The Seasonal Stability of the Factor Structure of Stock Returns
Author(s): D. Chinhyung Cho and William M. Taylor
Source: *The Journal of Finance*, Vol. 42, No. 5 (Dec., 1987), pp. 1195-1211
Published by: Blackwell Publishing for the American Finance Association
Stable URL: http://www.jstor.org/stable/2328522
Accessed: 26/11/2009 08:13

THE JOURNAL OF FINANCE • VOL. XLII, NO. 5 • DECEMBER 1987

The Seasonal Stability of the Factor Structure of Stock Returns

D. CHINHYUNG CHO and WILLIAM M. TAYLOR*

ABSTRACT

This paper investigates the month-by-month stability of (a) daily returns and correlation coefficients of stock returns, (b) correlation and covariance matrices, (c) number of return-generating factors, and (d) the APT pricing relationship. The results show that there is a January effect and a small-firm effect in stock returns. Correlation matrices are more stable than covariance matrices, but both types of matrices are not stable across months and across the sample groups. The number of return-generating factors is rather stable most of the time and for most of the sample groups, but there is some significant instability that is related to the average correlation coefficients among stocks. The APT pricing relationship does not seem to be supported by the two-stage process using the maximum-likelihood factor analysis.

THE SEASONALITY OF COMMON stock returns has received considerable attention in financial economics research. Studies by Officer [23], Rozeff and Kinney [29], Gultekin and Gultekin [11], and others have investigated the stock returns of the U.S. and other countries on a monthly basis. They generally have found that, on average, January returns are significantly higher than the returns for other months. Banz [1] and Reinganum [25, 26] have found abnormally high returns in small-market-value firms, and Keim [17] has related this to the first few trading days of January. Schwert [30] points out the current difficulty of explaining these empirical regularities and the need for further investigations.

Several studies have examined seasonality in risk premiums. Rozeff and Kinney [29] found that seasonality appeared in the risk premiums obtained from a two-parameter CAPM and that January displayed a relatively large risk premium compared with the other months. Keim [16] also found seasonality in the risk premiums. More recently, Tinic and West [32] examined the risk premiums of the two-parameter CAPM and observed the previously reported seasonality and high January premium. They also found that January is the only month in which there is a consistently positive, statistically significant relationship between expected return and risk. Very recently, using the two-stage procedure with the

* School of Business, University of Wisconsin, and School of Management, Boston University, respectively. We have benefited from the comments of the participants of the workshops at the Florida International University, Santa Clara University, University of Illinois, University of Michigan, University of Waterloo, and University of Wisconsin. An earlier version of this paper was presented at the University of Southern California Conference on Empirical Research on the APT and the 1986 Western Finance Association Meetings. We also wish to thank K. Jeon and J. Lee for their assistance. This project was started while W. M. Taylor was at the University of Wisconsin and completed while he was visiting Northwestern University.

maximum-likelihood factor analysis, Gultekin and Gultekin [12] show that the risk premia of the APT tend to be significant only in January.

In view of the above studies, we believe that it is worthwhile to investigate the factor structure of stock returns on a calendar-month basis. Studies by Kryzanowski and To [20], Dhrymes, Friend, and Gultekin [7], and Lehmann and Modest [21] show that factor structures are not invariant with respect to the sample size and the number of observations. However, these studies did not examine the factor structures on a calendar-month basis. This question of the stability of factor structures is also important because the previous empirical studies of APT that have been generated by the seminal theoretical paper by Ross [28] implicitly assume that factor structures are stationary across calendar months. These studies include Roll and Ross [27], Reinganum [24], Chen [4], Brown and Weinstein [3], Cho, Elton, and Gruber [6], Dhrymes, Friend, and Gultekin [7], Cho [5], Dhrymes, Friend, Gultekin, and Gultekin, [8], Gultekin and Gultekin [12], and Lehmann and Modest [21].

Hence, in this paper, we investigate the factor structure of daily common stock returns on a month-by-month basis to test the stability of (a) the correlation and covariance matrix, (b) the number of factors, and (c) the pricing of these factors. As in other studies, we use the maximum-likelihood factor analysis to estimate factor loadings and use a modified generalized least-squares regression analysis to test the APT pricing relationship. Because of the problem of factor rotations, we are not able to test whether the factors extracted are the same across months. Also, as indicated by Dhrymes, Friend, and Gultekin [7], we are not able to test the stability of individual factor prices from month to month. However, our paper should shed some light on the stability of factor structures, the applicability of maximum-likelihood factor analysis in testing the APT, and the APT pricing relationship. From this perspective, our study of factor-structure stability logically precedes the earlier studies.

The rest of the paper is organized as follows. Section I describes data and sample designs. Section II discusses the test methodology and test statistics. Section III presents the empirical results. Section IV concludes the paper.

I. Data and Sample Design

The data used in this study are daily stock returns on the New York or American Stock Exchanges obtained from the 1984 version of the CRSP Daily Returns File. We sample from those companies listed on the file from January 2, 1973, to December 30, 1983. The condition of being listed on the first and last trading days of this period and having not more than fifteen missing observations determined the eligible firms in our samples.

The samples were designed in the following manner. First, we obtained all companies on the COMPUSTAT Annual Industry Tape in each of six industry groups. These are the industries that contain sufficient numbers of companies for our analyses. These industry groups are foods, machineries, electronics, utilities, retail outlets, and financial institutions.[1] This industry-based sample design is prompted by previous studies that have indicated the possible importance of industry-related factors in explaining stock returns. An early factor-analysis study by King [18] has indicated the importance of an industry factor in explaining the covariability of stock returns.

Among a total of 577 companies that belonged to these industry groups, 459 companies were traded either on the New York or American Stock Exchanges. Among these 459 companies, 119 companies had to be eliminated because they either were not traded on January 2, 1973 or December 30, 1983, or had more than fifteen missing observations. We were left with 340 companies.

These 340 companies were then divided into five different market-value groups. Market values were obtained by multiplying the number of shares outstanding and the end-of-year closing share price as of December 1972 using the COMPUSTAT Annual Industry Tape. This gave us thirty different cross-classifications with respect to industry and market value. Table I summarizes the number of companies in each cell along with the market-value ranges. In addition, the table shows how many companies were selected from each cell to form our samples.

Three types of samples were formed by stratified random sampling of the companies in the cross-classification table. The three types are three random samples, six industry samples, and five market-value samples. Each of these fourteen samples contains thirty companies.

Table I

Cross-Classifications of the Number of Companies

	Market Values (Million Dollars)															
	1.4–36			36–122			122–281			282–688			694–18,214			Total
Industries	a	b	c	a	b	c	a	b	c	a	b	c	a	b	c	a
Foods	10	6	5	7	6	5	7	6	5	6	6	5	15	6	5	45
Machineries	23	7	5	11	7	5	7	6	5	7	6	5	4	4	4	52
Electronics	15	13	5	4	4	4	5	5	5	3	3	3	5	5	5	32
Utilities	5	5	5	24	6	6	35	7	5	34	6	6	22	6	6	120
Retails	9	6	5	11	7	5	5	5	5	6	6	5	9	6	5	40
Financials	6	6	5	11	6	5	9	6	5	12	6	6	13	6	5	51
Total	68			68			68			68			68			340

a: Total number of companies in each cell.
b: Number of companies selected from each market-value strata to form an industry sample.
c: Number of companies selected from each industry strata to form a market-value sample.

[1] The composition of these industry groups is available from the authors. Briefly, the SIC codes of each industry group are as follows.

foods: 2000,2010, 2020, 2030, 2041, 2048, 2050, 2065, 2070, 2082, 2085, 2086, 2099.
machineries: 3510, 3531, 3537, 3540, 3550, 3558, 3560, 3580.
electronics: 3570, 3573, 3600, 3670, 3674, 3679.
utilities: 4911, 4922, 4923, 4924, 4931.
retails: 5311, 5331, 5600, 5621, 5661, 5944, 5949, 5961, 5999.
financials: 6022, 6025, 6120, 6140, 6150, 6199, 6200.

The first type consists of three random samples obtained by randomly selecting one company without replacement from each of the thirty cells. Note that, since the electronics industry had only three companies in the market-value range of $282 to $688 million, we cannot generate more than three distinct random samples. We need to have distinct samples in order to compare the results across different samples in a given type. This approach randomizes the market values and industries so that any observed differences between the factor structures among random samples would not be attributable to the differences in market value or to differences in industry.

The second type consists of six industry samples that were constructed by obtaining one sample from each of the six industry groups. Each sample was constructed by randomly selecting about six companies without replacement from each of the five market-value strata. Note that the market-value distribution of companies in the electronics industry is more concentrated on small market values, and, hence, the electronics-industry sample contains more small-market-value firms. This approach randomizes market values so that any observed differences between the factor structures among industry samples would not be attributable to differences in market values.

The third type consists of five market-value samples that were constructed by obtaining one sample from each of the five market-value groups. Each sample was constructed by randomly selecting about five companies without replacement from each of the six industry strata. This approach randomizes industries so that any observed differences between the factor structures among market-value samples would not be attributable to differences in industries.

As we mentioned above, we used the daily return data from January 2, 1973, to December 30, 1983. In total, there were 2048 trading days. However, we did not use all of these trading days. Even though each company in our sample had, at most, fifteen missing observations, the missing dates did not coincide across companies. Hence, we decided to use only those days on which none of 340 companies had missing observations. This left us with 1718 observations.[2] The number of observations for each of the twelve months are, respectively, from January to December, 122, 128, 143, 132, 147, 135, 140, 157, 155, 152, 149, and 158 days.

II. Test Methodology

In this paper, we address the following four issues:

1. Do returns and correlation coefficients differ across calendar months and groups?
2. Do correlation matrices and covariance matrices differ across calendar months and groups?

[2] We refrain from transformations of the data including thin-trading adjustments such as those suggested by Shanken [31]. However, our elimination of days with missing data from all companies will mitigate the nonsynchronous trading problem. In this manner, any departures from factor-structure stability that we find can be readily related to the underlying assumptions and data treatment of previous studies rather than to differences in transformations or adjustments of the data.

3. Does the number of return-generating factors differ across calendar months and groups?
4. Do pricing relationships differ across calendar months and groups?

The first issue is addressed in order to verify the previous findings of seasonal patterns in stock returns. We are mainly interested in checking the consistency of our data with previous studies. Hence, we do not carry out any formal tests and just report numbers.

The second issue is addressed because the factor analysis is carried out on either the covariance matrix or the correlation matrix. First, we test whether January correlation matrices and covariance matrices are the same as those for other months within each group. Then, we test whether the matrices are the same for all twelve months within each group. The tests of equal correlation matrices and covariance matrices are carried out using Box's M-statistic [2]. (See Mardia, Kent, and Bibby [22].) If the null hypothesis of equal covariance matrices for q months is true, then Box's M-statistic, given by

$$M = \gamma \textstyle\sum_{i=1}^{q} (n_i - 1)\ln | S_i^{-1} S |,$$

where

$$\gamma = 1 - \frac{(2p^2 + 3p - 1)}{6(p + 1)(q - 1)} \left(\textstyle\sum_{i=1}^{q} \frac{1}{n_i - 1} - \frac{1}{N - q} \right)$$

and

$$S = \frac{1}{N - q} \textstyle\sum_{i=1}^{q} (n_i - 1) S_i,$$

has a χ^2 distribution with $p(p + 1)(q - 1)/2$ degrees of freedom. S_i is the ith-month covariance matrix, p is the number of variables in the covariance matrices, and N is equal to $\sum_{i=1}^{q} n_i$, where n_i is the number of observations for the ith month. S_i is replaced by a correlation matrix for the case of testing equal correlation matrices.

The third issue examines the number of factors for each calendar month and group. For these, we use the maximum-likelihood factor analyses with a significance level of 0.1. The maximum-likelihood factor analysis estimates the factor loadings B and unique variance Ψ by maximizing the likelihood function of $(n - 1)S$ or, equivalently, by minimizing

$$F_k(B, \Psi) = \ln | \Sigma | - \ln | S | + \mathrm{tr}(S \Sigma^{-1}) - p,$$

where $\Sigma = BB^t + \Psi$. S is the sample covariance matrix, which is an estimate of Σ obtained from a sample of n independent observations on p return variables, tr is the trace operator, and the superscript t represents the transpose matrix. Note that this function depends on the number of factors k. Also, note that B can be estimated uniquely only up to an orthogonal transformation because any orthogonally transformed matrix of B gives the same function value. The number of factors k is determined by examining the likelihood ratio, which is equal to $(n - 1)F_k(B, \Psi)$. If n is large, this likelihood ratio approaches a χ^2 distribution with $[(p - k)^2 - (p + k)]/2$ degrees of freedom. (See Jöreskog [14].)

Jöreskog and Sörbom [15] suggest the use of this χ^2-statistic as a goodness-of-fit measure rather than a test statistic in the sense that large χ^2 values correspond to bad fit and small values to good fit. Furthermore, they state that the χ^2 measure is sensitive to sample size and very sensitive to departures from multivariate normality of the observed variables. Large sample sizes and departures from normality tend to increase χ^2 and result in overestimating the number factors. Also, as documented by Dhrymes, Friend, and Gultekin [7] and Kryzanowski and To [20], the number of estimated factors tends to increase with an increase in the number of variables.

Once we estimate the number of factors for each calendar month and group, we test whether these numbers are stable across calendar months and across groups. Since we do not know the distributions of the number of factors, we use the rank method developed by Friedman [9]. This approach is similar to the analysis of variance except that, instead of using the original quantitative values, one uses the ranks that correspond to the sizes of these values. The major advantage of using ranks rather than the original values is that no assumptions about the distribution of the original variates are needed. Specifically, we need not be concerned with the normality assumptions of the analysis of variance.

Suppose we have a two-way table where the q columns represent the calendar months, the differences of which are to be tested, and the n rows represent the sample points, the elements of which are the ranks corresponding to the number of factors for a given calendar month. Each row represents different sample groups within the same type. The rank 1 is given to the smallest number of factors; the rank 2 is given to the next smallest number of factors, etc.[3] If the null hypothesis of an equal number of factors across q months were true, then the ranks entered for a particular calendar month should be random. In fact, the set of ranks in each column would represent a random sample of n items from the discrete rectangular distribution, the elements of which are 1 through q. The mean and variance of this distribution are $(q + 1)/2$ and $(q^2 - 1)/12$, respectively. This implies that the mean rank, $\bar{r}_j$, of the jth column will have a mean value $\rho = (q + 1)/2$ and variance $\sigma^2 = (q^2 - 1)/(12n)$. Hence, the null hypothesis of an equal number of factors across calendar months can be tested by calculating the following statistic:

$$\chi_r^2 = \frac{q-1}{q\sigma^2} \sum_{j=1}^{q} (\bar{r}_j - \rho)^2 = \frac{12n}{q(q+1)} \sum_{j=1}^{q} \left(\bar{r}_j - \frac{1}{2}(q+1)\right)^2.$$

The null hypothesis of an equal number of factors across groups can be tested as above by exchanging the roles of the rows and columns.[4]

[3] In some cases, two (or more) values in a row will be identical; i.e., there will be "tied" ranks. In these cases, one usually assigns the average rank to each of the tied observations (e.g., if two values are tied for the rank 2 and 3, then each value will be given the rank of 2.5). This modification does not affect the validity of χ_r^2.

[4] We should note that the Friedman procedure is very similar to the more well-known methodology developed by Kruskal and Wallis [19]. Both methodologies examine whether each of the q columns is generated from the same population. The difference is that the former treats the rows as fixed and views the problem as a matched-sample problem, whereas the latter assumes that the rows are independently drawn from the null distribution. Hence, the former assigns the ranks within each row, whereas the latter assigns the ranks over all rows.

For small values of q and n, exact distributions of the χ_r^2-statistics can be easily constructed. For $q \geq 3$ and $n \geq 6$, the distribution of the χ_r^2-statistic can be approximated by the χ^2 distribution with $q - 1$ degrees of freedom. For large q and small n, the χ_r^2 distribution can be approximated by the standard normal distribution as follows:

$$\frac{\chi_r^2 - (q - 1)}{2\,\dfrac{n - 1}{n}\,(q - 1)} \sim N(0, 1).$$

The last issue deals with the APT pricing relationships. For these, we follow the approach used by Roll and Ross [27]. More specifically, we use a modified generalized least-squares regression analysis on the following model.

$$\bar{R}_i = \lambda_0 + \lambda_1 b_{1i} + \lambda_2 b_{2i} + \cdots + \lambda_k b_{ki} + \varepsilon_i; \quad i = 1, 2, \cdots, p,$$

or in the matrix form

$$\bar{R} = B^* \Lambda + \varepsilon,$$

where $\bar{R}_i$ is the mean return for the ith stock for a particular month, λ_j is the risk premium for the jth factor, b_{ji} is the jth factor loading of the ith stock, $\bar{R} = (\bar{R}_1, \bar{R}_2, \cdots, \bar{R}_p)^t$, $\Lambda = (\lambda_0, \lambda_1, \cdots, \lambda_k)^t$, $B^* = (e, B)$, $e = (1, 1, \cdots, 1)^t$, B is the $p \times k$ factor-loading matrix, and $\varepsilon = (\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_p)^t$ are disturbance terms. The estimates of the regression coefficients are

$$\hat{\Lambda} = (\hat{B}^{*t}\hat{\Sigma}^{-1}\hat{B}^*)^{-1}\hat{B}^{*t}\hat{\Sigma}^{-1}\bar{R},$$

and the covariance matrix of these estimates is

$$V = \frac{1}{n}\,(\hat{B}^{*t}\hat{\Sigma}^{-1}\hat{B}^*)^{-1},$$

where $\hat{\Sigma} = \hat{B}\hat{B}^t + \hat{\Psi}$, $\hat{B}^* = (e, \hat{B})$, and $\hat{B}$ is the estimate of factor loadings.

In our study, the factor loadings used in the above regression model are estimated for each month using 0.1 significance level. Studies by Tinic and West [32] and Gultekin and Gultekin [12] estimate these systematic risks, or factor loadings, using the entire period. Hence, if the January expected returns are higher than those of any other months and if the standard errors of (risk premia) estimates are rather stable across months, then cross-sectional studies should yield more significant risk premia in January than in any other months. This may be the reason why they find a significant pricing relationship only in January.

In order to see which factors are priced, one would normally examine t-statistics for each coefficient λ_j. However, as pointed out by Dhrymes, Friend, and Gultekin [7], t-statistics are not invariant under orthogonal transformations and, hence, cannot be used in this situation. Recall that the factor loadings can be estimated uniquely only up to an orthogonal transformation, and, hence, the estimated risk premia are, at best, some linear combinations of true risk premia. Thus, t-statistics do not tell us which factors are priced. All that can be tested in this situation are whether the intercept term is significantly different from zero and whether the risk premia as a whole are significantly different from zero. The latter is possible because the sum of squares is invariant under orthogonal transformations. In this paper, we use Hotelling T^2-statistics to test whether the risk premia as a whole are significant at 0.05 significance level.[5] Hotelling T^2-statistic is obtained by

$$T^2 = \hat{\Lambda}^{*t} V^{*-1} \hat{\Lambda}^*,$$

and its significance can be tested using the following F-statistic:

$$F = \frac{n-k}{k(n-1)} T^2,$$

which has an F-distribution with k and $n - k$ degrees of freedom. $\hat{\Lambda}^*$ is a vector of $\hat{\lambda}_j$'s excluding $\hat{\lambda}_0$, and V^* is a submatrix of V excluding the first row and the first column of V. This Hotelling T^2-statistic is equivalent to the χ^2-statistic used by Dhrymes, Friend, Gultekin, and Gultekin [8].

III. Empirical Results

A. Returns and Correlation Coefficients

As a first step to investigating the seasonality of stock returns, we examine the returns and correlation coefficients of our sample groups on a calendar-month basis. Table II reports average daily percentage returns and average correlation coefficients for each of twelve months and for the entire twelve months under the heading of "Year". For each group, average returns are obtained by averaging the mean returns of thirty companies, and average correlation coefficients are obtained by averaging $30 \times 29/2 = 435$ pairwise correlation coefficients within a group.

In regard to returns, we first note some fluctuations across different months. For each group, the largest average return occurs in January, and January average returns are about four times more than average returns for the entire period. Also, we note that the smallest market-value group M1 has the largest January average return, and average returns tend to decrease as we move from M1 to the largest market-value group M5 for each column. We further note that three random groups, R1 through R3, and six industry groups, I1 through I6, have no discernible pattern except that January average returns are larger than those of other months. Hence, these results support the previous findings of the small-firm effect and the January effect.[6]

In regard to average correlation coefficients, there are some fluctuations across different months. January average correlation coefficients are higher than average correlation coefficients for the entire period, but the largest average correlation coefficients occur either in October or November for twelve out of our fourteen sample groups. Also, we note that average correlation coefficients for M1 are lower than average correlation coefficients for M5. In fact, average correlation coefficients tend to increase as we move from M1 to M5 for each

[5] We do not test the hypothesis of zero intercept because the appropriate null hypothesis should be whether the intercept equals the risk-free rate.

[6] As mentioned before, we do not carry out any rigorous tests on returns. For example, we do not consider risks.

Table II

Average Daily Percentage Returns and Average Correlation Coefficients[a]

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Year
					Panel A: Average Daily Returns								
R1	0.297*	0.023	0.093*	0.072*	0.090*	0.103*	0.063*	0.030	−0.055*	−0.042	0.125*	−0.025	0.062*
R2	0.321*	0.094*	0.062*	0.088*	0.093*	0.102*	0.037	−0.042	0.021	−0.012	0.096*	0.017	0.070*
R3	0.242*	−0.038*	0.109*	0.073*	0.076*	0.119*	0.041	−0.005	0.010	−0.058*	0.106*	0.018	0.056*
I1	0.250*	0.028	0.040	0.068*	0.050*	0.156*	0.045	−0.023	0.037*	−0.015	0.075	0.089*	0.065*
I2	0.387*	0.005	0.108*	0.050	0.136*	0.119*	0.017	−0.038	−0.058*	−0.095*	0.097*	0.097*	0.066*
I3	0.447*	0.061	0.092*	0.158*	0.158*	0.180*	−0.003	0.016	−0.020	−0.066*	0.137*	0.021	0.095*
I4	0.189*	−0.049*	−0.009	0.028*	0.057*	0.132*	0.033	0.062*	0.061*	0.037*	0.076*	−0.016	0.046*
I5	0.267*	0.031	0.253*	0.105*	0.056*	0.117*	0.038*	0.004	0.045*	−0.077*	0.080	−0.088*	0.066*
I6	0.347*	0.005	0.023	0.120*	0.028	0.136*	0.024	0.091*	0.012	−0.006	0.143*	−0.024	0.065*
M1	0.462*	0.115*	0.086*	0.083*	0.139*	0.178*	0.107*	−0.034	−0.021	−0.056	0.145*	0.004	0.096*
M2	0.366*	−0.003	0.069*	0.128*	0.044	0.128*	0.030	−0.029	0.009	0.046	0.113*	−0.047	0.060*
M3	0.275*	0.003	0.062*	0.047*	0.095*	0.136*	0.004	0.004	0.029	0.000	0.087*	0.020	0.063*
M4	0.205*	−0.009	0.103*	0.067*	0.080*	0.108*	−0.005	−0.036	0.026	−0.032	0.046*	0.035	0.048*
M5	0.157*	−0.015	0.064*	0.091*	0.048*	0.089*	−0.006	0.000	−0.052*	0.068*	0.069*	−0.012	0.041*

[a] 1. For each sample, Panel A represents daily percentage returns averaged over thirty companies in each group and Panel B represents the average of $30 \times 29/2 = 435$ pairwise correlation coefficients in the correlation matrix.
2. "Year" represents the entire twelve months.
3. R1, R2, R3: random samples
I1: foods-industry sample
I2: machineries-industry sample
I3: electronics-industry sample
I4: utilities-industry sample
I5: retail-industry sample
I6: financial-industry sample
M1: $1.4–36 million market-value sample
M2: $36–122 million market-value sample
M3: $122–281 million market-value sample
M4: $282–688 million market-value sample
M5: $694–18,214 million market-value sample

* Significant at 0.05 significance level. All average correlation coefficients are significant at 0.05 significance level.

Table II—Continued

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Year
					Panel B: Average Correlation Coefficients								
R1	0.133	0.070	0.117	0.086	0.094	0.079	0.100	0.137	0.137	0.163	0.140	0.104	0.115
R2	0.182	0.085	0.132	0.101	0.117	0.110	0.102	0.133	0.156	0.168	0.163	0.128	0.134
R3	0.147	0.083	0.122	0.086	0.094	0.092	0.111	0.124	0.134	0.177	0.167	0.121	0.124
I1	0.135	0.085	0.121	0.093	0.075	0.074	0.084	0.079	0.139	0.160	0.111	0.109	0.106
I2	0.160	0.106	0.138	0.104	0.119	0.092	0.118	0.133	0.156	0.178	0.184	0.139	0.138
I3	0.173	0.131	0.178	0.149	0.141	0.118	0.137	0.158	0.167	0.204	0.200	0.134	0.159
I4	0.159	0.069	0.080	0.106	0.075	0.112	0.061	0.122	0.116	0.157	0.105	0.114	0.110
I5	0.161	0.095	0.154	0.107	0.103	0.114	0.131	0.163	0.168	0.172	0.186	0.138	0.144
I6	0.176	0.122	0.141	0.136	0.142	0.131	0.148	0.157	0.164	0.209	0.201	0.168	0.161
M1	0.095	0.056	0.084	0.043	0.063	0.053	0.043	0.074	0.085	0.121	0.094	0.063	0.074
M2	0.116	0.056	0.092	0.069	0.067	0.084	0.063	0.083	0.105	0.125	0.105	0.096	0.091
M3	0.183	0.114	0.144	0.109	0.132	0.099	0.107	0.154	0.169	0.219	0.185	0.149	0.151
M4	0.175	0.109	0.145	0.121	0.131	0.119	0.134	0.171	0.181	0.215	0.192	0.164	0.157
M5	0.240	0.177	0.220	0.200	0.198	0.196	0.216	0.281	0.286	0.290	0.303	0.270	0.243

calendar month. It is interesting that average correlation coefficients for industry groups do not seem to be higher than average correlation coefficients for random groups, whereas M5 has the highest average correlation coefficients and M1 has the smallest average correlation coefficients among all sample groups for each column. These results seem to indicate that the degree to which firms move together depends on market values regardless of whether they are in the same industry (as we have defined them).[7] Why do large firms tend to move together? We do not have a good answer. However, we think that this is a very important finding and deserves further research attention.

B. Correlation Matrices and Covariance Matrices

We test the following hypotheses.

H1: The January correlation (covariance) matrix is the same as the correlation (covariance) matrix of other months.
H2: Correlation (covariance) matrices are the same for all twelve months.

These hypotheses are tested for each group using 0.05 significance levels.

Our results show that hypothesis H2 and the covariance-matrices part of H1 should be rejected for all cases.[8] In regard to the correlation-matrices part of H1, the hypothesis is rejected in 122 of the 154 cases. Table III reports the p-levels associated with Box's M-statistics. For groups of R1, R2, I3, I4, M4, and M5, January correlation matrices are significantly different from the correlation matrices of all other months. All the groups, except I1 and M3, show that January is different from at least six other months.[9]

In summary, our study shows that covariance matrices are not stable across calendar months. Correlation matrices show a little more stability than covariance matrices, but these are also not stable across calendar months.[10] Since factor analyses are carried out on either covariance matrices or correlation matrices, one may wonder whether the instability that has just been observed in these matrices may lead to instability in the estimated number of return-generating factors. This question is addressed next.

C. Number of Return-Generating Factors

The number of return-generating factors is estimated by the maximum-likelihood factor analysis using 0.1 significance level. Table IV reports the number of factors obtained for each group by calendar month and for the entire period.[11]

[7] This result for market-value groups is similar to that found by Huberman and Kandel [13] in a study that focused on a size-based-return model.

[8] The statistics discussed in this paragraph are not reported but are available upon request.

[9] We also examined whether correlation (covariance) matrices are the same between two groups within a given type. Note that we cannot test this hypothesis using all fourteen groups because the sample groups are independently constructed only within a given type. Our results show that covariance matrices between two groups are different in all cases and that correlation matrices between two groups are different in most of the cases.

[10] Gibbons [10] also reported that correlation matrices exhibit more stability than covariance matrices.

[11] In the following, Heywood cases are eliminated so that the reported factor structures are not affected.

Table III

Equal Correlation Matrices between January and Other Months[a]

	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
R1	0.001*	0.000*	0.000*	0.001*	0.046*	0.001*	0.007*	0.000*	0.000*	0.000*	0.001*
R2	0.009*	0.006*	0.019*	0.016*	0.008*	0.002*	0.006*	0.001*	0.013*	0.000*	0.000*
R3	0.032*	0.012*	0.080	0.001*	0.077	0.120	0.000*	0.024*	0.008*	0.004*	0.001*
I1	0.185	0.016*	0.060	0.156	0.049*	0.377	0.098	0.043*	0.055	0.004*	0.001*
I2	0.001*	0.001*	0.169	0.009*	0.405	0.055	0.255	0.001*	0.001*	0.001*	0.048*
I3	0.000*	0.000*	0.057*	0.001*	0.010*	0.004*	0.004*	0.000*	0.000*	0.000*	0.000*
I4	0.000*	0.001*	0.031*	0.000*	0.002*	0.002*	0.004*	0.000*	0.001*	0.035*	0.005*
I5	0.014*	0.001*	0.001*	0.074*	0.225	0.023*	0.052	0.002*	0.000*	0.195	0.000*
I6	0.150	0.001*	0.302	0.023*	0.017*	0.280	0.010*	0.000*	0.002*	0.009*	0.010*
M1	0.001*	0.061	0.251	0.001*	0.016*	0.001*	0.019*	0.000*	0.001*	0.000*	0.000*
M2	0.269	0.040*	0.702	0.001*	0.217	0.104	0.224	0.006*	0.024*	0.040*	0.000*
M3	0.049*	0.000*	0.035*	0.151	0.199	0.137	0.303	0.146	0.110	0.066	0.005*
M4	0.001*	0.000*	0.001*	0.001*	0.000*	0.000*	0.012*	0.002*	0.000*	0.000*	0.004*
M5	0.000*	0.015*	0.004*	0.002*	0.004*	0.000*	0.000*	0.013*	0.000*	0.001*	0.000*

[a] Numbers reported in this table are p-levels corresponding to Box's M-statistic.
* Significant at 0.05 significance level.

Table IV

Number of Return-Generating Factors[a]

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Year
R1	5	2	4	4	2	2	1	1	4	3	5	3	6
R2	5	1	3	3	2	2	4	3	5	4	3	5	7
R3	6	3	4	1	3	2	1	2	4	5	4	4	6
I1	3	1	3	2	1	1	2	1	1	2	6	4	4
I2	4	3	2	1	2	1	1	1	3	4	4	1	6
I3	6	5	6	2	4	5	4	5	5	6	4	7	8
I4	2	3	3	2	5	5	1	3	4	6	3	4	7
I5	4	3	4	2	3	3	3	4	7	6	2	8	7
I6	4	2	5	1	3	3	3	4	7	7	5	7	9
M1	4	4	2	1	4	3	3	1	4	5	5	3	5
M2	3	1	2	2	2	2	1	2	4	2	3	7	4
M3	2	2	4	1	3	2	2	5	5	3	2	2	6
M4	6	3	2	2	2	6	5	4	4	3	5	3	7
M5	5	9	6	2	4	5	7	7	6	6	4	6	10

[a] 1. Number of factors is estimated by the maximum-likelihood factor analysis using 0.1 significance level. The subroutine OFCOMM in the IMSL Library is used for factor analysis.
2. "Year" represents the entire twelve months.

We first note that, with few exceptions, for each group the number of factors for the entire period is more than the numbers of factors for calendar months. They range between four and ten factors, with a mode between six and seven factors. These results are consistent with previous findings that show that the number of factors tends to increase as the number of observations increases and that there are about five to seven factors for thirty stock groups.

However, Table IV suggests wide fluctuations in the number of factors across calendar months and groups. In order to investigate these fluctuations more formally, we test for an equal number of factors for the entire twelve months and between each pair of months and, also, for all groups and between each pair of groups within a given type of sample group.

First, we use the Friedman rank method, with groups in each type as sample points to test for fluctuations across months. The hypothesis of an equal number of factors across all months is rejected for the random groups and industry groups but not for the market-value groups. The p-levels are 0.023 for random groups, 0.002 for industry groups, and 0.259 for market-value groups.

p-Levels for the tests of equal number of factors between each pair of calendar months also support these findings. For each type of sample groups, sixty-six $\left(\binom{12}{2} = 66\right)$ pairs of calendar months are compared.[12] The results show that there are twenty-eight significant pairs for random groups, fourteen significant pairs for industry groups, and eight significant pairs for market-value groups. However, these significant results primarily originate from pairs that include-January, June, or September for random groups, April, July, or October for industry groups, and April or September for market-value groups. Examination of Table IV reveals that there are more factors in January and September for random groups, in October for industry groups, and in September for market-value groups. There are fewer factors in June for random groups, in April and July for industry groups, and in April for market-value groups. Otherwise, the number of factors seems to be rather stable across the remaining calendar months.

Next, Friedman's rank method is used, with the twelve months as sample points to test for fluctuations across groups within a given type. The hypothesis that all groups within a given type have an equal number of factors is rejected for the industry groups and the market-value groups but not for the random groups. The p-levels are 0.000 for the industry groups, 0.001 for the market-value groups, and 0.527 for the random groups.

p-Levels for the tests of equal number of factors between each pair of groups also support these findings. For random groups, the three pairwise comparisons show no significant differences. For industry groups, seven out of fifteen pairwise comparisons show significant results. These significant results originate from three groups, I1, I2, and I3. I3 has more factors, while I1 and I2 have fewer factors than other groups. For market-value groups, four out of ten pairwise comparisons show significant results. These significant results are mainly due to one group, M5, which has more factors than other groups.

Finally, a comparison of Tables IV and II suggests a positive relationship between the number of factors and the average correlation coefficient. For example, M5 has more factors and higher average correlation coefficients than M1. Also, all of our groups tend to have fewer factors and lower average correlation coefficients in midyear than at either the beginning or the end of the year. Kendall's τ_B rank correlation coefficient is used to test for a positive relationship between the number of factors in Table IV and the average correlation coefficients in Table II. The rank correlation coefficients are 0.462 for random groups, 0.400 for industry groups, and 0.398 for market-value groups. In all cases, the associated p-levels are 0.000. Hence, there is a significant positive

[12] The statistics discussed in this paragraph are not reported but are available upon request.

relationship between the number of factors and the average correlation coefficient among stocks. Note that, from a theoretical point of view, there is no reason why we should observe this positive relationship. For example, it is possible to explain large correlations by large factor loadings instead of a large number of factors.

In summary, the number of factors differs across months in two or three months (differing according to the type of group) and across some groups. This is contrary to our results in Section III, Subsection B, for correlation matrices. Recall that correlation matrices differ more often than not across months and across groups. This indicates that there is no relationship between the stability of correlation matrices and the stability of numbers of factors. However, there is a relationship between the number of factors and the average correlation coefficients among stocks. Again, we stress the need to understand why certain firms move together, especially in particular months. In the absence of a good explanation, we must incorporate the instability of the factor structure in testing the APT.

D. APT Pricing Relationships

The APT pricing relationship is investigated by running a modified generalized least-squares regression analysis of mean returns on factor loadings.[13] While there may be more powerful methods for investigating the APT pricing relationship, we use this method to permit easy comparison with previous studies. A 0.05 significance level is used for testing purposes. The analysis is carried out for each group by calendar month and for the entire period. We emphasize that our factor loadings are estimated on a monthly basis. Hence, each (monthly) cross-sectional regression has different sets of factor loadings rather than the same set of factor loadings as in Gultekin and Gultekin [12]. Table V reports p-levels associated with Hotelling T^2-statistics for the null hypothesis that the risk premia as a whole are not different from zero.

We first note that the APT does not seem to hold for the entire period. There are no groups that show any significant statistics. This result is similar to the findings of Dhrymes, Friend, Gultekin, and Gultekin [8]. When the analysis is repeated for each month, there are only nineteen out of 168 cases that show significant Hotelling T^2-statistics. This result, again, is not in support of the APT.

However, January is a little different from other months. For each type of group, there are two cases that show that at least one factor is priced. Our results are different from the results obtained by Gultekin and Gultekin [12]. They test the APT on a monthly basis using the same set of factor loadings that are obtained from the seven-factor maximum-likelihood factor analysis on the entire period. They find that these factors are priced for all groups in January and are rarely priced in other months. Hence, they conclude that the factor-analysis

[13] The use of noncontiguous days for estimation of parameters and testing of the APT may have reduced the biases in the second-stage estimates. However, as pointed out by Gultekin and Gultekin [11], some sources of biases still remain. In any case, the primary focus of our paper is on the factor structures rather than APT pricing relationships. Consequently, we decided not to carry out analyses, for example, on an even-odd-days basis. By so doing, our results are biased in favor of finding an APT pricing relationship.

Table V

Pricing of Factors[a]

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Year
R1	0.034*	0.462	0.074	0.728	0.816	0.849	0.990	0.998	0.680	0.249	0.831	0.921	0.968
R2	0.014*	0.969	0.032*	0.928	0.232	0.265	0.957	0.504	0.248	0.321	0.411	0.741	0.275
R3	0.248	0.571	0.167	0.504	0.716	0.243	0.575	0.699	0.997	0.081	0.571	0.388	0.673
I1	0.526	0.503	0.425	0.020*	0.001*	0.001*	0.498	0.018*	0.798	0.040*	0.688	0.010*	0.412
I2	0.000*	0.703	0.821	0.479	0.004*	0.040*	0.867	0.933	0.992	0.162	0.502	0.752	0.100
I3	0.136	0.768	0.977	0.185	0.497	0.004*	0.978	0.926	0.755	0.630	0.993	0.995	0.598
I4	0.253	0.591	0.988	0.604	0.373	0.886	0.151	0.460	0.532	0.644	0.925	0.391	0.694
I5	0.326	0.947	0.317	0.683	0.698	0.639	0.570	0.073	0.950	0.927	0.709	0.995	0.151
I6	0.000*	0.659	0.868	0.887	0.571	0.505	0.653	0.964	0.983	0.178	0.606	0.569	0.630
M1	0.015*	0.591	0.440	0.415	0.019*	0.393	0.251	0.814	0.349	0.644	0.565	0.356	0.092
M2	0.048*	0.403	0.045*	0.359	0.274	0.897	0.623	0.848	0.922	0.457	0.380	0.163	0.225
M3	0.202	0.212	0.569	0.644	0.162	0.067	0.915	0.757	0.809	0.778	0.063	0.934	0.216
M4	0.533	0.041*	0.286	0.618	0.779	0.325	0.863	0.135	0.726	0.876	0.871	0.167	0.554
M5	0.203	0.717	0.159	0.896	0.494	0.962	0.668	0.554	0.486	0.333	0.745	0.150	0.439

[a] 1. Numbers reported in this table are p-levels corresponding to the Hotelling T^2-statistic.
2. "Year" represents the entire twelve months.

* Significant at 0.05 significance level.

approach would imply that the APT is valid only in January. As mentioned before in Section II, their procedure might have been biased in favor of finding the significant relationship in January. On the other hand, our results are based on the factor loadings that are extracted for a given month. Still, we find that January shows different pricing behavior than other months.

Can it then be concluded that the APT pricing relationship holds only in January? Perhaps, but one should be cautious before jumping to this conclusion. Note that the methodology commonly used in testing the APT is rather weak. We know that the maximum-likelihood factor analysis tends to overestimate the number of factors in the first stage. Hence, there is no reason why all of these factors should be priced in the second stage. Unfortunately, due to the nonuniqueness of factor loadings, one cannot test which subset of these factors is priced. One can only test whether the risk premia as a whole are equal to zero. However, as more insignificant factor loadings are included in the regression analysis, the value of Hotelling T^2-statistic decreases. This problem may be enhanced when the estimated factor loadings are positively related to each other. Hence, one tends to reject the APT as more insignificant factor loadings are included. Therefore, it seems that further study is needed before one can conclude that the APT pricing relationship holds only in January.

IV. Conclusion

In this paper, we carried out month-by-month analyses on (a) returns and correlation coefficients of stock returns, (b) correlation and covariance matrices, (c) the number of return-generating factors, and (d) the APT pricing relationship. We employed the commonly used maximum-likelihood factor analysis and the modified generalized-regression analysis to examine their applicability in testing APT on a monthly basis. Our results show that there is a January effect and a small-firm effect on stock returns. Correlation matrices are more stable than covariance matrices, but both correlation matrices and covariance matrices are not stable across months. The number of return-generating factors is rather stable most of the time and for most of the groups. However, there is some significant instability among some groups and in two or three months. Also, the number of return-generating factors has a positive relationship with the average correlation coefficient among stocks. Finally, the APT pricing relationship does not seem to be supported by our testing methodology, and January seems to have a somewhat different pricing behavior than other months.

To the extent that the factor-structure instability is significant, tests of the APT should employ methods that accommodate the observed patterns. We leave the application of alternative methods for future research.

REFERENCES

1. R. Banz. "The Relationship between Return and Market Value of Common Stocks." *Journal of Financial Economics* 9 (March 1981), 3–18.
2. G. Box. "A General Distribution Theory for a Class of Likelihood Criteria." *Biometrika* 36 (December 1949), 317–46.
3. S. Brown and M. Weinstein. "A New Approach to Testing Asset Pricing Models: The Bilinear Paradigm." *Journal of Finance* 38 (June 1983), 711–43.

4. N. Chen. "Some Empirical Tests of the Theory of Arbitrage Pricing." *Journal of Finance* 38 (December 1983), 1393–1414.
5. D. Cho. "On Testing the Arbitrage Pricing Theory: Inter-Battery Factor Analysis." *Journal of Finance* 39 (December 1984), 1485–1502.
6. ———, E. Elton, and M. Gruber. "On the Robustness of the Roll and Ross APT Methodology." *Journal of Finance and Quantitative Analysis* 19 (March 1984), 1–10.
7. P. Dhrymes, I. Friend, and N. Gultekin. "A Critical Reexamination of the Empirical Evidence on the Arbitrage Pricing Theory." *Journal of Finance* 39 (June 1984), 323–46.
8. P. Dhrymes, I. Friend, M. Gultekin, and N. Gultekin. "New Tests of the APT and Their Implications." *Journal of Finance* 40 (July 1985), 659–74.
9. M. Friedman. "The Use of Ranks to Avoid the Assumption of Normality Implicit in the Analysis of Variance." *Journal of the American Statistical Association* 32 (December 1937), 675–701.
10. M. Gibbons. "An Empirical Examination of the Return Generating Process of the Arbitrage Pricing Theory." Working Paper, Stanford University, 1984.
11. M. Gultekin and N. Gultekin. "Stock Market Seasonality: International Evidence." *Journal of Financial Economics* 12 (December 1983), 469–81.
12. ———. "Stock Return Anomalies and the Tests of the APT." *Journal of Finance* 42 (December 1987).
13. G. Huberman and S. Kandel. "A Size Based Stock Returns Model." Working Paper, University of Chicago, 1985.
14. K. Jöreskog. "Some Contributions to Maximum Likelihood Factor Analysis." *Psychometrika* 32 (December 1967), 443–82.
15. ——— and D. Sörbom. *LISREL V.* Chicago: National Educational Resources, Inc., 1983.
16. D. Keim. "Asset Pricing Anomalies and Capital Market Seasonality: Empirical Evidence." Unpublished manuscript, University of Chicago, 1980.
17. ———. "Size-Related Anomalies and Stock Return Seasonality: Further Empirical Evidence." *Journal of Financial Economics* 12 (June 1983), 13–32.
18. B. King. "Market and Industry Factors in Stock Price Behavior." *Journal of Business* 39 (January 1966), 139–90.
19. W. Kruskal and W. Wallis. "Use of Ranks in One-Criterion Variance Analysis." *Journal of the American Statistical Association* 47 (December 1952), 583–621.
20. L. Kryzanowski and M. To. "General Factor Models and the Structure of Security Returns." *Journal of Financial and Quantitative Analysis* 18 (March 1983), 31–52.
21. B. Lehmann and D. Modest. "The Empirical Foundations of the Arbitrage Pricing Theory I: The Empirical Tests." Working Paper No. 1725, National Bureau of Economic Research, 1985.
22. K. Mardia, J. Kent, and J. Bibby. *Multivariate Analysis.* New York: Academic Press, 1979.
23. R. Officer. "Seasonality in Australian Capital Markets: Market Efficiency and Empirical Issues." *Journal of Financial Economics* 2 (March 1975), 29–52.
24. M. Reinganum. "The Arbitrage Pricing Theory: Some Empirical Results." *Journal of Finance* 36 (May 1981), 313–21.
25. ———. "Misspecification of Capital Asset Pricing: Empirical Anomalies Based on Earnings Yields and Market Values." *Journal of Financial Economics* 9 (March 1981), 19–46.
26. ———. "The Anomalous Stock Market Behavior of Small Firms in January: Empirical Tests for Tax-Loss Selling Effects." *Journal of Financial Economics* 12 (June 1983), 89–104.
27. R. Roll and S. Ross. "An Empirical Investigation of the Arbitrage Pricing Theory." *Journal of Finance* 35 (December 1980), 1073–1103.
28. S. Ross. "The Arbitrage Theory of Capital Asset Pricing." *Journal of Economic Theory* 13 (December 1976), 341–60.
29. M. Rozeff and W. Kinney, Jr. "Capital Market Seasonality: The Case of Stock Returns." *Journal of Financial Economics* 3 (October 1976), 379–402.
30. G. Schwert. "Size and Stock Returns, and Other Empirical Regularities." *Journal of Financial Economics* 12 (June 1983), 3–12.
31. J. Shanken. "Nonsynchronous Data and the Covariance-Factor Structure of Returns." *Journal of Finance* 42 (June 1987), 221–31.
32. S. Tinic and R. West. "Risk and Return January vs. the Rest of the Year." *Journal of Financial Economics* 13 (December 1984), 561–74.