicates some departures from normality. (The skewness, kurtosis, and studentized range are similar to those of the actual returns, summarized in Table I.) To examine the sensitivity of the null distributions to the use of normal disturbances, we repeat the simulations using a bootstrap approach (see Efron [11]), which samples randomly with replacement from the mean-centered residuals to generate disturbances. These experiments produce distributions of test statistics that are very similar to those using normal disturbances, as can be seen in the lower panel of Table III.

The null distributions of the test statistics summarized in Table III appear to be well approximated by normal distributions. The skewness and kurtosis are very close to what is expected given normality. Only three of eighteen studentized-range statistics for the individual trials are significant at the five percent level, and none appear significant at the one percent level. Close conformance to a normal distribution is reflected in the ten percent, five percent, and one percent tail areas as well. The pooled distributions with nine hundred samples conform even more closely to normal distributions than do the three individual trials. The pooled, bootstrapped null distributions are used to evaluate the sample statistics, but virtually identical inferences are obtained using the normal simulations.

Table IV summarizes the results of tests of three asset-pricing models. The tests do not reject the hypothesis $K \leq 1$, $a_t \neq 0$ in any individual subperiod. The evidence against exact pricing with a single premium ($K = 1$, $a_t = 0$) is also very weak; the smallest right-tail p-value is 0.304 in the first subperiod. Thus, given the alternative hypothesis, the tests cannot reject asset pricing based on a single, time-varying risk premium when that premium is not constrained to depend on market weights. The tests of conditional mean-variance efficiency ($K = 0$, $a_t \neq 0$) in Table IV produce different results. In the first subperiod (when a size effect on average returns is most pronounced in these data), conditional mean-variance efficiency of the value-weighted index is strongly rejected. The fourth-subperiod p-value (0.118) is not quite significant at conventional levels; the third-subperiod p-value exceeds twenty-five percent, and the second-subperiod statistic lies close to the center of the null distribution.

Table IV also presents (in the right-hand column) an overall p-value for each hypothesis. These overall tests assume that each of the four subperiods provides an independent draw from the same null distribution of the test statistic. The mean of the four sample values is compared with the null distribution of the mean, assuming normality. The overall tests reject conditional mean-variance efficiency at the three percent level and produce large p-values for the other hypotheses.

The simulated null distribution for the case of conditional mean-variance efficiency uses the fitted a_t's. As Table II indicates, those fitted values exhibit substantial volatility, which allows the fitted expected returns in the simulations to fluctuate substantially over time. For the small-firm portfolio, the time-series variance of the fitted expected excess returns equals twenty-four percent of the variance of the ex post excess returns, averaged over the three trials. For the portfolio of the largest firms, the figure is fifty-two percent.

Table IV

Tests of Asset Pricing[a]

Null Hypothesis[b]	First Subperiod January 9, 1963 through December 20, 1967		Second Subperiod December 27, 1967 through February 7, 1973		Third Subperiod February 14, 1973 through January 25, 1978		Fourth Subperiod February 1, 1978 through December 3, 1982		Overall Tests	
	Sample Value of LRT	p-value[c]	Sample Value of LRT	p-value	Sample Value of LRT	p-value	Sample Value of LRT	p-value	Average Value of LRT	p-value[d]
$K = 0, a_t \neq 0$	1806.75	0.011	1639.14	0.582	1695.67	0.278	1734.17	0.118	1718.93	0.029
$K = 1, a_t = 0$	1671.06	0.304	1529.24	0.926	1453.94	0.999	1539.39	0.907	1548.41	0.993
$K = 1, a_t \neq 0$	1436.74	0.598	1287.67	0.993	1323.25	0.977	1364.43	0.914	1353.02	0.999

[a] Test assets are ten portfolios of common stocks grouped by firm-size decile (weekly data). K is the number of time-varying, "nonmarket" risk premiums. If $a_t = 0$, market weights do not directly enter the pricing relation.

[b] The null hypotheses are versions of equation (1).

[c] LRT is the likelihood-ratio test statistic. The p-value is the proportion of nine hundred bootstrapped LRTs summarized in the lower panel of Table III exceeding the sample value of the LRT.

[d] The overall p-value is the probability that a normal random variable distributed as the mean of three draws from the pooled bootstrapped null distribution will exceed the mean sample value of the LRT.

Further simulations suggest that the rejection of conditional mean-variance efficiency is not sensitive to the amount of time-series variation in expected excess returns that the null distributions assume. We generate null distributions where the fitted a_t's are subjected to varying degrees of smoothing, up to the point where a_t is held constant over the period (in which case variation in the expected excess returns comes only from w_t). The smoothing is performed by taking moving averages of the fitted a_t's over progressively longer periods. For each of these simulations, the conditional covariance matrix is chosen so that the unconditional covariance matrix is held fixed at the same value. The null distributions and the implied p-values for the tests in Table IV are very similar for all the cases simulated.[6]

B. Interpreting the Results

There are several possible explanations for rejecting conditional mean-variance efficiency while not rejecting a general single-premium model. One interpretation is that expected returns vary with a single risk premium that is not a market portfolio. For example, a maximum-correlation portfolio with consumption (e.g., Breeden [3] and Breeden, Gibbons, and Litzenberger [4]) could be mean-variance efficient but have portfolio weights that differ from market-portfolio weights.

A rejection of mean-variance efficiency can be consistent with the CAPM, given that the "true" market portfolio is not measured and tests are conducted on subsets of assets. For example, it is possible that a constant combination of the size portfolios is a better proxy for the unobserved market portfolio than is a value-weighted combination. It is also likely that tests of mean-variance efficiency and exact pricing with K risk premiums have different sensitivities to the problem of missing assets. If the omitted assets are correlated with the included assets, then expected returns on the subset of included assets may be given by equation (1), with $K > 0$, even if the "true" market portfolio is mean-variance efficient. However, if exact pricing with K risk premiums ($a_t = 0$, $K > 0$) holds for all assets, then it will hold also on a subset of assets.

The tests examine a joint hypothesis including the statistical assumptions, such as the constant conditional covariances. Of course, it is possible that the rejection of mean-variance efficiency reflects departures from these assumptions or other statistical misspecifications.

Some previous studies appear to find that more than a single factor is priced in the sense of earning an average excess-return premium (e.g., Roll and Ross [31]). A comparison with studies that focus on average or unconditional risk premiums, however, can be misleading. Such tests typically make stronger assumptions about stationarity, factor structure, and independence than does the present approach. Perhaps a single time-varying premium could appear as multiple premiums when these restrictions are imposed.

[6] In another set of experiments, we formed moving averages of the product $a_t w_t$ using 2 to 128 two-week periods in the average. (The latter corresponds to constant expected risk premiums satisfying unconditional mean-variance efficiency.) Using these simulations as null distributions also produces p-values very close to those reported in Table IV.

C. Power of the Tests

The evidence above indicates that mean-variance efficiency of the value-weighted index can be rejected and that this rejection appears to be insensitive to the variability in expected risk premiums. It is possible that the exact single-premium pricing hypothesis is not rejected because of low power of these tests.

Table V reports simulation evidence on the power of the tests against alternative hypotheses with larger numbers of time-varying risk premiums. Three trials are conducted for each alternative using estimates of the parameters under the alternative model to generate the artificial data. For each trial, three hundred samples of data are generated by adding to the fitted means a normal random-disturbance term with covariance equal to the estimated V.[7] The empirical power reported in Table V is the fraction of three hundred samples in which a test of a given size rejects the null hypothesis when a particular alternative generates the data. The null distributions presented previously determine the critical values. Ten percent and one percent tests are summarized.

The results in Table V reinforce the findings in Tables III and IV. Like the distributions of the test statistics in Table III, those in Table V for a given hypothesis are similar in most cases across trials, indicating little sensitivity to the choice of parameters. (As in Table III, the parameters are chosen from the first three subperiods.) Table IV indicates that the null hypothesis of a market premium plus one nonmarket premium ($K = 1$, $a_t \neq 0$) cannot be rejected, and the bottom panel of Table V suggests that, if the data were generated as in our simulations with K larger than one [($K = 2$, $a_t = 0$) or ($K = 2$, $a_t \neq 0$)], the test would be very likely to reject the null hypothesis. The smallest empirical power against these alternatives exceeds eighty percent. The test for exact single-premium asset pricing ($K = 1$, $a_t = 0$) also seems to have some power against the two-premium model ($K = 1$, $a_t \neq 0$). The power for a one percent test ranges from 0.350 to 0.847 across the three trials. As might be expected, the test of single-premium pricing has virtually no power against the alternative of conditional mean-variance efficiency. The results in Table V also confirm the power of the test of conditional mean-variance efficiency. When the data are generated according to a single time-varying risk premium but the disturbances do not satisfy the covariance restriction (3) (i.e., $K = 1$, $a_t = 0$ is used), then this test rejects the null hypothesis in more than ninety-eight percent of the nine hundred samples.

V. Conclusions

This paper develops tests of asset-pricing models that allow expected risk premiums and market betas to vary over time. Using weekly data on ten portfolios of common stocks formed according to firms' equity capitalizations, conditional mean-variance efficiency of a value-weighted stock index is rejected for the

[7] Several cases were also conducted using the bootstrap approach to generate the data. As occurred with the null distributions (Table III), the results were virtually identical to those based on normally generated data.

Table V

Simulation Results for Power of Tests of Asset Pricing[a]

Null Hypothesis	Alternative Hypothesis	Trial	Power[b] of Test of Size		Summary Statistics of Simulated LRT under Alternatives		
			$\alpha = .10$	$\alpha = .01$	Mean	Standard Deviation	Studentized Range
Mean-Variance Efficiency ($K = 0, a_t \neq 0$)	$K = 1, a_t \neq 0$:	1	1.00	0.993	2014.3	74.29	6.23
		2	1.00	0.987	1992.1	74.03	6.70
		3	1.00	1.000	2021.9	75.93	5.71
		Pooled	1.00	0.993	2009.5	75.73	6.77
	$K = 1, a_t \neq 0$:	1	1.00	0.997	2027.4	75.89	5.27
		2	1.00	0.993	2007.5	72.89	6.06
		3	1.00	1.000	2023.8	78.54	5.72
		Pooled	1.00	0.998	2019.6	76.22	6.25
Exact Single-Premium Pricing ($K = 1, a_t = 0$)	$K = 0, a_t \neq 0$:	1	0.077	0.000	1626.9	67.10	5.91
		2	0.077	0.007	1628.0	69.51	7.42
		3	0.080	0.013	1623.8	71.29	5.53
		Pooled	0.078	0.010	1626.2	69.27	7.45
	$K = 1, a_t \neq 0$:	1	0.963	0.847	1856.7	68.06	5.64
		2	0.970	0.810	1862.2	70.45	5.83
		3	0.723	0.350	1765.2	75.76	6.35
		Pooled	0.886	0.658	1828.0	84.14	7.04
"Market" plus Additional Premium ($K \leq 1, a_t \neq 0$)	$K = 2, a_t = 0$:	1	0.987	0.860	1691.9	72.09	5.81
		2	0.997	0.923	1701.5	68.24	5.55
		3	0.997	0.913	1705.6	68.78	6.13
		Pooled	0.993	0.889	1699.7	69.88	6.31
	$K = 2, a_t \neq 0$:	1	1.000	0.973	1727.4	65.62	4.93
		2	0.993	0.950	1719.4	70.42	5.52
		3	0.997	0.923	1706.4	67.72	5.64
		Pooled	0.997	0.949	1717.7	68.42	6.06

[a] Based on three hundred samples for each null and alternative hypothesis. Pooled results are for nine hundred samples.
[b] Critical values are from the pooled simulated null distributions in Table III, assuming normality.

overall period, 1963 through 1982. The strongest evidence is observed in the 1963 through 1967 subperiod, when a "size effect" on average returns is most pronounced. Simulation evidence indicates that these results are insensitive to the amount of variation through time in expected risk premiums. A single risk-premium model of expected returns is not rejected if the premium is allowed to vary over time and if the risk measures associated with that premium are not constrained to equal market betas.

REFERENCES

1. Y. Z. Bergman. "Time Preference and Capital Asset Pricing Models." *Journal of Financial Economics* 14 (March 1985), 145–60.
2. T. P. Bollerslev, R. F. Engle, and J. M. Wooldridge. "A Capital Asset Pricing Model with Time-Varying Covariances." Working Paper, University of California, San Diego, September, 1985.
3. D. T. Breeden. "An Intertemporal Asset Pricing Model with Stochastic Consumption and Investment Opportunities." *Journal of Financial Economics* 7 (September 1979), 265–96.
4. ———, M. R. Gibbons, and R. H. Litzenberger. "Empirical Tests of the Consumption-Oriented CAPM." Working Paper, Stanford University, 1986.
5. J. Y. Campbell. "Stock Returns and the Term Structure." Forthcoming, *Journal of Financial Economics.*
6. G. Chamberlain and M. Rothschild. "Arbitrage, Factor Structure and Mean-Variance Analysis on Large Asset Markets." *Econometrica* 51 (September 1983), 1281–1304.
7. N. Chen. "Some Empirical Tests of Arbitrage Pricing." *Journal of Finance* 38 (December 1983), 1393–1414.
8. G. Connor. "A Unified Beta Pricing Theory." *Journal of Economic Theory* 34 (October 1984), 13–31.
9. J. C. Cox, J. E. Ingersoll, Jr., and S. A. Ross. "An Intertemporal General Equilibrium Model of Asset Prices." *Econometrica* 53 (March 1985), 463–84.
10. P. H. Dybvig. "An Explicit Bound on Individual Assets' Deviations from APT Pricing in a Finite Economy." *Journal of Financial Economics* 12 (December 1983), 483–96.
11. B. Efron. *The Jackknife, the Bootstrap, and Other Resampling Plans.* Philadelphia: Society for Industrial and Applied Mathematics, 1982.
12. E. F. Fama. *Foundations of Finance.* New York: Basic Books, 1976.
13. ——— and K. R. French. "Permanent and Temporary Components of Stock Prices." Working Paper, University of Chicago, 1986.
14. E. F. Fama and G. W. Schwert. "Asset Returns and Inflation." *Journal of Financial Economics* 5 (November 1977), 115–46.
15. J. A. Frankel and W. T. Dickens. "Are Asset-Demand Functions Mean Variance Efficient?" Manuscript, University of California, Berkeley, 1984.
16. M. R. Gibbons. "Multivariate Tests of Financial Models: A New Approach." *Journal of Financial Economics* 10 (March 1982), 3–27.
17. ——— and W. Ferson. "Testing Asset Pricing Models with Changing Expectations and an Unobservable Market Portfolio." *Journal of Financial Economics* 14 (June 1985), 217–36.
18. M. Grinblatt and S. Titman. "Factor Pricing in a Finite Economy." *Journal of Financial Economics* 12 (December 1983), 497–508.
19. R. E. Hall. "Intertemporal Substitution in Consumption." Working Paper, Stanford University, 1981.
20. L. P. Hansen and R. J. Hodrick. "Risk Averse Speculation in the Forward Foreign Exchange Market: An Econometric Analysis of Linear Models." In J. J. Frenkel (ed.), *Exchange Rates and International Macroeconomics.* Cambridge, MA: National Bureau of Economic Research, 1983, 113–46.
21. L. P. Hansen and K. J. Singleton. "Generalized Instrumental Variables Estimation of Nonlinear Rational Expectations Models." *Econometrica* 50 (September 1982), 1269–86.

22. J. E. Ingersoll, Jr. "Some Results in the Theory of Arbitrage Pricing." *Journal of Finance* 39 (September 1984), 1021–39.
23. J. D. Jobson and B. Korkie. "Potential Performance and Tests of Portfolio Efficiency." *Journal of Financial Economics* 10 (December 1982), 433–66.
24. D. B. Keim and R. F. Stambaugh. "Predicting Returns in the Bond and Stock Markets." *Journal of Financial Economics* 17 (December 1986), 357–90.
25. J. Lintner. "The Valuation of Risky Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets." *Review of Economics and Statistics* 47 (February 1965), 13–37.
26. J. B. Long, Jr. "Stock Prices, Inflation, and the Term Structure of Interest Rates." *Journal of Financial Economics* 1 (July 1974), 131–70.
27. D. Mayers. "Nonmarketable Assets and Capital Market Equilibrium under Uncertainty." In M. Jensen (ed.), *Studies in the Theory of Capital Markets.* New York: Praeger, 1972, 223–48.
28. R. C. Merton. "An Intertemporal Capital Asset Pricing Model." *Econometrica* 41 (September 1973), 867–87.
29. R. K. Rayner. "Generalized Instrumental Variables Estimation under Rational Expectations on First and Second Moments." Working Paper, Ohio State University, 1986.
30. R. Roll. "A Critique of the Asset Pricing Theory's Tests, Part I: On Past and Potential Testability of the Theory." *Journal of Financial Economics* 4 (March 1977), 129–76.
31. ——— and S. A. Ross. "An Empirical Examination of the Arbitrage Pricing Theory." *Journal of Finance* 35 (December 1980), 1073–1103.
32. S. A. Ross. "The Arbitrage Theory of Capital Asset Pricing." *Journal of Economic Theory* 13 (December 1976), 341–60.
33. ———. "The Capital Asset Pricing Model (CAPM), Short Sale Restrictions and Related Issues." *Journal of Finance* 32 (March 1977), 177–83.
34. M. S. Rozeff. "Dividend Yields are Equity Risk Premiums." *Journal of Portfolio Management* 10 (Fall 1984), 68–75.
35. W. F. Sharpe. "Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk." *Journal of Finance* 19 (September 1964), 425–42.
36. R. F. Stambaugh. "Arbitrage Pricing with Information." *Journal of Financial Economics* 12 (November 1983), 357–69.