NEW TESTS OF THE APT AND THEIR IMPLICATIONS

by

Phoebus J. Dhrymes *
Irwin Friend **
Mustafa N. Gultekin ***
N. Bulent Gultekin **

Abstract

This paper provides new tests of the arbitrage pricing theory (APT). Test results appear to be extremely sensitive to the number of securities used in the two stages of the tests of the APT model. New tests also indicate that unique risk is fully as important as common risk. While these tests have serious limitations, they are inconsistent with the APT.

* Columbia University.
** University of Pennsylvania.
*** New York University.

1. Introduction and Summary

In this paper we report on a number of new tests on the empirical relevance of the arbitrage pricing theory (APT). These tests address more comprehensively the question of stationarity, i.e. the ability of risk measures from one period to "explain" returns in another, as well as other issues we have raised in our previous work. Thus, in DFG [1] and DFGG [2], we have shown that there is a general non-equivalence between factor analyzing small groups of securities and factor analyzing groups of securities sufficiently large for the APT to hold. As one increases the number of securities to which the factor analytic procedures are applied, the number of factors "discovered" increases, and this result cannot be readily explained by a distinction between "priced" and "non-priced" risk factors; in addition, we have also shown that it is generally impermissible to carry out tests on whether a given risk factor is "priced," though such tests are invariably found in the standard factor analytic literature used to test the APT.

Abstracting from the basic conceptual and empirical limitations of the factor analytic techniques used to test the APT, our prior analysis carried out comprehensive tests of the two key implications of APT -- those relating to (1) the irrelevance of unique (variance or standard deviation) as contrasted with common (covariance) measures of risk in the pricing of risky assets and (2) the risk-free or zero-beta rate of return interpretation of the constant term or intercept in the (linear) expected return function. Overall, the results obtained are, in large part, inconsistent with the APT model.

In a response to our recent findings, Roll and Ross (RR) [5] argue that the relevant point in tests of the APT is not the number of factors found (through factor analysis) but whether these factors are priced in the second stage of the tests. They claim that since a number of factors found to be significant in the first-stage time-series analysis of individual securities are not significant in the second-stage cross-return analysis across securities, it is quite likely that these factors are not "priced," and thus the number of "priced" factors might still be invariant to the number of securities factor-analyzed.

In this paper, we have repeated for much larger groups of securities the usual two-stage procedures used by RR [4] and other authors, and we find that the number of "priced" factors rises with the size of the group of stocks being factor-analyzed. As we have argued in our previous work, this procedure lacks rigor in that tests of significance on individual risk premia are invalid in this factor analytic context. Nonetheless, we carry out these tests so as to examine the issues raised in our earlier work, and, also, to appraise the response to it in the context advocated by RR and those who follow their lead. What we find is that, abstracting from these problems, the number of "priced" factors is not invariant to the number of securities factor analyzed; similarly, most of the factors extracted (from the first stage) appear to be "insignificant" at the second stage. The general tenor of our findings is that, as we increase the number of securities in each group, both the number of factors "discovered" at the first stage as well as the number of "priced" factors at the second stage increase, although most factors are not "priced." Needless to say, these findings do not constitute evidence favorable to the empirical relevance of APT.

In addition to updating our earlier results, we further expanded our analysis to pursue several new lines of inquiry, using data over the period 7/3/62 to 12/31/81. Three other new results seem worth mentioning. First, when factors (and unique or total standard deviation) estimated from one half of the period are used to explain returns from the other half, unique or total standard deviation performs as well as or better than the factor loadings. Second, when the number of time-series observations used by us (and earlier writers) to derive the relevant number of factors is increased (nearly doubled), the number of factors discovered also increases; this is consistent with the results we obtained in our earlier work when we broke the period originally covered into two subperiods. Third, tests on the constant term or intercept seem to depend both on the number of observations and the group size of securities factor-analyzed.[1]

2. The APT Model and Updating Our Earlier Empirical Tests

The APT model of Ross (1976) starts with the return-generating process for securities

$$r_{t\cdot} = E_{t\cdot} + f_{t\cdot}B + u_{t\cdot} \quad , \tag{1}$$

where $r_{t\cdot}$ is an m-element row vector containing the observed rates of return at time t for the m securities. $E_{t\cdot}$ is an m-element row vector containing the expected (mean) returns, while $f_{t\cdot}$ is a k-element vector of common (but unobservable) factors affecting security returns, both at time t. B is a $k\times m$ matrix of parameters, indicating the sensitivity of securities to the common factors. $u_{t\cdot}$ is the idiosyncratic component of the error term. Ross shows that if the number of securities (m) is sufficiently large, there exists a (k+1)-element row vector $c_{t\cdot}$ such that

$$E_{t\cdot} = c_{t\cdot}B^* \quad , \quad t = 1,2,\ldots,T \quad , \tag{2}$$

where $B^{*\prime} = [e\colon B']$ and e is an m-element column of ones. The empirical tests of the APT model are based upon a two-stage factor analytic approach.

In the first step, one determines the number of factors (k) and estimates the elements of B, denoted by $\tilde{B}$. In the second stage, using (the rows of) $\tilde{B}$ as "independent variables," one estimates the vector $c_{t\cdot}$, whose elements have the interpretation that c_{ti} is the "risk premium" attached to the i^{th} factor i = 1,2,...k, while c_{t0} is the risk-free rate (or possibly the return on a zero-beta asset). One estimates, at the second stage, the vector $c_{t\cdot}$ by GLS methods as

$$\tilde{c}'_{t\cdot} = \left(\tilde{B}^*\tilde{\Psi}^{-1}\tilde{B}^{*\prime}\right)^{-1}\tilde{B}^*\tilde{\Psi}^{-1}r'_{t\cdot} \quad , \quad t = 1,2,\ldots,T \quad . \tag{3}$$

There are two critical testable hypotheses implied by the APT model. First, the intercept term in (3) is the risk-free (or zero-beta) rate. Second, the risk premium vector is not null. In addition to these two hypotheses above, one could also test the APT model alternatively by introducing other explanatory variables in the model implied by (1) and (2). The restriction on empirical evidence implied by APT is that no other (relevant) economic/financial variables should have any effect on the determination of expected rates of return.

The data in this paper consist of daily stock returns from the CRSP tapes for the July 3, 1962-December 31,1981 period. Securities with more than 100 missing observations are deleted. This resulted in 900 New York and American Stock Exchange stocks, providing a sample size (number of observations) ranging from 4793 to 4893 daily returns per security. We rank these securities alphabetically to form groups of 30, 60 and 90 securities each.

The results of the first stage factor analysis tests are summarized in Table 1 for the entire sample period and two equal half periods; we do this separately for each of the three different group sizes. As showed in our earlier papers, the number of factors determined increases as both the number of observations and number of securities increase. We determine a 5-factor model for groups of 30-stocks, an 8-factor model for groups of 60 stocks, and a 13-factor model for groups of 90 stocks for each of the two subperiods. While such representation of security returns is adequate for the first half period, more factors are needed for the second half period for all group sizes. For the entire period, about 20% of the groups require more than 7 factors for groups of 30 stocks, 66% require more than 11 factors for groups of 60 stocks, and 50% require more than 17 factors for groups of 90 stocks.

The second-stage results are summarized in Table 2 for the two half periods and for the entire period. Column (1) of Table 2 shows that in the first period, at the 5% level, joint χ^2 tests conclude that the risk premium vector is significantly different from zero for only 5 (16.7%) out of 30 groups, for one group (3%) in the second period, and for 4 groups (13.3%) over the entire period (see last row of columns 6 and 12). This is in agreement with the results of our earlier papers which were carried out essentially for the first of these two periods and show that common risks are "priced" in very few of the groups. In sum, this analysis, based on using the customary groups of 30 stocks each, provides very little support for the key implication of the APT model.

Table 2 also presents results relevant to the question of whether 5 or 7 factor decomposition exhausts the "explanation" of the expected return process for groups of 30 stocks, (using the same methodology as in our previous work but covering a much longer time period). We test this by including (total) standard deviation of stock returns and, separately, the square root of the residual (specific) variance from the first stage as additional explanatory variables in the second stage.[2] Columns (2), (4), (7), (9), (12) and (14) in Table 2 show the relevant statistics for testing the null hypothesis of zero-risk premia when the factor risk premia are estimated in conjunction with other extraneous variables for subperiods and the entire period. Once (total) standard deviation (σ) or residual standard deviation (Ω) is included, the null hypothesis of a zero risk premia vector is rejected for only one group at the 5% level over the two subperiods and is uniformly accepted for the entire period. Both extraneous variables, however, are "priced" at least in 5 or 6 groups for the entire period and from 1 to 5 groups in the two subperiods (see columns (3), (5), (8), (10), (13) and (15)). The results are similar at the 10% level.

Overall, the implications of the updated test results reviewed in this section are similar to those obtained in our earlier papers and are, generally, not in accord with the implications of the APT model.

3. New Tests of Contributions to Asset Returns of Common Versus Unique Measures of Risk

In this section, we report on new tests of the basic implication of the APT model that only common (factor) risks are priced. In Section 2, common and unique variance measures are estimated within the same sample period, in which they serve as "explanatory" variables.

To minimize further the problem of "spurious" correlation between stock returns and risk measures and to test the robustness of such results, we derived the factor and unique measures of risk from the daily time-series observations in the first half period (1962-72) and used them to "explain" the daily cross-section returns for the second half period (1972-81). This technique should not only greatly reduce any concern about "spurious" correlations but should also indicate whether either the common or unique risk measures have any predictive value in assessing investors' return requirements in different assets.

The results summarized in Table 3 for three group sizes again indicate that both the common and unique measures of risk based on the 1962-72 data provide only extremely limited insights into prospective returns.[3] However, they also indicate that either total or residual standard deviation seems to be a more important determinant of stock returns than factor risk premia. Thus, the null hypothesis that none of the risk premia estimated in the first period is priced in the second period is rejected for only 1 out of 30 groups at the .05 level of significance (see column 2). Moreover, when total or residual standard deviation is added to the risk premia as explanatory variables, it is not possible to reject the null hypothesis for any group (see column 3). In contrast, in these last regressions, both own and residual standard deviation are significant for 5 groups at the .05 level. (columns 4 and 6). The results are qualitatively identical for groups of 60 and 90 stocks.

In separating the 1962-81 period into two halves for estimating risk and returns independently, it is clear why risk measures observed or at least available at the end of the first half are assumed to determine required returns for the second half rather than the other way around. However, we have replicated the analysis summarized in Table 3 by inverting the two sub-periods used to measure risk and returns though the rationale for this procedure would seem to be much weaker than for the results presented above.[4] The new results summarized in Panel B of Table 3 suggest a somewhat more important (though still relatively weak) role for factor risk premia in risky asset pricing, but again their effect on expected returns largely disappears when standard deviation is used as an additional explanatory variable (columns 3 and 5).

To summarize these results on the effect of unique versus common risk on explaining stock returns, unique risks as measured by residual standard deviation seem at least as important as common risks measured by factor risk premia. However, neither measure of risk contributes appreciably in explaining returns on individual securities. These results are inconsistent with the APT (as well as the CAPM), both of which as usually formulated deny any role to unique risk in the pricing of risky assets.

4. Effect of Increased Number of Assets on Estimated Number of "Priced" Factors

In this section, we attempt to determine the empirical relationship, if any, between the number of assets factor analyzed and the estimated number of "priced" factors from such a two-stage procedure. The results are presented in Table 4. The first-step factor analysis indicated that generally 7 factors were sufficient in explaining individual security returns for groups of 30 securities, 11 factors for groups of 60 securities, and 17 factors for groups of 90 securities (see Table 1). The second-stage cross-section GLS regressions for the groups of 30, 60 and 90 securities, respectively, were initially based on the 7, 11 and 17 factors determined from the relevant factor analysis. However, these tests are repeated by constraining the number of factors to 7 in the first-stage factor analysis for all three group sizes.

When we examined the observed percentages of groups with significant risk premia for a given factor, we do not find a monotonic negative relation between the proportions of significant risk premia and the ordering of factors; i.e., it is not always the first factor that is "significant" and the last few that are "insignificant." Rather, the relationship has an inverted U-shape or occasionally a rectangular shape. This is true whether we use a uniform number of factors (7) for each of the groups of 30, 60 and 90 securities or a different number of factors (7, 11, and 17) suggested by the factor analysis.

Greater insight into the nature of the relationship between the number of "priced" factors and the size of the group of assets factor analyzed can be obtained from an examination of the observed percetages of groups of 30, 60 and 90 securities respectively with at least a specified number of significant risk premia (either 1,2,...,7 for the groups of 30 securities, 1,2,...,11 for the group of 60 securities, and 1,2,...,17 for the groups of 90 securities, or 1-7 for each of the three sets of groups). Table 4 shows that when the factors are arrayed in natural order, the second-stage GLS regressions are more likely to yield at least one factor which is "priced" in the 90 stock groups (100% of the group coefficients have significant t-values at the .95 level) than is true for the 60 stock groups (73%) or for the 30 stock groups (43%). In the 60 stock groups, like the 30 stock, there are no groups with at least three "priced" factors. However, 10% of the 90 stock groups have at least three "priced" factors. When the cross-section regressions for the 30, 60 and 90 stock groups are based on the same number of 7 factors as found for the groups with the smallest number of securities factor analyzed (30), one finds that at least one or two factors are priced for the larger size groups (especially for the 90 stock groups) than for the 30 stock groups. In fact, there is considerable evidence in this analysis that in the 90 securities groups we have at least two significant factors at the .05 level; the evidence is less compelling that in the 30 security groups we have at least one significant factor.[5] Nevertheless, the number of "priced" factors found in the second-stage cross-section regressions is much smaller than the number of factors determined in the first-stage (factor) analysis.

Clearly, the evidence above suggests a positive relation between the number of "priced" factors and the number of assets in the groups of assets being factor analyzed. There are, however, two qualifications. First, it is conceivable, though not likely, that the difference in the observed results, while systematic and very large, might occur by chance. One, not very satisfactory, test of this explanation is to compare them with the expected number of groups with at least one, two, three, etc. "significant" factor risk premia on the null hypothesis of no factor effect on returns, using for this purpose a binomial distribution with probability of success $p = .05$ and the number of tosses (n) equal to the number of factors.[6] The expected value for the binomial distribution with different values of p and n are presented in Table 4 for comparison with the relevant observed results. Such a comparison indicates uniformly that the larger the number of assets factor-analyzed, the more substantial the difference between the observed and expected proportions of groups with at least one, two or three priced factors.

The second and more important qualification is that, for reasons explained in our earlier papers, it is inappropriate to use t-tests for the individual slope coefficients derived from the second-stage cross-section analysis as a measure of the significance of individual factor risk premia. A joint chi-square (χ^2) test, which can be used to determine whether the risk premia vector is null, is the appropriate one to use, though neither it nor the t-tests can indicate whether an individual factor risk premium is significant. The results obtained from such an analysis are summarized in Table 5 for 30, 60 and 90 security groups. This is done first using the 7, 11 and 17 factors determined without constraint from the time-series analysis of individual securities, and then constraining the number of factors to be the same, viz. seven, for each of the three groups of securities. These results are shown in Panel B.

The analysis based on χ^2 tests again shows a difference in the implications for the "priced" factors when we analyze groups with different numbers of assets. The proportion of 90 stock groups with a significant χ^2 statistic, which indicates that at least one of the risk factors is "priced," was much higher than the corresponding proportion for 60 stock groups, which in turn was higher than the proportion for 30 stock groups. This is true regardless of whether the unconstrained or constrained number of factors from the first-stage factor analysis of individual securities is used in the second-stage cross-section regressions.

These findings provide strong support to the conclusion that there is a positive relation (association) between the number of assets in the groups being factor-analyzed and "priced" factors. Once again, however, when own or residual variance is included, the risk premia vector is not "priced" and these extraneous variables are priced relatively more often (columns 3-6).

5. Effect of Increased Group Size on Intercept Test

Another important implication of the APT model, which it shares with other capital asset pricing models, is that the constant term in (3) corresponds to the risk-free rate (or at least the return on a zero-beta asset). This section provides tests about this second implication of the APT model and investigates whether intercept tests are affected by the size of the groups of assets. The relevant results are summarized in Table 6.

Using the entire period, we first test whether the intercept terms are jointly significant. Column (1) in Panel A of Table 6 indicates that we reject the hypothesis that all intercepts are zero.

Column (2) presents tests for the equality of intercepts to the risk-free rate. We use, as the risk-free rate, the seventh root of the (one plus) weekly Treasury Bill yield observed every Thursday. In this formulation, we assume that the seventh root of the Treasury Bill observed on a Thursday is the daily risk-free rate for Thursday and the next four trading days. Testing directly the hypothesis that the intercept is the risk-free rate when the weekly Treasury Bill rate is used for this purpose, we accept this hypothesis at the 5% level for the groups of 30 and 60 stocks but reject it for groups of 90 stocks at the 10% level. The test statistic in this case is shown in Column (2).

We also test whether all intercepts are equal. Chi-square values in Column (3) clearly show that we cannot reject this hypothesis.

We finally test the hypothesis that the intercept terms are equal to the risk-free rate using only observations on every Thursday to avoid the assumption that the Treasury Bill rates are constant for a week. This procedure results in a smaller number of observations (one-fifth of the previous number). Interestingly, we reject the hypothesis in this case for groups of 30 and 60 stocks at the 5% level and for all groups at the 10% level.[7] We also reject the equality of intercepts as well as equality to zero for groups of 30 and 60 stocks at the 10% level, but not for groups of 90 stocks.

As we indicated in our earlier paper (DFGG [2]), the intercept tests are somewhat mixed. When we use observations for the entire period with the exception of groups of 90 stocks, results are not inconsistent with the second implication of the APT model. Further work would be required to determine whether rejection of the risk-free rate interpretation at the 10% level for groups of 90 stocks is a random aberration or reflects the effects of the increased size of the groups factor-analyzed. Similarly, further work is needed to determine whether the rejection of the risk-free rate and the zero-beta interpretation of the intercept when we use observations on every Thursday is merely a result of the diminished number of observations.

6. A Rejoinder to RR's Reply and Summary

The purpose of this section is to provide a rejoinder to the "reply" by RR [5] appearing in the same issue of the Journal of Finance in which our original paper (DFG [1]) was published. In that paper, we made three major points: (a) the procedure used by RR in factor-analyzing 30 securities cannot be expected to yield reliable information on the factor loadings and hence on the crucial questions regarding risk premia, risk-free rates, zero-beta rates and the like; (b) testing for the "significance" of individual risk premia is not meaningful; and (c) that the number of factors "discovered" increases with the size of security groups analyzed. We shall indicate that their response to each of these points is either misleading or incorrect.

In response to (b), RR say that they were well aware of this point all along. Now, it is true that RR note in many places that the factor loadings one extracts are subject to rotation -- a fact found in all textbooks. However, the ancillary consequence that testing individual risk premia is therefore meaningless does not appear in such textbooks, and RR certainly do not mention this point. In fact, they provide statistics on how many groups had "one significant," "at least two significant," ... factors; to this effect, see Table III on page 1092. Their confusion on the issue most clearly emerges from the following quote (p. 1091):

> "In our case, however, constraining the sample design to the independent case is especially important because the λ's [estimated risk premia] at best are some unknown linear combination of the true λ's and testing for the number of priced factors or non-zero λ_j's is thereby reduced to a simple t-test."

One wonders how testing that an unknown linear combination of parameters is zero gives one information about how many of the underlying parameters are zero.

The same sort of confusion is also evident in other aspects of their reply. For example, in dismissing our finding that as the number of securities analyzed increases the number of factors also increases, as a result "expected" by them, RR state (p. 329 in their Reply [5]):

> "We want to take this opportunity to emphasize the irrelevance of the point that factor analysis extracts more factors with larger groups of securities or with larger time-series sample sizes.... To illustrate, suppose that a group of 30 securities contains just one cosmetics company. Factor analysis produces, say, three significant factors. If the sample is large enough we would certainly anticipate finding a fourth significant factor, a factor for the cosmetics industry."

If we take this at face value, it is just not clear what becomes of APT in this context. How could one argue for the validity of the model as originally presented and at the same time maintain that the number of "common" factors is indeterminate? If it has been conceded earlier that individual significance tests are meaningless, who is to decide whether with the addition of the "other" cosmetics firm what we are getting as a "significant" risk premium is not for the "irrelevant" cosmetics factor? What if we replace cosmetics by automobiles or oils? A proper procedure in each case would be to allow for the presence of group factors.

One wonders, of course, whether we would have been treated to a different menu if the original paper used 60 security groups; perhaps we would have been told that 5-7 factors were significant. What is more difficult to surmise is how they would have reacted to the 17 factors or so which would have resulted if their original paper used 90 security groups.

In this context, their reply also creates an unfortunate obfuscation; see footnote 2 (p. 349). In discussing the relation of factor loadings obtained from 240-security groups and those obtained from the constituent 30-security groups, we observed that the impression is widespread that the two sets are related by an orthogonal transformation. We pointed out that if the two matrices of factor loadings are related by an orthogonal transformation, then <u>as a matter of mathematical requirement</u> the columns[8] of the two matrices must have the same length, since an orthogonal transformation is distance preserving. This is <u>not a statistical test</u>, as we explicitly stated in our paper; it was simply an exercise to disabuse those who hold the view that one gets the same information from the two procedures except for the fact that the normalization is a bit different in the two cases. Thus, we certainly are not making an "unbelievable assertion" that this "particular test" has "infinite statistical power."

In their Reply, we also detect a substantial modification of what is to be meant by Arbitrage Pricing Theory. In the original paper by RR which we had criticized, the model being tested is the very soul of simplicity and parsimony. It claims that the return-generating function is composed of a "systematic" and a "random" component. The random component is then said to be composed of an "idiosyncratic" and a "common" component. Very few objections can be raised against such a framework. It is then the great virtue of the arbitrage pricing hypothesis, in conjunction with other technical requirements, that it yields a very appealing conclusion about the pricing of risky assets. In the process of this reasoning, it is essential that we leave the specification of the systematic component open since it will be the subject of the conclusion from the no-arbitrage hypothesis. If the number of "factors" responsible for the common components is "large," the parsimonious aspect of this model is lost. Alternatively, if we are to claim that these "factors" are ab initio concrete variables which the investigator is to specify, then we are straying away from the simplicity (and parsimony) of the original model and into the complexities of the usual multi-index model, where the role of arbitrage pricing is rather tenuous. For, if we are to specify ab initio the systematic component, then we are dealing with a situation far different from that addressed by Ross' original paper. We have questioned the empirical methodology in the paper by RR, but not the contribution to the literature made by Ross' earlier theoretical contribution.

Finally, we would like to comment on a point made at the end of the RR Reply, regarding the presentation of some of our results. In particular, they maintain:

> "We cannot fully discuss the tests of Section VII since they are not reported in full, but it is interesting to note that DFG adopted tests "like those used in RR' even though such procedures are 'subject to the basic limitations...discussed earlier in the paper' (p. 345). Despite these alleged limitations, however, DFG rely on them to produce results which they interpret as '...inconsistent with the APT model,' (p. 345). So having spent their entire paper criticizing the RR test procedures, DFG finally report results for which the tests are apparently satisfactory."

In our work, we stressed that testing for "significant" individual risk premia parameters is not meaningful in the context under consideration; if RR had not committed this error, we would not have found it necessary to point out this fact. Nonetheless, since the RR methodology has found wide acceptance, we did provide in our earlier paper (DFGG [2]) a number of results based on the RR methodology which turn out to be nonsupportive of the implications of arbitrage pricing theory models -- contrary to the assertions of RR. In fact, results reported in this paper show that, again in the 30-security context, whether we use a 10-year time series sample or a 20-year time series sample, the proper (joint) test of significance for the risk premia vector rejects its nullity in, at most, 10 out of 30 groups. When own (total) or residual standard deviation is introduced as an additional variable, then the hypothesis that the vector of risk premia is null is accepted by the proper (joint) test in 30 out of 30 groups, i.e. uniformly for the entire sample. It is difficult to imagine a more complete rejection of the crucial implication of such APT models, using the flawed methodology of splitting the universe of assets into 30-security groups.

FOOTNOTES

1 Due to space limitations, most of our results are summarized briefly. A working paper with more detailed results is available from the authors.

2 Specific variance is the $\tilde{\Omega}$ term in the covariance matrix, $\tilde{\Psi} = \tilde{B}'\tilde{B} + \tilde{\Omega}$, estimated by factor-analytic methods.

3 Due to space limitations, we only present the summaries of the tables. Complete tables are available in a working paper. The numbers in Table 3 correspond to the similar summaries in the last row of Table 2.

4 The rationale would presumably be based on an extremely long-term stationarity of the relative riskiness of stock returns.

5 If one uses 10% level of significance, the number of "priced" factors rises more dramatically with increasing numbers of stocks in a group.

6 One problem with this test is that the slope coefficients in our second-stage regressions are not statistically independent, but the alternative would be the use of a fairly arbitrary value for the constant term. As a result, it would be necessary to test the sensitivity of results to the constant term selected (presumably some measure of the risk-free rate), and the test would be conditional on the validity of the assumption that the constant term has the value indicated.

7 Similar results are reported by Gultekin and Rogalski [3] using government bonds.

8 Or, alternatively, the rows, depending on one's point of view.

REFERENCES

1. Phoebus J. Dhrymes, Irwin Friend and N. Bulent Gultekin, "A Critical Reexamination on the Empirical Evidence on the Arbitrage Pricing Theory," Journal of Finance 39, (June 1984), 323-346.

2. Phoebus J. Dhrymes, Irwin Friend, Mustafa N. Gultekin and N. Bulent Gultekin, "An Empirical Examination of the Implications of Arbitrage Pricing Theory," Journal of Banking and Finance, (March 1985), forthcoming.

3. N. Bulent Gultekin and Richard J. Rogalski, "Interest Rate Risk and the APT," Journal of Finance, (March 1985), forthcoming.

4. Richard Roll and Stephen A. Ross, "A Critical Reexamination of the Empirical Evidence on the Arbitrage Pricing Theory: A Reply," Journal of Finance 39, (June 1984), 347-350.

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

TABLE 1

Chi-Squared Tests of the Hypothesis That k-Factors Generate the Daily Stock Returns - Summary Results (number of times the null hypothesis is accepted at 5% level)

Group Size	# of Groups (1)	# of Factors for Half Periods (2)	# of Factors for Entire Period (3)	# of Groups with Significant Chi-Square Value: First Period (4)	Second Period (5)	Entire Period (6)
30 stocks	30	5	7	24 (80%)	9 (30%)	24 (80%)
60 stocks	15	8	11	9 (60%)	1 (7%)	6 (40%)
90 stocks	10	13	17	7 (70%)	1 (10%)	5 (50%)

The first half period covers 7/3/62-4/23/72, the second half period covers 4/24/73-12/31/81, and the entire period is 7/3/62-12/31/81. Groups are formed from alphabetically-ranked 900 securities from the daily CRSP tapes. Each group is then factor-analyzed. The number of groups in columns (4), (5) and (6) indicate that k-factor generating models shown in columns (2) and (3) for the half and entire periods, respectively, are adequate at 5% level, i.e. that chi-squared values are significant at 5% level for this many groups out of total number of groups shown in column (1). Figures in parentheses show percent of groups with significant chi-square in each set.

TABLE 2: TESTS OF SIGNIFICANCE FOR RISK PREMIA AGAINST SPECIFIC ALTERNATIVES - STANDARD DEVIATIONS AND RESIDUAL STANDARD DEVIATIONS AS ALTERNATIVES TO APT MODEL: SUMMARY STATISTICS FOR GROUPS OF 30 STOCKS

Group	1st Half Period (7/3/62-4/23/72)					2nd Half Period (4/24/72-12/31/81)					Entire Period (7/3/62-12/31/81)				
	B (a)	B & σ (b)		B & Ω (c)		B (a)	B & σ (b)		B & Ω (c)		B (a)	B & σ (b)		B & Ω (c)	
	χ_1^2	χ_2^2	t(σ)	χ_3^2	t(Ω)	χ_1^2	χ_2^2	t(σ)	χ_3^2	t(Ω)	χ_1^2	χ_2^2	t(σ)	χ_3^2	t(Ω)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
1	4.049	1.765	1.30	1.245	1.42	4.379	4.712	1.48	4.779	1.47	6.765	4.532	1.61	5.626	1.60
2	9.329*	3.428	2.10**	3.862	2.11**	3.770	3.512	-.53	3.620	-.62	12.961*	4.856	1.03	7.366	0.90
3	9.661*	1.738	2.12**	1.542	2.11**	4.036	2.309	1.32	2.913	1.30	15.790**	4.682	2.15**	5.953	2.15**
4	8.192	3.667	1.19	3.839	1.15	5.674	5.095	0.45	4.911	0.37	9.854	5.046	1.02	5.403	1.03
5	6.779	3.666	0.41	6.716	0.56	11.233**	7.695	0.13	8.076	0.04	9.757	3.142	0.47	7.222	0.78
6	5.090	3.613	0.18	4.117	0.19	5.078	4.392	0.71	4.143	0.65	6.448	3.669	0.48	4.504	0.29
7	12.398**	5.979	0.57	6.297	0.51	9.426*	7.117	0.58	7.008	0.62	13.848*	6.161	0.52	6.328	0.51
8	8.019	6.786	2.44**	8.627	2.38**	6.388	3.328	0.46	2.557	0.67	14.393**	6.460	1.46	7.749	1.33
9	16.166**	4.091	0.75	4.682	0.76	6.536	5.298	0.42	5.336	0.37	16.083**	5.603	1.05	7.457	0.94
10	10.198*	4.883	0.38	8.120	0.40	5.862	4.975	0.98	4.256	1.00	7.934	0.869	1.02	1.952	0.99
11	8.326	3.663	0.35	6.535	0.22	7.929	6.013	1.22	6.443	1.06	12.717*	6.655	1.59	5.867	1.49
12	4.140	2.800	0.64	3.254	0.51	5.059	2.690	0.63	2.717	0.46	7.325	4.561	1.07	4.393	0.71
13	5.326	1.775	1.16	2.247	1.20	3.475	3.337	0.48	3.444	0.54	6.444	4.038	0.90	6.461	0.91
14	5.499	4.170	0.14	5.550	0.21	6.578	2.588	2.36**	4.520	2.20**	8.979	3.716	1.87*	8.259	1.88*
15	8.052	6.379	1.02	7.177	0.96	2.135	1.666	0.65	1.196	0.69	5.945	3.178	0.66	3.283	0.62
16	14.783**	8.036	0.64	8.624	0.75	4.980	4.875	1.25	4.537	1.32	12.057*	5.870	1.21	5.943	1.20
17	2.699	1.545	0.73	1.093	0.77	7.157	0.838	1.62	1.237	1.61	7.259	0.799	1.73*	1.010	1.77*
18	6.411	5.145	0.30	5.150	0.24	3.891	3.903	-.22	3.983	-.32	6.411	5.789	0.59	6.882	0.68
19	3.601	0.638	2.51**	0.705	2.47**	10.488*	5.072	1.05	3.402	1.24	10.904	2.117	2.05**	2.563	2.05**
20	16.807**	4.224	1.62	4.611	1.66*	4.324	2.417	2.14**	2.428	2.14**	19.027**	2.300	1.71*	5.903	1.80*
21	7.391	1.279	1.94*	2.576	1.90*	7.753	5.043	0.72	4.570	0.74	13.531*	2.693	1.27	3.270	1.16
22	9.266*	4.002	1.44	4.836	1.49	4.639	3.819	1.02	3.836	0.92	9.560	3.850	1.65*	4.398	1.56
23	4.799	1.766	1.86*	2.575	1.99**	4.594	4.217	-.21	4.030	-.28	6.479	1.623	1.08	2.330	0.85
24	6.647	2.845	2.02**	2.472	2.02**	1.222	0.322	0.49	0.381	0.44	4.346	1.892	2.00**	5.238	2.09**
25	2.886	1.488	1.06	1.372	0.98	9.809*	12.570**	2.83**	12.716**	2.88**	12.087*	9.472	2.78**	11.551	2.66**
26	6.756	3.422	1.60	3.330	1.53	7.479	7.279	1.02	7.667	1.07	6.421	4.922	1.92*	5.483	1.94*
27	5.276	0.530	1.08	0.713	1.13	2.426	1.897	1.64	2.119	1.55	9.433	2.376	1.27	3.260	1.23
28	6.788	2.933	0.46	5.392	0.37	4.753	4.763	-.33	4.551	-.28	8.445	5.437	0.26	7.334	0.18
29	7.805	2.762	0.85	2.675	0.91	5.742	6.041	1.00	5.983	1.09	11.420	6.905	1.17	6.159	1.25
30	16.463**	8.973	1.71*	10.135*	1.72*	8.930	6.892	1.00	5.984	0.89	10.790	3.089	2.25**	3.208	2.19**
Number of Significant Statistics															
at 10%	9	0	8	1	9	4	1	3	1	3	10	0	10	0	9
at 5%	5	0	5	0	6	1	1	3	1	3	4	0	5	0	5

a/ χ^2 (chi-square) values test the null hypothesis that the risk premia vector is null. The risk premia is estimated by the GLS model $\tilde{c}'_{t.} = (\tilde{B}^*\tilde{\Psi}^{-1}\tilde{B}^{*\prime})^{-1}\tilde{B}^*\tilde{\Psi}^{-1}r'_{t.}$. $B^{*\prime}$ is the [e:B] matrix of factor loadings with unit vector e.

b/ Own standard deviation of daily stock returns are included as an additional explanatory variable in B^*. Chi-square value tests the nullity of the risk premia vector while t(σ) is the t-ratio for the regression coefficient for own standard deviation.

c/ Square root of residual (or specific) variances (Ω) is included as an additional explanatory variable in B^*. Residual variance corresponds to the diagonal elements of $\tilde{\Omega}$ in $\tilde{\Psi} = \tilde{B}'\tilde{B} + \tilde{\Omega}$ from the factor analysis. t(e) is the t-ratio for the regression coefficient for residual variance.

The number of factors is 5 for the half periods and 7 for the entire period, respectively, and these number should be used for the degrees of freedom for chi-square tests. * and ** indicate groups for which nullity of risk premia is rejected at 10% and 5% levels respectively.

TABLE 3: Tests of Significance for Risk Premia Against Specific Alternatives Using Risk Measures From One Half Period and Returns from the Other Half

Period: Number of groups with significant test statistics at 5% level [(a)]

Group Size	# of Factors (1)	Independent Variables: B [(a)] (2)	B (3)	σ [(b)] (4)	B (5)	Ω [(c)] (6)
A. Risk measures are estimated from the first half period [(e)]						
30	5	1 (3%)	0	5 (17%)	0	5 (17%)
60	8	1 (7%)	0	3 (20%)	0	3 (20%)
90	13	0	0	2 (20%)	2	2 (20%)
B. Risk measures are estimated from the second half period						
30	5	6 (20%)	0	6 (20%)	0	4 (13%)
60	8	5 (33%)	1 (7%)	7 (47%)	3 (20%)	8 (53%)
90	13	4 (40%)	4 (40%)	6 (60%)	3 (30%)	6 (60%)

(a) These are summary results showing number of groups with significant risk premia using a chi-square test. These numbers correspond to the similar summaries in the last row of Table 1. Detailed tables are available from the authors.

(b) Factor loadings are the only set of explanatory variables.

(c) Own standard deviation is included as an additional independent variable to the factor loadings.

(d) Squared root of the residual variance is added as an additional independent variable (see footnotes b and c in Table 2).

(e) Risk measures (i.e. factor loadings, own standard deviation and residual variance) are estimated from the daily return during the first half of hte period (7/3/62–3/23/72). These parameters are independent variables in the GLS model in (3) using the daily returns in the second half period (3/24/72–12/31/81). In Panel B, this order is reversed.

TABLE 4: TESTS OF SIGNIFICANCE FOR INDIVIDUAL RISK PREMIA USING t-TESTS AT 5% LEVEL: SUMMARY RESULTS (7/3/62-12/31/81)

	Number of Factors											
	1	2	3	4	5	6	7	8	9	10	14	17
Expected percent of groups with at least this number of risk premia significant using 7-factor model [a]												
	30.160	4.4360	0.3760	0.02000	0.00125	0.000658	0.00648					
Observed percent of groups with at least this number of risk premia significant using 7-factor model [b,c]												
Groups of 30 stocks	43.33	10.00	0.00	0.00	0.00	0.00	0.00					
Groups of 60 stocks	93.33	13.33	0.00	0.00	0.00	0.00	0.00					
Groups of 90 stocks	100.00	50.00	0.00	0.00	0.00	0.00	0.00					
Percent of groups with this factor's risk premium significant in natural order from factor analysis [d]												
Groups of 30 stocks	6.67	10.00	10.00	3.33	13.33	10.00	0.00					
Groups of 60 stocks	13.33	0.00	40.00	33.33	13.33	6.67	6.67					
Groups of 90 stocks	10.00	20.00	40.00	20.00	20.00	20.00	30.00					
Percent of groups with at least this number of risk premia significant [c]												
Groups of 60 stocks (11-factor model)												
expected	43.1199	10.1894	1.5235	0.1552	0.0111	0.0005	$.12 \times 10^{-4}$	$.7 \times 10^{-4}$	$.1 \times 10^{-6}$			
observed	73.33	40.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
Groups of 90 stocks (17-factor model)												
expected	58.1879	20.7772	5.0253	0.8801	0.1164	0.0119	0.0009	$.6 \times 10^{-4}$	$.3 \times 10^{-5}$	$.1 \times 10^{-6}$		
observed	100.00	60.00	10.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Percent of groups with this factor's risk premia significant in natural order from factor analysis [d]												
Groups of 60 stocks	6.67	0.00	46.67	20.00	13.33	13.33	6.67	0.00	0.00	6.67	0.00	
Groups of 90 stocks	10.00	10.00	10.00	30.00	30.00	20.00	20.00	20.00	10.00	0.00	10.00	10.00

(a) Probability of observing at least this many significant regression coefficients in (3) given that risk premia vector is null.

(b) Number of factors are constrained to seven for each group at the first stage.

(c) Using t-tests, many of the regression coefficients are significant at 5% level in the GLS model (3).

(d) The factor with this number from the factor analysis has this percent significant t-ratios.

TABLE 5: Tests of Significance for Risk Premia Against Specific Alternatives Using Different Group Sizes (Number of groups with significant test statistics at 5% level)

Group Size	# of Factors (1)	Independent Variables: $B^{(a)}$ (2)	B (3)	$\sigma^{(b)}$ (4)	B (5)	$\Omega^{(c)}$ (6)
A. Number of Factors is Not Constrained						
		I. First Period: 7/3/62-3/23/72				
30	5	5 (17%)	0	5 (17%)	0	6 (20%)
60	8	4 (27%)	1 (7%)	5 (33%)	1 (7%)	5 (33%)
90	13	4 (40%)	1 (10%)	5 (50%)	1 (10%)	5 (50%)
		II. Second Period: 3/24/72-12/31/81				
30	5	1 (3%)	1 (3%)	3 (10%)	1 (3%)	3 (10%)
60	8	0	0	2 (13%)	1 (7%)	2 (13%)
90	13	0	0	2 (20%)	0	2 (20%)
		III. Entire Period: 7/3/62-12/31/81				
30	7	(13%)	0	5 (17%)	0	(17%)
60	11	3 (20%)	0	(40%)	0	(40%)
90	17	5 (50%)	0	(40%)	0	(40%)
B. Number of Factors Constrained to Seven						
30	7	4 (13%)	0	5 (17%)	0	5 (17%)
60	7	7 (47%)	0	9 (60%)	0	9 (60%)
90	7	9 (90%)	0	6 (60%)	2 (20%)	7 (70%)

(a,b,c) See Table 3 for explanations

TABLE 6: Joint Chi-Square Tests for the Intercepts (7/3/62-12/31/81)

Stocks Per Group	Null Hypotheses: $c_{t0}^{(i)} = 0$ (a) (1)	$c_{t0}^{(i)} = r_{f_t}$ (b) (2)	$c_{t0}^{(i)} = c_{t0}^{(1)}$ (c) (3)
	A. Using All Daily Returns		
30	54.523**	24.974	14.706
60	49.336**	21.292	11.609
90	46.809**	17.642*	5.819
	B. Using Returns on Every Thursday (d)		
30	43.304*	44.931**	43.258**
60	23.657*	26.720**	23.521*
90	14.261	17.657*	14.017

(a) We jointly test whether all intercepts are equal to zero.

(b) The intercept term is compared to the seventh root of the (one plus) weekly Treasury Bill yield observed every Thursday. It is assumed that this daily yield is constant for the next five trading days. The number of observations is 3280.

(c) The equality of intercepts is tested by subtracting the daily intercept for the first group from the rest of the groups and then testing whether the difference is equal to zero jointly.

(d) The intercept term on each Thursday is compared to the seventh root of the (one plus) weekly Treasury Bill observed every Thursday. Intercepts for Monday-Wednesday and Friday are deleted. The number of observations is 702.

NOTE: The degrees of freedom for the chi-square tests are 30, 15 and 10 for the groups of 30, 60 and 90 stocks in Columns (1) and (2). The corresponding numbers are 29, 14 and 9 for Column (3). * indicates that we reject the null hypothesis at the 10% level and ** indicates rejection at the 5% level.