Stability of the Arbitrage Pricing Theory Model Factors

Patrick A. Hays
Western Carolina University

David E. Upton
Virginia Commonwealth University

Carroll D. Aby, Jr.
Northwestern State University

This study examines the stability of empirically estimated arbitrage pricing theory (APT) model factor sensitivities. The Ashley stability test, which is exact and powerful, is applied to the coefficients of factors extracted from large returns datasets for both individual securities and portfolios. In most previous tests of APT models the stability of the factor sensitivities over the analysis period is implicitly assumed. The results from this study suggest that parameter instability exists for all factors in the APT model for both within group and holdout samples. The parameter instability exhibits no identifiable pattern or trend and appears to be in the form of random discrete jumps or shifts.

INTRODUCTION

The arbitrage pricing theory (APT) model suggested by Ross (1976) and extended by Huberman (1983), Shanken (1982), Dybvig (1983), Grinblatt and Titman (1983), and Stambaugh (1983), among others, assumes that investors behave as if security returns are generated by a k factor generating function. The generating function is assumed to be of the form:

(1) $R = E + BF + d$

where:

R	=	n vector of asset returns;
E	=	n vector of expected returns;

0747-5535/97/1300-0071 $2.50

F	=	k vector of the random deviations of the factor outcomes from their expected zero values;
B	=	$n \times k$ matrix of the sensitivity of the n asset returns to the k deviations in F; and
d	=	n vector of mean zero random error terms.

Other than stipulating that the number of factors is small, the model is mute on the number and nature of the factors.

Because the factors are exogenous to the model, tests of the statistical significance of a given factor structure are not tests of the APT. Tests of the APT focus on the implication that the expected returns E are linearly related to the factor sensitivities B:

(2) $E = a + bF + \varepsilon$

These tests, however, are confounded joint tests of both the APT and the specified factor structure. Stability of the factor sensitivities over the observation period is implicitly assumed. Factor analysis of security return samples has been used widely to empirically estimate the factor structure and associated sensitivities. The results of this empirical estimation have been unsatisfactory. The number of factors has varied widely both between studies and among data sets within a study. Roll and Ross (1980) use five factors in their model and conclude that at least three, but not more than four, factors are significant. Other researchers suggest that the number of economically significant factors is lower.

The focus of this paper is not on the statistical or economic significance of estimated factors. Rather, we focus on the implicitly assumed stability of empirically estimated factor sensitivities which underlies APT tests. The Ashley (1984) stability test (STAB) is used to examine the stationarity of factor sensitivities. In addition to being powerful, the STAB test is exact and "assumes little about the form of the instability, thus eliminating the need for separate tests against outliers, discrete parameter jumps, deterministic parameter drift, etc." (Ashley, 1984, p. 257) The factor sensitivities are based on five factor models estimated by methodology similar to that of Roll and Ross (1980). Factor structures are estimated from several sample sizes to investigate any sample size effects. Tests are applied to both individual assets and portfolios, both within sample and to holdout samples.

Testing proceeds in two stages. In the first stage tests are applied to all factors using two different arbitrary equal length subintervals. While these tests are sufficient to establish instability, they are insensitive and may understate the amount of instability. In order to examine the pattern of, and further verify, instability, subintervals are empirically inferred by combining contiguous periods with similar factor sensitivities. Stability tests are applied to first factor of the refined subintervals.

While tests of factor structure stability are not tests of the APT, such tests are important for several reasons. First, stability tests are necessary to validate the APT tests: instability would bias tests of economic significance downward, causing fewer factors to be accepted as significant. Second, understanding instability may allow future researchers to identify with more precision the underlying variables affecting asset returns. Third, any operational application depends upon correct parameter values. This paper also investigates the effect on stability of estimation from different sample sizes and uses portfolios to reduce the variation inherent in individual assets.

EMPIRICAL PROCEDURES

DATA

Two separate daily return data sets are used in the study. For the first sample 128 nonfinancial stocks are randomly chosen from those on the CRSP tape having complete data from January 2, 1980 to December 21, 1989. The returns over the week of October 19-23, 1987, are removed from the data set. To minimize the effects of skewness and serial correlation only every fourth daily return is retained (i.e., returns for days 1, 5, 9, ... , 2521 are retained), resulting in 630 observations for each security. This procedure was suggested to Roll and Ross by McEnally (Roll and Ross, 1980, p. 1095). Maximum likelihood factor analysis is used to extract a five factor model:

$$(3)\ R_{it} = a_i + b_{i1}F_{1t} + b_{i2}F_{2t} + b_{i3}F_{3t} + b_{i4}F_{4t} + b_{i5}F_{5t} + \varepsilon_{it}.$$

In order to judge the effect of sample size, factor structures are extracted using the returns for the first 50, and then the first 75, and then the first 100 stocks.[1] Factor scores are computed for the stocks in each group. Factor scores also are computed for the remaining 28 stocks, providing a holdout sample of stocks which are not used in factor structure estimation.

Individual assets often display a high variation in returns, resulting in relatively low R^2 values. This individual asset variation also degrades the stability tests. In order to reduce the effects of idiosyncratic risk and make the role of the factors more clear, factor scores for six equally weighted portfolios also are computed. Portfolios I, II, and III are created from stocks 1-10, 1-20, and 1-30 and are drawn from within the stocks used in estimating the factor structure (within sample). Portfolios IV, V, and VI are created from stocks 101-110, 101-120, and 101-128. None of the stocks in the latter portfolios is used in factor structure estimation, so that these portfolios are holdout samples.

Some authors have suggested that factors can only be successfully extracted from samples containing large numbers of firms.[2] Larger and more recent samples of returns over the period from January 3, 1989 to December 30, 1994 also are analyzed. In order to further test the effect of sample size while staying within the bounds of practicality, factor structures are extracted for independent samples of 200 and 300 stocks. Again, only every fourth observation is retained, resulting in 379 observations. Holdout samples of 20 individual stocks are independent of the factor structure extraction. Four equally weighted portfolios of 50 stocks also are used as holdout samples. The holdout portfolio stocks for the 200 stock sample are drawn from the 300 stock sample, while the holdout portfolio stocks for the 300 stock sample are drawn from the 200 stock sample.

---

[1] Models estimated from different sizes of data sets are referred to as *factor 50, factor 75*, and so on.

[2] We would like to thank an anonymous referee for urging us to add this further testing, which led to considerable strengthening of the results.

TEST METHODOLOGY

The usual procedure in APT testing is to examine the economic significance of the factor sensitivities. Because the focus of this paper is the stability of factor sensitivities, we instead compute the series of factor scores (values of $F_{\cdot t}$), which are tested for stability using the Ashley STAB test. Ashley (1984) analyzes the power of the STAB test relative to the Chow (1984) test, the Garbade (1977) varying parameter regression (VPR) test, and the Brown-Durbin-Evans (1975) cusum of squares (BDE) test for random walk, stable Markov process, and discrete jump types of parameter variability. The STAB test is superior to the BDE and the Chow tests for each type of parameter variation. The STAB test and VPR test have about the same power to detect random walk and stable Markov process parameter variation, while the STAB test appears to be superior for discrete jump parameter variation. Additionally, the STAB test assumes little about the form of the instability and is more generally applicable.

In order to apply the STAB test for the stability of the coefficient of a given factor, dummy variables D_j are created as follows:

D_1 = 1 for observations 1, 2, ... , n and 0 otherwise;
D_2 = 1 for observations n + 1, n + 2, ... , 2n and 0 otherwise;
. .
. .
. .
D_j = 1 for observations N - n + 1, N - n + 2, ... , N and 0 otherwise.

The test is optimized when interest is focused on the stability of one, or a small subset, of the coefficients. To examine the stability of factor 1, for instance, we estimate the following:

$$(4)\quad R_{it} = a_i + b_{i2}F_{2t} + b_{i3}F_{3t} + b_{i4}F_{4t} + b_{i5}F_{5t} + \gamma_{i1}D_1F_{it} + \ldots + \gamma_{ij}D_jF_{1t} + \varepsilon_{it}$$

where:

D_j = 1 for the j^{th} subinterval[3] and 0 otherwise.

The model given by equation (4) allows the coefficient for factor 1 to change over the j subperiods. The null hypothesis is:

$$H_o\colon \gamma_{i1} = \gamma_{i2} = \ldots = \gamma_{ij},$$

i.e., that the coefficient of factor 1 is stable for the subject stock or portfolio.

Under the null hypothesis the test statistic:

---

[3] The number of observations within each subinterval need not be the same. Ashley (1984, p. 259) notes: " ... where the parameter variation (or one's interest in it) is localized in one part of the sample, that part can be examined at higher resolution by using smaller subperiods ..."

(5) $STAB = [(SSE_R - SSE_F)/(j - 1)] / [SSE_F/(N - k - j + 1)]$

where:

SSE_R	=	Residual sums of squares from equation (3);
SSE_F	=	Residual sums of squares from the full model (4); and
k	=	Number of variables in equation (4).

STAB is distributed $F(j - 1, N - k - j + 1)$.

A graphic representation of the parameter variation, called a *stabilogram* by Ashley (1984), also is used. The stabilogram is a plot of a coefficient ± its standard error [e.g., $\gamma_1 \pm$ standard error (γ_1)] for each of the subintervals. This type of plot "plays a role analogous to that of the correlogram in Box-Jenkins modeling—it can be used to test formally the need for further specification modifications and it can be used informally to suggest the form such modifications should take" (Ashley, 1984, p. 257). This visual aid can be a helpful informal tool that suggests the form of the instability.

CHOICE OF SUBPERIODS

The tradeoff in the STAB test is between increasing the number of subperiods for more power (a better picture of coefficient variation or resolution) and the loss of degrees of freedom.[4] Because of this tradeoff, three different sets of subperiods are used. The first set of subperiods is naively chosen as five equal subperiods for each stock or portfolio. This permits only a coarse test of stability; a second naive set of ten equal subperiods for each stock or portfolio permits a finer resolution. To reduce the possibility that the naive equal subperiods may mask instability, an empirically determined set of subperiods is used to more closely examine the stability of the first factor. The five and ten subperiod tests are applied to all factors, but due to the large number of tests only the first factor is used for the empirical subperiods. The first factor is a logical choice for investigation because this factor is typically found to have an explanatory power much larger than the other factors, is highly correlated with the market portfolio, and tended to show greater stability than the other factors in the preceding tests.

To determine the empirical subperiods, the samples for the selected stocks and portfolios are partitioned into subperiods with ten observations each.[5] Where a coefficient plus half its standard error overlaps the next coefficient plus half the standard error, a model combining the two subperiods is tested. The combined subperiod model is retained if: 1.) the coefficient for the combined periods is at least as significant as the most significant of the separate periods, and 2.) the SSE is reduced. A stabilogram is used each time the equation (3) full model is estimated in order to visually inspect the overlap between the coefficients in contiguous periods.

The empirically inferred subintervals also provide some insight into the amount and pattern of instability. The procedure used is admittedly somewhat ad hoc. No claim is

[4] There is no criterion for assuring an optimal tradeoff. The relative power of the test using different subperiod lengths is discussed in Ashley (1984, p. 258).

[5] An exception is stock 110, which exhibited extreme instability. For this stock the number of starting subperiods is increased to 126 subperiods of five observations each.

made that this partitioning represents the optimal division of the samples in terms of the explanatory power or reduction of the residual sums of squares. The purpose of this test, however, is to ascertain whether significant instability appears to be present in factor 1. Any error introduced by this procedure would err on the side of conservatism. If significant instability in factor 1 is indicated by this ad hoc procedure, then a more optimal partitioning of the sample would result in a stronger indication of parameter instability in the model.

EMPIRICAL RESULTS

The results for the both individual stocks and portfolios indicate that the factor models are successful in explaining returns. For the first sample of 128 stocks all factor models are significant at or above the 0.05 level, and over 96 percent are significant at or above the 0.01 level. Table 1 gives the adjusted R^2 and regression F-values for the equation (3) APT models for six typical stocks (three within sample and three holdout) and all portfolios. The APT model typically explain 10 percent to 33 percent of the variation in the within sample individual stock returns and 8 percent to 25 percent of the variation in the holdout sample of individual stocks. Explanatory power for the portfolios is much higher, at 84 percent to 94 percent for within sample portfolios, and 57 percent to 68 percent for holdout portfolios. For both individual stocks and portfolios the within sample explanatory power of the factor scores decreases somewhat as the size of the estimation data set increases. The holdout portfolios exhibit increasing R^2 and F-values as the size of the estimation sample increases, but the improvement is small.

A possible explanation for the within and holdout sample differences is that the estimated factor structure partly reflects the idiosyncratic risk of the sample. As the estimation set increases in size, the assets in the within portfolios are a smaller proportion of the set and the estimated factor structure becomes more general. This would produce a poorer fit for the within portfolios, but a better fit for stocks in general. This capitalization on idiosyncratic variance indicates the importance of using holdout samples.

Table 1 also presents average R^2 and F-values for the 200 and 300 stock samples.[6] For the 20 stocks within sample, all models are significant at or above the 10 percent level. For the 20 individual holdout stocks 18 from the 200 stock model and 19 from the 300 stock model are significant at or above the 10 percent level. The R^2 for individual stocks ranges from about 2 percent to 54 percent, and as indicated by the averages, tends to be higher within sample than for holdout stocks. Both models are significant for all portfolios, with much higher R^2 than for individual stocks.

STABILITY TESTS USING EQUAL SUBPERIODS

For the 50, 75, and 100 factor scores parameter instability appears to be widely present in all individual stocks in the sample. Different factors, however, exhibit significant instability for different stocks, and the number of unstable sensitivities differs between stocks. Even the relatively low resolution five subperiod tests are generally sufficient to indicate instability in the factor model. Results from increasing the number of subperiods are mixed. Increasing the number of subperiods generally provides somewhat stronger

[6] Due to the large number of tests, only average results are presented.

Table 1

Panel A: R^2 and F-Values* for Individual Stocks

Stock	Factor 50 Scores Adjusted R^2	Factor 50 Scores F-value	Factor 75 Scores Adjusted R^2	Factor 75 Scores F-value	Factor 100 Scores Adjusted R^2	Factor 100 Scores F-value
10	0.329	62.43	0.262	45.68	0.226	37.78
20	0.282	50.48	0.334	63.98	0.300	54.99
30	0.188	30.13	0.201	32.58	0.200	32.42
101	0.115	17.31	0.121	18.34	0.127	19.34
110	0.233	39.14	0.235	39.74	0.226	37.69
120	0.082	12.20	0.110	16.49	0.113	16.99

Panel B: R^2 and F-Values* for Portfolios

Stock	Factor 50 Scores Adjusted R^2	Factor 50 Scores F-value	Factor 75 Scores Adjusted R^2	Factor 75 Scores F-value	Factor 100 Scores Adjusted R^2	Factor 100 Scores F-value
I	0.837	645.80	0.782	452.98	0.766	412.73
II	0.936	1835.91	0.895	1070.44	0.885	969.84
III	0.915	1348.80	0.886	982.11	0.876	887.75
IV	0.570	167.86	0.575	171.48	0.579	174.23
V	0.682	270.26	0.707	304.67	0.718	320.91
VI	0.701	295.89	0.724	331.68	0.733	346.96

Panel C: R^2 and F-Values* for Large Samples

	Average - Factor 200 Adjusted R^2	Average - Factor 200 F-value	Average - Factor 300 Adjusted R^2	Average - Factor 300 F-value
20 Stocks Within	0.214	24.241	0.232	27.656
20 Stocks Holdout	0.209	23.956	0.174	17.774
Four Portfolios Within	0.875	267.65	0.765	311.55
Four Portfolios Holdout	0.680	198.03	0.827	296.25

* All F-values significant at > .0001 level

indications of instability, but the results are not always consistent for a given factor and stock. This inconsistency may reflect the large variability shown by individual assets or the arbitrary subperiods used. The parameter instability is somewhat higher in the holdout samples, as may be expected.

Results for the portfolio tests also indicate widespread instability, with all of the portfolios exhibiting instability in at least one of the factor sensitivities. Again, the results of increasing the number of subperiods is mixed, and the results are not consistent. The holdout sample again exhibits a greater degree of instability. The larger portfolios exhibit a greater degree of instability in the holdout sample and a smaller degree of instability in the within sample, but the difference is small. These tendencies are consistent with the differences among R^2 and F-values noted above.

The rate of instability in the 10 subperiod large sample tests is shown in Table 2.[7] Significant instability is present and again occurs more often in the holdout samples. The stability of the model parameters is not enhanced by increasing the size of the data set. The rate of occurrence of instability in the 200 and 300 stock samples is at least equal to, and possibly higher than, the rate of instability in the smaller samples. Increasing the size

[7] Some firms in the sample exhibit unusually large returns, causing poorly conditioned matrices. Runs are made both on the original data set and with a data set in which firms with returns greater than 33 percent are removed. The reported results are on the restricted data set, which did not systematically differ from the results using the unrestricted data set.

Table 2—Rate of Occurrence of Instability in Large Samples, Ten Equal Subperiods

Panel A: Factor 200 Scores	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
20 Firms Within					
F Significant @ 0.1	17	17	18	18	18
Unstable @ 0.1	7	13	9	10	11
Unstable @ 0.05	6	8	3	9	10
20 Firms Holdout					
F Significant @ 0.1	19	19	19	20	20
Unstable @ 0.1	8	8	12	9	10
Unstable @ 0.05	5	5	7	9	10
Four Portfolios Within					
F Significant @ 0.1	4	4	4	4	4
Unstable @ 0.01	2	3	2	3	2
Unstable @ 0.05	1	2	2	2	2
Four Portfolios Holdout					
F Significant @ 0.1	4	4	4	4	4
Unstable @ 0.01	2	2	0	2	1
Unstable @ 0.05	2	0	0	1	1
Panel B: Factor 300 Scores					
20 Firms Within					
F Significant @ 0.1	17	17	18	17	18
Unstable @ 0.1	12	11	11	8	3
Unstable @ 0.05	9	9	9	5	2
20 Firms Holdout					
F Significant @ 0.1	19	19	19	20	20
Unstable @ 0.1	12	9	13	9	10
Unstable @ 0.05	10	7	10	6	7
Four Portfolios Within					
F Significant @ 0.1	4	4	4	4	4
Unstable @ 0.1	2	0	0	2	2
Unstable @ 0.05	2	0	0	1	1
Four Portfolios Holdout					
F Significant @ 0.1	4	4	4	4	4
Unstable @ 0.1	0	1	2	1	4
Unstable @ 0.05	0	0	2	1	3

of the generating data set from 50 to 75 or 100 increases the parameter stability of some stocks while decreasing the parameter stability of others. Therefore, no general statement about the effect of the size of the data set generating the factors on the parameter instability for individual stocks would be appropriate, except that increasing the generating data set's size above a well-diversified portfolio of 50 does not generally result in increased parameter stability. This suggests that the finding of instability is not caused by the size of the samples from which the factors are extracted.

REFINED STABILITY TESTS

The initial stability tests partition the sample into arbitrary subperiods because there is no *a priori* theory calling for any particular partitioning of the data set. These unsophisticated tests are adequate to indicate the presence of parameter instability in the APT model. Inconsistency in the results from different subperiods, however, suggests that the resolution of these tests is relatively low. Episodes of instability may be masked within the arbitrary periods. This suggests that refined tests based on inferred periods should be

Table 3—Refined Subperiods and Stability

Panel A: Factor 1, Stocks

	Stock 10	Stock 20	Stock 30	Stock 101	Stock 110	Stock 120
# Subperiods	15	15	13	9	19	17
Average Length	42	57.3	48.5	70	33.2	37.1
Maximum Length	170	250	160	280	215	130
Minimum Length	10	10	10	10	5	10

Panel B: Factor 1, Portfolios

	Portfolio 1	Portfolio 2	Portfolio 3	Portfolio 4	Portfolio 5	Portfolio 6
# Subperiods	10	14	14	6	12	10
Average Length	63	45	45	105	52.5	63
Maximum Length	130	80	130	200	130	160
Minimum Length	20	20	20	40	10	10

Panel C: Large Sample Tests

	Factor 200 Scores		Factor 300 Scores	
	Four Portfolios Within	Four Portfolios Holdout	Four Portfolios Within	Four Portfolios Holdout
Average # Subperiods	15.25	17.75	15.75	8
Average Length	25.18	21.35	24.06	92.08
Maximum Length	170	80	160	260
Minimum Length	10	10	10	10

used. Also, these coarse tests can indicate instability, but do not reveal its nature. Once instability is established, however, the stabilograms used in determining the subperiods also can be examined to obtain some idea of how the coefficients have varied over time. This can help determine whether the instability is in the form of a slow drift or an abrupt change and whether the instability is periodic or random. It also can suggest the size of the change in the factor sensitivities.

Due to the large number of tests, the refined tests are restricted to the first factor. The first factor appears to be stable in the majority of the arbitrary subperiods and exhibits greater stability than the other factors. This is consistent with previous studies. A finding of instability for this factor would imply instability in the higher order factors.

Results are shown in Table 3 for the factor 50, factor 200, and factor 300 models.[8] For both individual stocks and portfolios the number and length of derived subperiods varies widely. For the factor 50 models factor 1 is significant in the majority of the subperiods. Subperiods typically are short, with the minimum length of ten observations[9] the modal result. There are, however, widespread occurrences of long subperiods. In the large sample results all subperiods are significant at the 10 percent level. The length of the inferred periods usually is quite short. The average is just over 20 observations, with the minimum period length of ten observed most often. The distribution of period lengths is highly skewed, however, and long periods are observed infrequently.

Although the first factor is always significant over the total observation period, the results of the relatively high resolution stability tests point to parameter instability in the

[8] Factor 75 and factor 100 score stability are extensively sampled and are similar to the factor 50 score stability.

[9] The minimum subperiod length arises from the procedure used to refine the periods, which combine blocks of ten observations. This minimum is actually 40 trading days, as only every fourth observation is retained.

first factor. All of the stocks, both within sample and holdout, produce STAB values significant at or above the 1 percent level for the first factor. This pattern of significant instability continues in the portfolios. The higher occurrence of instability than in previous coarse tests indicates that the arbitrary subperiods mask instabilities.

Because the STAB tests confirm the presence of instability, the changes in factor sensitivity are examined for all 128 stocks and for all six portfolios. Few monotonic series of changes are observed, indicating that the instability is in the form of a random jump or shift rather than a drift. The size of the changes varies; in a few cases reversals to a significant negative sensitivity occur. No trend in either rate of occurrence or in size of change is evident, either for individual stocks or across stocks. The lack of a pattern across stocks seems to rule out market causes for the shifts. The same pattern is obtained for the portfolio instabilities, although the range of parameter variation is somewhat narrower.

SUMMARY AND CONCLUSIONS

This study uses the Ashley stability test to examine the factor structures extracted by factor analysis of returns. Several sizes of data set are used to extract the factor structures to determine if this affects stability. The focus of the paper is the stability of the estimated factor structure; no effort is made to ascertain the optimal number of factors or to measure the economic significance of the factors. Both individual stocks and stock portfolios are tested, and tests are applied both within sample and to holdout samples. Analysis of the R^2 and F-values of the factor structures indicates that the factor models are highly significant. Increasing the size of the estimation data set produces minor improvement in explanatory power for the holdout samples, but a minor degradation for the within samples. This may reflect idiosyncratic risk in the estimated factor structure. The presence of some inconsistency between the five and ten subperiod results suggests the need for more refined tests based on inferred subperiods.

Coarse tests using arbitrary subperiods indicate substantial instability for all factors. This instability is evident both within sample and for the holdout sample and for both individual stocks and portfolios. Analysis using inferred subperiods focuses on the first factor because

- It would be expected to be significant no matter how many other factors might be extracted according to the criteria employed;
- It explains more of the variability of the matrix of returns than other factors; and
- It is normally highly correlated with the appropriate market index return.

The tests indicate greater instability than the original tests, strongly suggesting that the factor sensitivities are not stationary over time. The parameter instability does not appear to have a discernible pattern or trend. The coefficients change via random discrete jumps or shifts. Finally, various sample sizes (up to 300 firms) to extract factors indicate that sample size is not the cause of instability.

Cross-sectional tests of the APT have generally found a significant relationship between returns and one or more factors extracted from factor analysis. These APT tests have generally been valid if, as Copeland and Weston (1983) state, "the factor loadings are stationary across time." The results of this analysis suggest that the assumption of stability of the factors over time may be inappropriate. This study is exploratory in nature and suggests several questions for future research. It would be particularly interesting to compare the stability of an individual factor with the factor significance in tests of the economic implications of the model. Finally, the present study should be extended to the stationarity of prespecified factors.

REFERENCES

1. Ashley, Richard A., "A Simple Test for Regression Parameter Instability," *Economic Inquiry*, 12 (1984), pp. 253-268.

2. Brown, R.L., J. Durbin, and J.M. Evans, "Techniques for Testing the Constancy of Regression Relationships Over Time," *Journal of the Royal Statistical Society*, Series B, 37 (1975), pp. 149-162.

3. Chow, G.C., "A Test of Equality Between Sets of Observations in Two Linear Regressions," *Econometrica*, 28 (1960), pp. 591-605.

4. Copeland, T.E., and J.F. Weston, *Financial Theory and Corporate Policy*, second edition (Reading, Mass: Addison-Wesley Publishing Company, Inc., 1983).

5. Dybvig, P., "An Explicit Bound on Deviations From APT Pricing in a Finite Economy," *Journal of Financial Economics*, 12 (1983), pp. 483-496.

6. Garbade, K., "Two Methods for Examining the Stability of Regression Coefficients," *Journal of the American Statistical Society*, 72 (1977), pp. 54-63.

7. Grinblatt, M., and S. Titman, "Factor pricing in a Finite Economy," *Journal of Financial Economics*, 12 (1983), pp. 495-507.

8. Huberman, G., "A Simple Approach to Arbitrage Pricing," *Journal of Economic Theory*, 28 (1983), pp. 183-191.

9. Roll, R., and S. Ross, "An Empirical Investigation of the Arbitrage Pricing Theory," *Journal of Finance*, 35 (1980), pp. 1073-1103.

10. Ross, S.A., "The Arbitrage Theory of Capital Asset Pricing," *Journal of Economic Theory*, 13 (1976), pp. 341-360.

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

12. Singleton, J. Clay., D. Solberg, and T.S. Zorn, "Instability in the Arbitrage Pricing Model," *The Northeast Journal of Business and Economics*, 12, no. 2 (1986), pp. 1-10.

13. Stambaugh, R., "Arbitrage Pricing With Information," *Journal of Financial Economics*, 12 (1983), pp. 357-369.