The Asset Pricing Mechanism in Japan's Stock Market: A New Test of Arbitrage Pricing Theory*

CHIHIRO SAKURABA**

I. Introduction

The main objective of this paper is to analyze empirically the asset pricing mechanism in Japan's stock market by use of Arbitrage Pricing Theory (APT). The paper will propose a new method of testing APT and enhance the power of the test by using large size panel data.

APT was initially proposed by Ross (1976) as an alternative to the Capital Asset Pricing Model (CAPM), the explanatory and predictive power of which had been increasingly questioned not only in the United States but also in Japan. The major difference between APT and CAPM lies in the different asset return generating processes they postulate: while APT allows more than one common factors to explain the variances of individual asset returns, CAPM allows only the expected market return to influence the expected returns on individual assets. This means that, as the return generating process assumed in APT is less restrictive, APT can more readily be made to fit the observed price behavior.

APT, however, has a fundamental indeterminacy problem for the factor structure of the entire asset market (Shanken 1982). In order to overcome this problem, our methodology will restrict the set of assets to the range of stocks whose returns are

* A number of people provided many valuable comments on various drafts of the paper. In particular, thanks are due to Mamoru Aoyama of Yokohama National University, Saburo Horimoto of the University of Shiga, Susumu Saito of the University of Ibaraki, Shinji Takagi of the Institute for Monetary and Economic Studies, Bank of Japan, and Yoshiyuki Takeuchi of the University of Tokyo, a Visiting Student at the Institute. The author also thanks Fumiko Konya of the Japan Securities Research Institute for kindly supplying the panel data of Japanese stock returns. However, the author is responsible for any remaining errors.
** Economist, Institute for Monetary and Economic Studies, Bank of Japan.

subject to the approximately same type of linear multi-factor structure. This restriction would make it unnecessary to consider the entire equilibrium of asset markets and to know the "true" market portfolio that may not be observable (Roll 1977).

In general, APT has not been successful in explaining the exact nature of stock pricing in the past. For example, Horimoto (1985), the only major empirical study of APT using Japanese stock market data, rejected APT as a satisfactory model for individual stock pricing. However, no conclusive evidence against APT can be said to exist because of methodological inadequacies in previous studies. In fact, one can show that, whenever CAPM is rejected, multivariate normality can not be assumed for asset returns. Thus, an ideal test of APT should use least squares estimators for factor loadings and not maximum likelihood estimators. Moreover, the choice of the number of common factors in the model becomes important. It turns out that the new testing procedure proposed in the paper is much more flexible than the more common maximum likelihood method.

In order to enhance the power of the new testing procedure, the paper will use least squares factor analysis to directly calculate the sample covariance matrices of asset returns. By not using the usual portfolio grouping procedure, we can maintain the size of matrices large enough to do meaningful hypothesis testing of APT. These methodological improvements will make it possible to successfully apply APT to the stock returns in Japan. In fact, our results will show that the linear multi-factor return generating process describes the stock pricing mechanism in the Japanese stock market fairly well.

The paper is organized as follows. Section II presents a brief comparison between CAPM and APT as well as the basic conditions required for testing Ross (1976)'s APT theorem. Section III discusses the methodological problems of previous studies, and proposes least squares factor analysis as a new computational procedure for testing APT; it will also discuss information criteria for choosing the number of common factors in the model. Section IV presents empirical results. Section V compares our results with those of other previous studies of APT. Section VI presents a summary and concluding remarks. Finally, the Appendix contains a brief explanation of factor analysis.

II. Modern Portfolio Theory and APT

1. CAPM and Its Explanatory Power

Analysis of the pricing mechanism in the stock market requires a model of portfolio selection and investment behavior under uncertainty. Such a model, however, is likely to be unnecessarily complicated for empirical application, because it requires a state-preference framework to treat incomplete markets and asymmetric information. Thus, modern asset pricing theory has generally adopted a simpler approach which focuses on the trade-off between risk and return.[1] During most of the 1970s, CAPM dominated the financial theory because it seemed capable of explaining the real world well.

At the outset, it may be useful to review CAPM for the purpose of providing background information on APT. CAPM, which is based on the notion of "mean-variance" efficiency, imposes the following restrictions on the structure of preferences or asset returns (Ross 1978).

(a) Investors' risk tolerance (that is, a reciprocal of absolute risk aversion) is a linear function of wealth.[2] A quadratic utility function, for example, would satisfy this condition.

(b) The distribution of returns on individual assets, $\tilde{r}_i$, is a linear combination of the riskless interest rate, r_f, and the rate of return on a risky fund, $\tilde{z}$. This condition can be symbolically stated as

$$\tilde{r}_i = (1-\beta_i)r_f+\beta_i\tilde{z}+\tilde{\varepsilon}_i,$$

where $E(\tilde{\varepsilon}_i|\tilde{z})=0$ for any i. Multivariate normality is one such two-fund separable distribution.

In a mean-variance efficient economy, CAPM shows that the expected return on an individual asset is a linear function of the expected return on the market portfolio $\tilde{r}_M$ as long as expectations are homogeneous. For example, this can be stated symbolically as[3]

1. See Mossin (1977) and Kobayashi (1984) for further discussion on the pricing mechanism and market equilibrium in asset markets.

2. The set of utility functions that satisfy condition (a) are said to display hyperbolic absolute risk aversion (HARA) because of the following definition,

$$ARA = -\frac{u''(W)}{u'(W)} = \frac{1}{a+bW}$$

HARA utility functions are categorized into the following three types:

$$u(W) = \begin{cases} -e^{-\frac{W}{a}} \\ \log(W+a) \\ \frac{1-b}{b}(a+bW)^{1-\frac{1}{b}} \end{cases}$$

3. Equation (1) is the basic formulation of the two-parameter (r_f and β_i) CAPM. See Sharpe (1986, Chapter 7) for its derivation.

$$E(\tilde{r}_i) = r_f + \beta_i\{E(\tilde{r}_M) - r_f\}, \tag{1}$$

where $\beta_i = \dfrac{\mathrm{cov}(\tilde{r}_i, \tilde{r}_M)}{\sigma^2(\tilde{r}_M)}$.

An inspection of equation (1) suggests that, if an economy is characterized by multivariate normality, CAPM cannot be rejected. That is to say, even if the real economy is mean-variance inefficient, the assumption of multivariate normality of individual returns makes it difficult to reject CAPM. This point becomes important in testing APT.

Empirical evidence suggests that CAPM has not been successful in explaining the real variances of individual asset returns; in fact, the "beta" factor has explained little of the return variances in many empirical studies (Fama and Macbeth 1973; Black, Jensen and Scholes 1972). The poor explanatory power of the beta factor has also been confirmed in studies based on Japanese data (Maru and Royama 1974; Konya 1978; Aoyama 1979; Sakakibara 1983; Sato 1984). One study even suggested that there was no verifiable trade-off between beta and the expected return (Sakakibara 1986). This may mean that the assumptions of multivariate normality and quadratic utility are inadequate descriptions of reality; CAPM may not be an appropriate model of capital asset pricing.

2. The Basic Characterization of APT

Testing CAPM involves methodological difficulties as well. Roll (1977), for example, questioned the observability of the "true" market portfolio, hence the feasibility of testing CAPM.[4] Partly responding to this criticism, some researchers have relaxed the assumption of one factor and considered the possibility of a multi-factor return generating process. Rosenberg and Gay (1976) were among the first to construct a multi-index CAPM with "extra-market" components of return (such as macroeconomic variables) to improve the explanatory power of CAPM. However, the choice of a proper alternative to the two-parameter pricing mechanism remains an open question.[5]

APT differs from any other modifications of CAPM in one important respect: APT postulates a return generating process without including the market return. The intuitive idea behind APT is that a linear multi-factor return generating process may

4. Shanken (1987) has shown that CAPM can be rejected even if a possible deviation of the true return from the observed market return is taken into account.

5. See Elton and Gruber (1981, Chapter 6) for a discussion on the return generating process of the multi-index CAPM.

represent some aspect of the asset market. This idea does not depend on the existence of any particular relationship between individual and market returns, making it possible to test APT on any combination of individual assets whose return generating process is a linear combination of several common factors.

The ability of APT to limit the range of assets by the characteristics of the return generating process is not only a merit but also a shortcoming. On the one hand, this makes it unnecessary to compose an arbitrary market portfolio; on the other hand, it does not allow description of the market equilibrium in which all types of assets (including bonds and real assets) with different return generating processes are included. However, for the purpose of considering a narrow set of common stocks, APT appears to be a promising alternative to CAPM and its previous variants.

APT maintains the linear trade-off between risk and return and postulates that the expected return is equal to the riskless interest rate plus a risk premium. Although the risk-return trade-off is a basic feature of modern portfolio theory, APT differs from CAPM in its disregard for market risk such as CAPM's beta. Instead of postulating a specific risk factor, APT assumes several common factors which affect the evaluation of individual assets. The prominent conclusion of APT is that any remaining risks idiosyncratic to individual assets do not need to be priced, because such risks can be diversified by market efficiency, which is defined as the absence of arbitrage opportunities. Therefore, only the risks against common factors are priced in APT.

3. The Fundamental Theorem and Its Testing Framework

This idea of APT can be more formally stated as a theorem (Ross 1976), for which two assumptions are required:

Assumption 1 (linear multi-factor return generating process):

The returns on a particular subset of assets, Ω, are generated by a linear combination of k common factors and one idiosyncratic disturbance, with k smaller than the number of all assets in the subset Ω. Symbolically, this assumption is stated as,

$$\tilde{r}_i = E(\tilde{r}_i) + \beta_{i1}\tilde{\delta}_1 + \ldots + \beta_{ik}\tilde{\delta}_k + \tilde{\varepsilon}_i, \quad (2)$$

where all common factors $\tilde{\delta}_j(j=1,\ldots,k)$ and all idiosyncratic disturbances $\tilde{\varepsilon}_i(i=1,\ldots,n)$ have zero expectations and are uncorrelated with each other, and the variances of all factors are bounded.[6] (Without loss of generality, the variances of common factors can be assumed to be one.) Finally, all investors hold the same

6. The first term on the right-hand side of equation (2) is initially a parameter. However, once the expectations of both sides are calculated, this parameter is shown to be equal to the expectation of the return itself.

subjective expectations about the rates of return for all assets that belong to the subset Ω.[7]

Assumption 2 (absence of arbitrage opportunity):

In the market, there exists no arbitrage opportunity, defined as an environment in which at least one arbitrage portfolio has positive expected return with zero variance.

The Fundamental Theorem of APT:

Under Assumptions 1 and 2, the expected returns on individual assets are approximately

$$E(\tilde{r}_i)=\rho+\lambda_1\beta_{i1}+\dots+\lambda_k\beta_{ik}. \tag{3}$$

The proof of the theorem can be briefly outlined along the lines suggested by Huberman (1982). Recognizing the orthogonality among factors and disturbances for all assets $i\in\Omega$, express the parameter $E(\tilde{r}_j)$ as the linear regression formula of k factor loadings $\beta_{i1},\dots,\beta_{ik}$ and 1 (which is the coefficient of each idiosyncratic disturbance $\tilde{\varepsilon}_i$), such that

$$E(\tilde{r}_i)=\rho+\lambda_1\beta_{i1}+\dots+\lambda_k\beta_{ik}+c_i\cdot 1,$$

where ρ is the fixed term. In this formulation, c_i is equivalent to the regression error term, which satisfies $\sum_{i\in\Omega} c_i=0$, because any disturbances are uncorrelated with each other.

Then regard a vector $c=[c_i]$ as a portfolio whose components are equal to the dollar amount put into each asset i. Since it costs nothing to acquire a vector c, it is

7. This assumption says that the expectations operator $E(\cdot)$ is applicable not only to the market but also to any individual investor.

8. Suppose a portfolio αc. It is an arbitrage portfolio for any $\alpha \neq 0$ such that

$$\Sigma \alpha c_i = 0,$$

$$E(\alpha c\cdot\tilde{r}) = \alpha\Sigma c_i^2,$$

$$\text{and } \operatorname{var}(\alpha c\cdot\tilde{r}) = \alpha^2\Sigma c_i^2.$$

If $\alpha=(\Sigma c_i^2)^{-\frac{3}{4}}$ and Σc_i^2 approaches infinity,

$$\lim_{\Sigma c_i^2\to\infty} E(\alpha c\cdot r) = \lim_{\Sigma c_i^2\to\infty} (\Sigma c_i^2)^{\frac{1}{4}} = \infty,$$

$$\text{and } \lim_{\Sigma c_i^2\to\infty} \operatorname{var}(\alpha c\cdot r) = \lim_{\Sigma c_i^2\to\infty} (\Sigma c_i^2)^{-\frac{1}{2}} = 0.$$

This result indicates that the expected return or the arbitrage portfolio αc approaches infinity while its variance remains zero as Σc_i^2 approaches infinity.

one of the arbitrage portfolios. Suppose now that the squared sum of c_i, $\sum_{i\in\Omega} c_i^2$, increases to infinity. In this environment, Huberman (1982) shows the existence of arbitrage portfolios.[8]

The contrapositive proposition of this case states subsequently that, if there exists no arbitrage opportunity, the squared sum $\Sigma\, c_i^2$ cannot increase to infinity, so that most of c_i's are small and may be approximately zero when the number of assets belonging to the subset $\Omega(n)$ is large.[9] Finally, we conclude that the evaluation formula given by equation(3) holds for most of the assets belonging to the subset.

It is obvious that ρ in equation(3) represents the riskless interest rate because all common factor loadings are zero for riskless assets, making the expected return on such riskless assets equal to ρ. Consequently, the fundamental theorem of APT implies that the expected returns on individual assets are equal to the riskless interest rate plus a risk premium against common factors in the absense of arbitrage opportunities.

Though definitionally clear, APT has two conceptual poblems: the indeterminacy of the factor structure (Shanken 1982) and the acceptability of the linear multifactor return generating process as a process characterizing the entire asset market. First, the indeterminacy problem can be solved if we assume that the variances of factors are unity and the variances of idiosyncratic disturbances are unique.[10] Although the standardization of the variances of random effect factors is difficult without knowing the distribution of asset returns, the variances in a subset of the entire asset market are computable. Thus, the indeterminacy of the factor structure can be overcome, and factor analysis applied to the limited range of asset returns remains useful as a test of APT.

The second problem can also be solved if we accept the fundamental theorem of APT for some subset of assets. It is true that the assumption of multi-factor linearity is as restrictive as that of two fund separability for real assets and human wealth. However, acceptance of the fundamental theorem for some subset of assets allows us to assume that all observed assets in the study are subject to the same linear multifactor return generating process.

Let us discuss the interpretation of the combined hypotheses of multi-factor linearity and the absence of arbitrage opportunity. The theorem describes a price

9. Note that the finiteness of Σc_i^2 is a necessary condition for accepting the linear multi-factor return generating process. But the infinity of the number of assets in subset Ω is irrelevant to the theorem.

10. To satisfy this requirement, data are often normalized to N(0,1). However, such a normalization is possible only if the data are subject to the asymptotic normal distribution. Thus, in order to avoid specifying the distribution of asset returns, our standardization was made only in terms of the observed mean and variance (See the Appendix).

formulation in the asset market which satisfies the two underlying assumptions, but it says nothing about the market equilibrium that results from investors' optimization. Although it can be shown that equilibrium prices do exist with multi-factor linearity, such linearity cannot be proved because the factor representation is too general to be tested. However, if we assume that the fundamental theorem holds for a subset of assets and that there is no arbitrage opportunity, we can accept multi-factor linearity for such a limited range of assets. Thus, we apply the theorem to the Japanese stock market as a representative stock market on the assumption that it is efficient enough to preclude the existence of arbitrage opportunity.

Such tests of APT must ideally satisfy the following three requirements:

⟨I⟩ No particular distribution should be specified for asset returns. For example, if multi-variate normality is assumed, CAPM will be accepted and tests of APT will be redundant.

⟨II⟩ Similarly, no distributional or other qualitative properties should be specified for individual factors.

⟨III⟩ The number of sample assets should be large enough to guarantee the unique variances for idiosyncratic disturbances.

Though all of these requirements are required for successful empirical tests of APT, one or two of them have not been satisfied in most previous studies.

III. The Method of Testing APT

1. Problems with Previous Methods

Traditional tests of APT, first employed by Roll and Ross (1980), have three main features: 1) decomposition of the covariance matrix of stock returns by maximum likelihood factor analysis; 2) determination of the number of common factors by a chi-square test; and 3) division of samples into smaller subsamples by portfolio grouping procedure. Each of these, however, is conceptually inadequate for testing APT.

First, decomposing the covariance matrix of stock returns by maximum likelihood estimation conflicts with requirement ⟨I⟩, because the distribution of stock returns is unknown and asymptotic normality may not be specified for its likelihood function (Roll and Ross 1980).[11] If asymptotic normality is postulated, rejection of

11. "(H)owever, there may be some problems attendant to the M.L.E. method because the likelihood function involved is that of a multivariate gaussian distribution. To the extent that the data have been generated by a non-gaussian probability law, unknown biases and inconsistencies may be introduced." (Roll and Ross 1980, p.1087)

equation(3) may be due to the misspecification of the return distribution rather than to the inappropriateness of the joint hypothesis of the multi-factor linearity of the return generating process and the absence of arbitrage opportunity. Maximum likelihood estimation is too restrictive as a method of testing APT.[12]

Second, a chi-square test is not powerful enough to determine the number of common factors, because it assumes that all common factors satisfy multivariate normality; such an assumption is not allowed by requirement ⟨II⟩. Since the sensitivity of the chi-square test to nonnormal data is unknown, several tests should be applied to determine the number of common factors.[13]

Third, the procedure in which assets are grouped into disjoint equally-weighted portfolios would present several problems in testing APT.The smaller the number of assets is, the more difficult portfolio diversification becomes and the more likely it is to reject the hypothesis of no arbitrage opportunity. In the context of APT, undiversified portfolios increase the idiosyncratic disturbances, which conflicts with requirement ⟨III⟩. Thus, the portfolio grouping procedure does not guarantee a unique factor structure. In other words, the characteristics of each factor of one subgroup may be different from that of another subgroup. For instance, suppose that the original sample of 100 stocks is divided into two samples of 50 each. In this case, even if five common factors are adequate for each subgroup, all factors may not be the same. Therefore, we should not divide the original sample into small groups for testing APT.

2. New Method of Testing APT: Least Squares Factor Analysis

To avoid these potential problems, we propose a new method of testing APT. This method can be characterized by the following three features: 1) use of least squares factor analysis to decompose the covariance matrix of asset returns so as to minimize the error between population and sample correlation matrices; 2) use of several statistical tests to determine the number of common factors; and 3) use of original individual data that are not subdivided. The procedure of the test can be outlined as follows.

12. In practice, maximum likelihood is often the most straightforward method. Thus, if testing APT is not the objective of the study, it is convenient to decompose a covariance matrix of asset returns by maximum likelihood estimation.

13. Kryzanwsky and To (1983) and Trzcinka (1986) suggest the scree test of characteristic roots of the covariance matrix and the chi-square test for determining the number of common factors. Other studies of APT use only the chi-square test (Roll and Ross 1980; Brown and Weinstein 1983; Cho 1984; Drymes, Friend and Gultekin 1984; Drymes, Friend, Gultekin and Gultekin 1985).

Step (1): Calculate the sample covariance matrix for the panel data of individual asset returns. The sample covariance matrix is equal to the sample correlation matrix if the data are standardized with mean zero and variance one.

Step (2): Set the number of common factors at k=1, and apply least squares factor analysis to the above sample correlation matrix in order to estimate equation(2). In working with the sample correlation matrix S and the population correlation matrix Σ_s, we define the measure of distance between the two correlation matrices as

$$L=\frac{1}{2}\mathrm{tr}\{(S-\Sigma_s)S^{-1}\}^2 \tag{4}$$

where tr is the trace operator. Then, minimization of L with respect to Σ_s will yield the first order conditions as

$$(S-\Sigma_s)\Sigma_s^{-1}\hat{B}_s = 0 \tag{5}$$

$$\hat{\psi}_s = \mathrm{diag}(S-\hat{B}_s\hat{B}_s') \tag{6}$$

where $\hat{B}_s$ is the matrix whose components β_{ij} are factor loadings of equation(2) for any i=1,..., and j=1,...,k (however, in this step k=1). $\hat{\psi}_s$ is the covariance matrix of idiosyncratic disturbances, and diag is the diagonal operator. The solutions for the above simultaneous equations, $\hat{B}_s$ and $\hat{\psi}_s$, must satisfy the following relation,

$$\Sigma_s = \hat{B}_s\hat{B}_s'+\hat{\psi}_s, \tag{7}$$

which is the fundamental equation of factor analysis. For the convenience of computation, give an initial value to ψ_s in equation(6), and repeat solving equations(5) and (6) until the solutions converges to equation(7).[14] Then, equation(2) is estimated for k=1.

Step (3): Next, increase the number of factors k and repeatedly estimate $\hat{B}_s$ by the same procedure as in the second step. Then, use several criteria to select the adequate factor representation for different values of k. Since different criteria may suggest different factor representations, we choose the criterion that gives the minimum number of factors. Here the chi-square test, Akaike's Information Criterion (AIC), and Schwarz's Bayesian Criterion (SBC) are applied to estimate $\hat{B}_s$.

Step (4): Next, repeat the same procedure by choosing different time spans and stock compositions, and examine whether there are any significant differences in the estimated value of $\hat{B}_s$ and the selected number of factors.

Step (5): Rewrite equation(3) with equation(2) for the cross-sectional data from i=1 through n, and obtain the regression model

14. In this paper, Joreskog (1977)'s algorithm is adopted in the SAS program. See Okamoto (1986) for discussion on algorithms used in factor analysis.

$$r=\lambda_i\beta_i+\ldots+\lambda_k\beta_k+\rho+\beta_i\delta_i+\ldots+\beta_k\delta_k+\varepsilon \tag{8}$$

where r is the dependent variable, $\beta_i,\ldots,\beta_k$ are independent variables, ρ is a fixed term, and $\beta_i\delta_i+\ldots+\beta_k\delta_k+\varepsilon$ are error terms. Since the variance of the sum of error terms is equal to the covariance of r ($=B_s\, B_s'+\psi_s$) and its estimator has already been obtained in the third step, estimate regression coefficients $\lambda_i,\ldots,\lambda_k$ by generalized least squares (GLS). Finally, test whether any of the GLS estimators, $\lambda_i,\ldots,\lambda_k$ are zero. Because only the chi-square test is available for this test, the significant level of the test should not be strict.

Step (6): Last of all, estimate the factor scores of $\delta_i,\ldots,\delta_k$ by the usual least squares method. Use of least squares (instead of maximum likelihood) satisfies requirement ⟨II⟩ described in section II.3. Since the least squares method does not require that idiosyncratic disturbances be subject to the asymptotic normal distribution as the maximum likelihood method does, least squares estimation is better for testing APT.

Some words of caution are in order concerning the potential problems with the least squares method. The most serious is the measurement problem regarding the error between population and sample correlation matrices. Although there are several types of fit function for measuring the error, we adopt the one which introduces the same estimator of factor loading as the maximum likelihood method does. Comparison of our least squares estimators with the other estimators in past studies will suggest that the previous estimates based on maximum likelihood are the same as the least squares estimates. (See the Appendix for the details on least squares estimation.)[15]

Relative to step (4), the idea of adopting different sample covariance matrices of asset returns was first suggested by Brown and Weinstein (1983).[16] The objective of this procedure is to generalize the estimated factor representation. If the factor-loading estimates are stable across different samples, the set of estimates will be unique and robust. Hence, unless we obtain stable factor-loading estimates, we will not be able to use a single set of estimates to test the fundamental theorem of APT. Because it is difficult to make a complete comparison of different samples, partial change technique is applied.

15. Note that translating maximum likelihood estimators into least squares estimators requires additional tests for nonnormality. Therefore, the interpretations of the statistical significance or the factor number may need to be reconsidered.

16. The main point of Brown and Weinstein (1983) lies in the bilinear test of asset pricing. Although we agree with their generalization technique, we do not follow their paradigm itself.

IV. Empirical Results

1. Data Description

The data are obtained from the Japanese stock return file compiled by the Japan Securities Research Institute. Individual rates of return are adjusted for stock splits and dividends. Monthly rates of return are represented as percentage change per annum from the end of one month to the next.

The original data file covers all individual stocks listed in the first section of the Tokyo Stock Exchange for the period between November 1951 (or the first trading month for new stocks) and December 1984. Because the number of observations per stock is at most 400, factor analysis requires that the number of stocks be much less than 400; thus, we need to limit the number of stocks in the sample to no greater than 350. Given other restrictions, we finally decided to limit the sample size to 225 stocks that are included in the Nikkei-Dow stock index; these stocks are representative of stocks traded in the Japanese stock market in terms of daily trading activity. Moreover, the selection was further motivated by the following considerations: (1) no one industry is to be over-represented; (2) some construction stocks (listed on the exchanges for the first time in 1961) are to be included, though this would limit the number of observations to about 250;(3) the same sample is to be maintained throughout the period.

The classification used in the Nikkei-Dow stock index sufficiently meet these considerations, except for the stocks of 16 financial institutions whose prices were in effect pegged during the whole period and for the stock of Taito, a sugar manufacturing company. Thus, our sample consists of 208 stocks out of the 225 stocks in the Nikkei-Dow index.[17]

2. Determining the Number of Common Factors

Table 1 compares results of different factor structures in terms of the number of common factors; the average statistics of communalities among all sample stocks, chi-square statistics, AIC statistics, and SBC statistics are calculated in order to determine the adequate factor structure for our sample data. A communality represents the portion of the individual return's variance which is due to common factors. When each return is standardized with mean zero and variance one, the communality is equal to the squared sum of factor-loading estimates.

The sequence of the average statistics of communalities implies that as the

17. Whenever there is a missing observation for a stock, the missing value was set equal to the sample mean of all remaining observations. Each stock had 5 to 10 such missing values out of 264 monthly observations.

Table 1. Comparison of Number of Common Factors (full-sized sample)

Sample: Returns of 208 Nikkei-Dow classified stocks, standardized with mean 0 and variance 1.
Period: January 1963 to December 1984; the number of observations is 264.

Hypothesized model	Average of communalities[1]	Chi-square statistics (degrees of freedom)[2]	AIC statistics	SBC statistics
Non-factor model	–	51154. (21528)	–	–
1-factor model	0.223	41342. (21320)	57628.	29558.
2-factor model	0.291	37937. (21113)	53546.	27887.
3-factor model	0.328	35796. (20907)	51179.	27072.
4-factor model	0.360	34237. (20702)	49597.	26648.
5-factor model	0.378	33104. (20498)	48595.	26511.
6-factor model	0.399	32047. (20295)	47686.	26419.
7-factor model	0.418	31023. (20093)	46813.	26344.
8-factor model	0.431	30297. (19892)	46346.	26470.
9-factor model	0.444	29610. (19692)	45929.	26619.
10-factor model	0.456	28963. (19493)	45558.	26789.
11-factor model	0.468	28351. (19295)	45231.	26980.

Note: 1. The average of all individual communalities, given by the squared sum of factor loadings.
2. The hypothesis that the number of common factors is sufficient is rejected by likelihood ratio tests at the .01 significance level.

Table 2. Comparison of Number of Common Factors (1st half-sized sample)

Sample: Returns of 104 Nikkei-Dow classified stocks, standardized with mean 0 and variance 1.
Period: January 1963 to December 1984; the number of observations is 264.

Hypothesized model	Average of communalities[1]	Chi-square statistics (degrees of freedom)[2]	AIC statistics	SBC statistics
Non-factor model	–	19724. (5356)	–	–
1-factor model	0.246	13478. (5252)	16102.	8423.
2-factor model	0.317	11407. (5149)	13937.	7525.
3-factor model	0.351	10207. (5047)	12776.	7126.
4-factor model	0.386	9347. (4946)	12003.	6921.
5-factor model	0.411	8753. (4846)	11536.	6866.
6-factor model	0.435	8102. (4747)	10996.	6773.
7-factor model	0.454	7640. (4649)	10674.	6787.
8-factor model	0.469	7205. (4552)	10378.	6812.
9-factor model	0.485	6836. (4456)	10115.	6873.
10-factor model	0.496	6536. (4361)	10012.	6971.
11-factor model	0.509	6256. (4267)	9887.	7077.

Note: See notes in Table 1.

Table 3. Comparison of Number of Common Factors (2nd half-sized sample)

Sample: Returns of 104 Nikkei-Dow classified stocks, standardized with mean 0 and variance 1.
Period: January 1963 to December 1984; the number of observations is 264.

Hypothesized model	Average of communalities[1]	Chi-square statistics (degrees of freedom)[2]	AIC statistics	SBC statistics
Non-factor model	–	19459. (5356)	–	–
1-factor model	0.221	13995. (5252)	16704.	8724.
2-factor model	0.295	12107. (5149)	14754.	7933.
3-factor model	0.342	10927. (5047)	13619.	7548.
4-factor model	0.376	9942. (4946)	12702.	7270.
5-factor model	0.407	9181. (4846)	12041.	7118.
6-factor model	0.430	8571. (4747)	11550.	7050.
7-factor model	0.449	8070. (4649)	11182.	7041.
8-factor model	0.467	7586. (4552)	10831.	7039.
9-factor model	0.481	7213. (4456)	10606.	7098.
10-factor model	0.496	6869. (4361)	10411.	7170.
11-factor model	0.509	6534. (4267)	10221.	7243.

Note: See notes in Table 1.

Table 4. Comparison of Number of Common Factors (full-sized sample)

Sample: Returns of 208 Nikkei-Dow classified stocks, standardized with mean 0 and variance 1. The missing values in the rate of return series of the stocks of 10 construction companies before 1963 are replaced by the average rates of return for the period 1963–84.
Period: January 1954 to December 1984; the number of observations is 372.

Hypothesized model	Average of communalities[1]	Chi-square statistics (degrees of freedom)[2]	AIC statistics	SBC statistics
Non-factor model	–	62758. (21528)	–	–
1-factor model	0.237	46363. (21320)	58290.	29960.
2-factor model	0.297	41868. (21113)	53249.	27845.
3-factor model	0.330	39235. (20907)	50499.	26874.
4-factor model	0.355	37384. (20702)	48709.	26831.
5-factor model	0.373	35909. (20498)	47377.	26115.
6-factor model	0.389	34584. (20295)	46224.	25936.
7-factor model	0.407	33339. (20093)	45161.	25800.
8-factor model	0.422	32212. (19892)	44239.	25733.
9-factor model	0.435	31275. (19692)	43549.	25780.
10-factor model	0.446	30469. (19493)	43017.	25903.
11-factor model	0.467	29663. (19295)	42479.	26022.

Note: See notes in Table 1.

Table 5. Comparison of Number of Common Factors (1st half-sized sample)

Sample: Returns of 104 Nikkei-Dow classified stocks, standardized with mean 0 and variance 1. The missing values in the rate of return series of the stocks of 10 construction companies before 1963 are replaced by the average rates of return for the period 1963–84.
Period: January 1954 to December 1984; the number of observations is 372.

Hypothesized model	Average of communalities[1]	Chi-square statistics (degrees of freedom)[2]	AIC statistics	SBC statistics
Non-factor model	–	25577. (5356)	–	–
1-factor model	0.256	15964. (5252)	18152.	9483.
2-factor model	0.320	13470. (5149)	15617.	8418.
3-factor model	0.349	12097. (5047)	14319.	7969.
4-factor model	0.382	10870. (4946)	13177.	7596.
5-factor model	0.408	10012. (4846)	12441.	7423.
6-factor model	0.430	9218. (4747)	11770.	7282.
7-factor model	0.448	8492. (4649)	11171.	7175.
8-factor model	0.464	7936. (4552)	10758.	7158.
9-factor model	0.478	7405. (4456)	10368.	7151.
10-factor model	0.491	7040. (4361)	10162.	7234.
11-factor model	0.500	6758. (4267)	10046.	7361.

Note: See notes in Table 1.

Table 6. Comparison of Number of Common Factors (2nd half-sized sample)

Sample: Returns of 104 Nikkei-Dow classified stocks, standardized with mean 0 and variance 1.
Period: January 1954 to December 1984; the number of observations is 372.

Hypothesized model	Average of communalities[1]	Chi-square statistics (degrees of freedom)[2]	AIC statistics	SBC statistics
Non-factor model	–	25272. (5356)	–	–
1-factor model	0.240	16420. (5252)	18658.	9737.
2-factor model	0.310	13748. (5149)	15927.	8573.
3-factor model	0.347	12467. (5047)	14732.	8175.
4-factor model	0.376	11262. (4946)	13615.	7815.
5-factor model	0.405	10209. (4846)	12661.	7534.
6-factor model	0.426	9436. (4747)	12015.	7405.
7-factor model	0.443	8789. (4649)	11505.	7342.
8-factor model	0.459	8258. (4552)	11120.	7339.
9-factor model	0.472	7800. (4456)	10814.	7374.
10-factor model	0.485	7402. (4361)	10571.	7439.
11-factor model	0.500	6976. (4267)	10294.	7485.

Note: See notes in Table 1.

number of common factors increases, the individual return's communality increases. It means that the 11th factor will explain only one percent (=.509-.496) of the total variance, but the explanatory power of the additional factors will not be negligible until the 30th factor.

The same implication is obtained by both sequences of chi-square statistics and AIC statistics, but the most striking result observed in Table 1 is that SBC statistics reach the minimum value when there are seven common factors and, therefore, a seven-factor model is the most adequate model to explain the variances of the original sample.

Because there is no specific reason to believe that the SBC criterion is the best for determining the number of common factors, it might be helpful to examine the sensitivity of SBC statistics to the choice of data by applying APT to two different types of samples. The first type of samples differ from the original one in the composition of stocks. In particular, the original sample is equally divided into two by type of industry; for example, in one subsample, chemical and food firms are included, while machinery and trading firms are in the other half. The test results for the first subsample indicate that, while the sequences of average statistics of communalities, chi-square statistics, and AIC statistics do not converge to a minimum value, SBC statistics attain a minimum value with 6 or 7 factors (Table 2). The results for the other half are broadly similar: SBC statistics take virtually the same minimum value with 6 to 8 factors (Table 3). These results seem to suggest that the optimal number of common factors implied by SBC statistics are stable at 7 even if the composition of samples is cross-sectionally different.

The sensitivity of SBC statistics can also be tested by composing an artificial sample. This can be done by using the sample average values from the period 1963-84 for the missing observations of the returns of 10 construction companies during 1954-62. If the statistics are stable, we would expect them to indicate an additional common factor. In fact, the results for the full-sized (208 stocks), the first half-sized, and the second falf-sized samples show that all but SBC statistics suggest an increase in the optimal number of common factors from 6-7 to 8-9 for the full and first half-sized samples which include artificially fixed returns of 10 construction firms. On the other hand, the optimal number of factors remains the same for the second half-sized sample even if the sample period is extended for more than a hundred months. This sample is not affected by the addition of some 100 artificial returns for the constructing firms.

Thus, the sequence of SBC statistics seems more reliable than those of average statistics of communalities, chi-square statistics, or AIC statistics as a method of determining the number of common factors. Moreover, a seven factor representation seems to be a reasonable representation of the return generating process for 208 Japanese stocks.

3. Stability of the Seven-factor Model

Before actually testing APT, it may be useful to check the stability of the factor-loading estimates across different samples. Since the seven-factor representation for the variances of stock returns is equivalent to neglecting marginal common factors, we need to confirm the uniqueness of the factor representation. First, we prepare the standard deviation of individual factor-loading estimates obtained from several partly different samples. Second, we calculate the cross-sectional average of those standard deviations for each factor loading.

Table 7 shows the comparison among factor-loading estimates which are different in the periods. For the full-sized sample, three factor-loading estimates are provided for this test: the estimates for the period 1963-84, those for 1963-82 and those for 1965-84. The cross-sectional average of the individual standard deviations is .020 for the first factor-loading estimates, and below .03 for the third through seventh factor loadings, but the average is as large as .256 for the second factor-loading. Since the value of each factor is standardized by mean zero and variance one, these results imply that most factor-loading estimates are stable.

For the half-sized samples, we can obtain more factor-loading estimates by shortening the sample period. For the first half-sized sample, four factor-loading estimates are compared. Three of the four estimates are obtained from the same periods as the full-sized samples, and the other from the period 1970-84. The cross-sectional average of the standard deviations of four individual factor-loading estimates are below .082 for all but the sixth factor-loading, for which the average statistic is .122.

The second half-sized sample does not have the period constraint because the new stocks, such as those of 10 construction firms, are not included. Thus, we can obtain another estimate for the period of 1955-69 as well as for the original period of 1963-84. The result that any cross-sectional averages of the standard deviations of five individual factor-loading estimates are below .134 is surprising because the estimates are very stable even though sample periods are completely different. Although it may not be possible to generalize these results, the choice of a sample period seems to have only a small effect on the factor-loading estimates when the sample is as large as ours.

Next we compare two factor-loading estimates which are estimated from the two different covariance matrices: one is the full-sized, and the other the half-size covariance matrix. Comparison of these two matrices shows that the composition of the sample covariance matrices is virtually irrelevant to the factor-loading estimation. According to Table 8, all average deviations between the full- and half-sized estimates for the period of 1963-84 are .1, with the average deviation of the second factor-loading estimates being the only exception. Thus, it is concluded that the

Table 7. Test of the Time Series Stability of Factor Loading Estimates

	Average of the standard deviations of factor loading estimates		
Size of sample	Full size	1st half-size	2nd half-size
Compared estimates	3	4	5
1st factor loading	0.020	0.030	0.057
2nd factor loading	0.256	0.042	0.074
3rd factor loading	0.029	0.059	0.127
4th factor loading	0.017	0.033	0.099
5th factor loading	0.019	0.044	0.120
6th factor loading	0.018	0.122	0.134
7th factor loading	0.018	0.082	0.126

Table 8. Test of the Cross-sectional Stability of Factor Loading Estimates

	Average of the standard deviations of differences between full- and half-sized factor loading estimates	
Size of sample	1st half size	2nd half-size
1st factor loading	0.027	0.030
2nd factor loading	0.107	0.289
3rd factor loading	0.095	0.158
4th factor loading	0.093	0.108
5th factor loading	0.091	0.119
6th factor loading	0.075	0.133
7th factor loading	0.114	0.122

Table 9. Tests of the Fundamental Theorem of APT

Number of tests of equation (8): 264 for each sample, total 792

	Number of times the null hypothesis that $\lambda_1 = \cdots = \lambda_7 = 0$ is rejected (as percent of total)			
Significance level	5%	10%	25%	50%
Size of sample				
Full-size	43 (16%)	49 (19%)	75 (28%)	118 (45%)
1st half-size	55 (21%)	77 (29%)	99 (38%)	160 (61%)
2nd half-size	67 (25%)	78 (30%)	107 (41%)	160 (61%)
Total	165 (21%)	204 (26%)	281 (35%)	438 (55%)

choice of a sample covariance matrix—whether in terms of sample period or sample composition—has little effect on factor-loading estimation. Therefore, any set of factor-loading estimates will yield broadly similar results in testing the fundamental theorem of APT.

4. Tests of the Fundamental Theorem of APT

Factor analysis indicates that the seven-factor model has more explanatory power than any regression model with seven or less exogenous variables. It remains to be seen, however, whether idiosyncratic disturbances accounted for the remaining variance or they are negligible by virtue of diversification.

In order to examine the adequacy of the seven-factor model, we perform the fifth step of the test by a) preparing three samples—one full-sized, and two half-sized—for the period between January 1963 and December 1984, b) regarding the factor-loading estimates of each sample as explanatory variables, c) applying generalized least squares to equation(8) in section III.2. for all 264 months from 1963 until 1984, and d) testing the significance of the regression coefficients of the factor loadings by chi-square statistics.

The resuts show that, out of 264 tests for each sample, only 21% could be accepted at the 5% level of significance, and 55% could be accepted at the 50% level (Table 9). Although the choice of the significance level is not clear in a sample that is subject to nonnormality, the seven-factor model seems to represent the return generating process of Japanese stocks quite reasonably.

V. Some Discussion on Estimating the Stock Return Generating Process

1. Comparison with Previous Empirical Studies

The empirical results reported in the previous section suggests that APT is a reasonable model of stock pricing in the Japanese stock market. These results are much more favorable to APT than such previous studies as Roll and Ross (1980) and Brown and Weinstein (1983) in terms of statistical significance, the adequacy of seven common factors to characterize a large number of sample stocks, and independence from variations in sample period and stock composition.[18]

In contrast, Roll and Ross (1980) found three factors to be adequate for 30 stocks, Cho, Elton and Gruber (1984) found five to six factors, and Brown and

18. The results of these studies may be altered if the SBC criterion is used. Moreover, the results obtained for U.S. data may not be applicable to Japanese data.

Weinstein (1983) found the number to be insensitive to the size of the sample. Our results, however, are consistent with Horimoto (1986) who could not reject the hypothesis of more than five common factors for Japanese stocks. Other studies have shown that the number of factors increases with the number of stocks in the sample; some of these studies have suggested the number to be 13 to 17 for 90 stocks on the basis of chi-square tests (Drymes, Friend, and Gultekin 1984; Drymes, Friend, Gultekin and Gultekin 1985). The use of the SBC criterion, however, has suggested the number to be relatively small and stable, allowing for some general conclusions to be drawn.

Although at a sufficiently high significant level,·even our methodology would reject APT. Nevertheless, recognizing that even the previous study most favorable to APT (Drymes, Friend, Gultekin and Gultekin 1985) accepted the fundamental theorem only 25% of the times at the 5% significant level,[19] one can say that our results are quite favorable to APT even at the 50% significant level.

2. Estimation of Factor Scores

To conclude this section, it may be useful to clarify the quantitative features of the seven common factors by estimating factor scores. The factor scores, standardized with mean zero and variance one, are estimated from the original covariance data, the estimated factor loadings and variances of idiosyncratic disturbances.

Figures 1 through 7 show the image of the least squares estimates for each factor score. It should be noted that, since the estimates for factor loadings are independent of sample differences, those for factor scores are also similar across the samples. The most remarkable feature is that the patterns of the first and second factor scores are analogous to the macroeconomic trend. In particular, variance patterns of these two factor scores are different depending on whether the rate of economic growth was higher or lower, or whether the sample period was before or after the two sharp increases in oil prices in 1973 and 1979. These patterns, however, were not observed for the third or higher factor scores, suggesting that the higher factors do not have the characteristics of macroeconomic trend but are related to microeconomic or non-economic variables.

Although it is difficult to identify factor characteristics, a factor that is directly related to macroeconomic activity is likely to be either the first or second factor; however, considering the lack of correlation between any two factors,one should note that both the first and second factors cannot be a macroeconomic-related factor

19. In a specific set of samples, Drymes, Friend, Gultekin and Gultekin (1985) accepted APT in 90% of the times at the 5% significant level.

Figure 1. Trend in First Factor Score

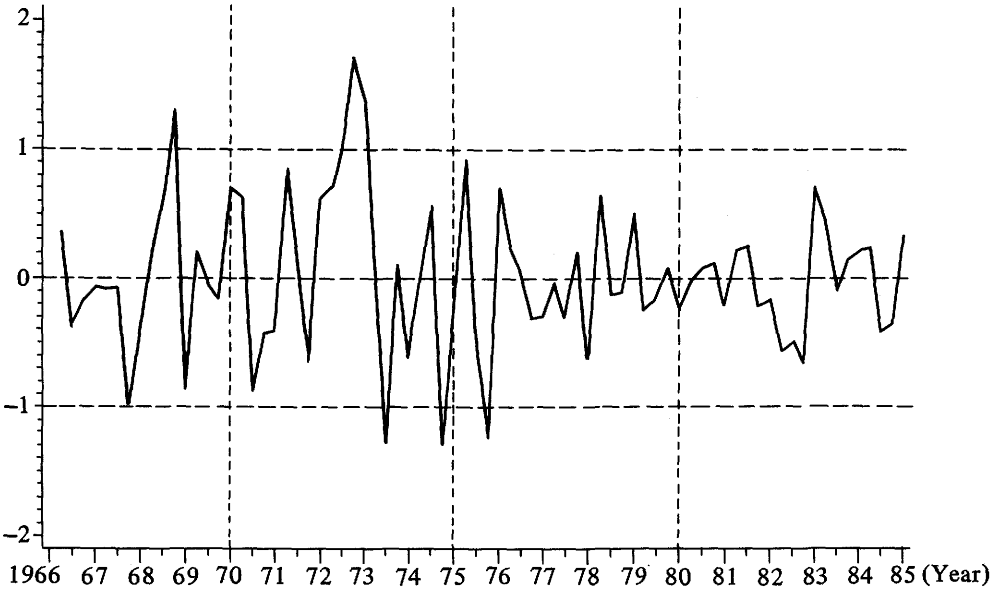


Figure 2. Trend in Second Factor Score

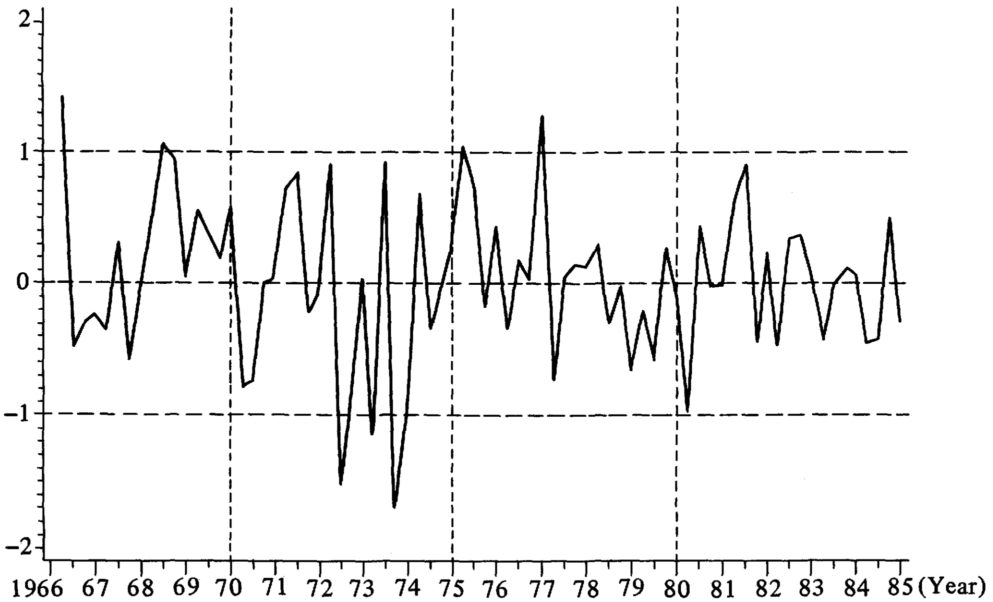


Figure 3. Trend in Third Factor Score

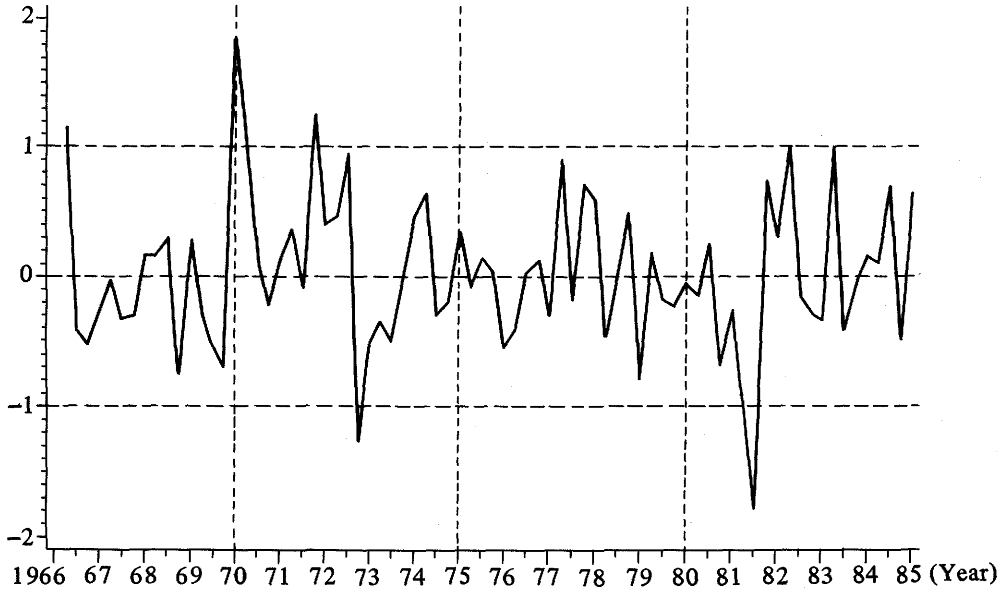


Figure 4. Trend in Fourth Factor Score

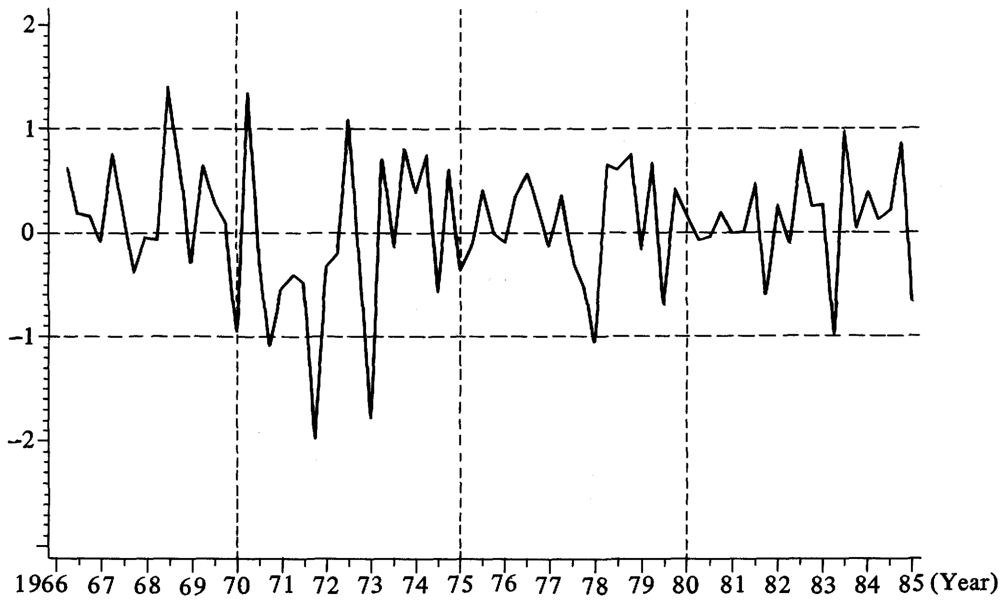


Figure 5. Trend in Fifth Factor Score

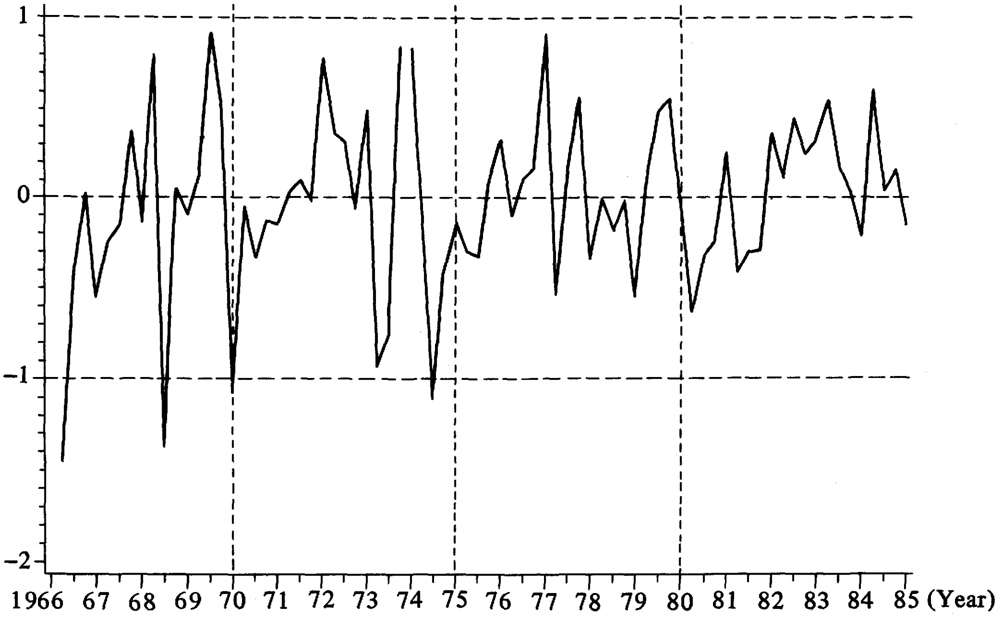


Figure 6. Trend in Sixth Factor Score

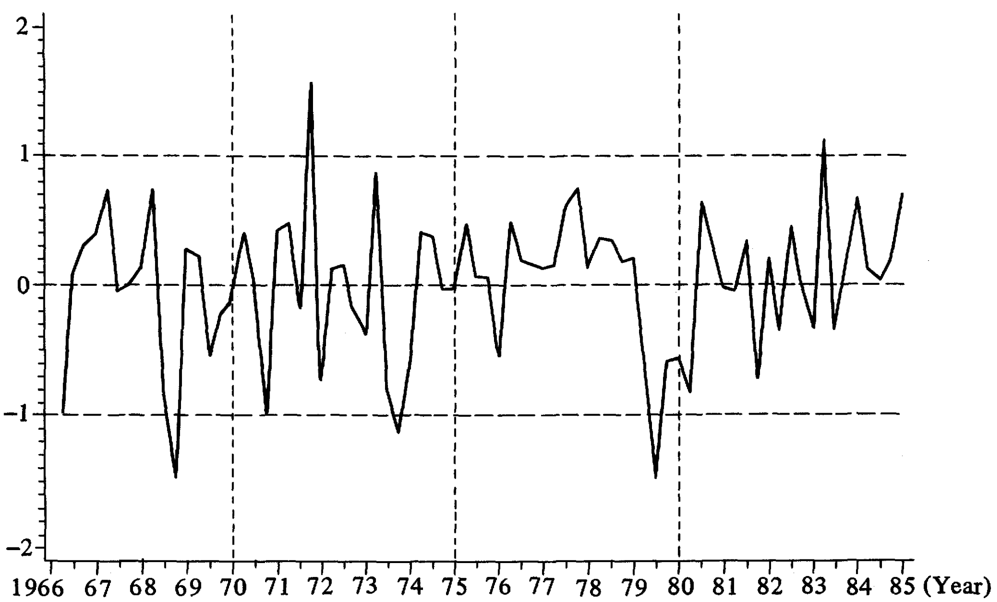


Figure 7. Trend in Seventh Factor Score

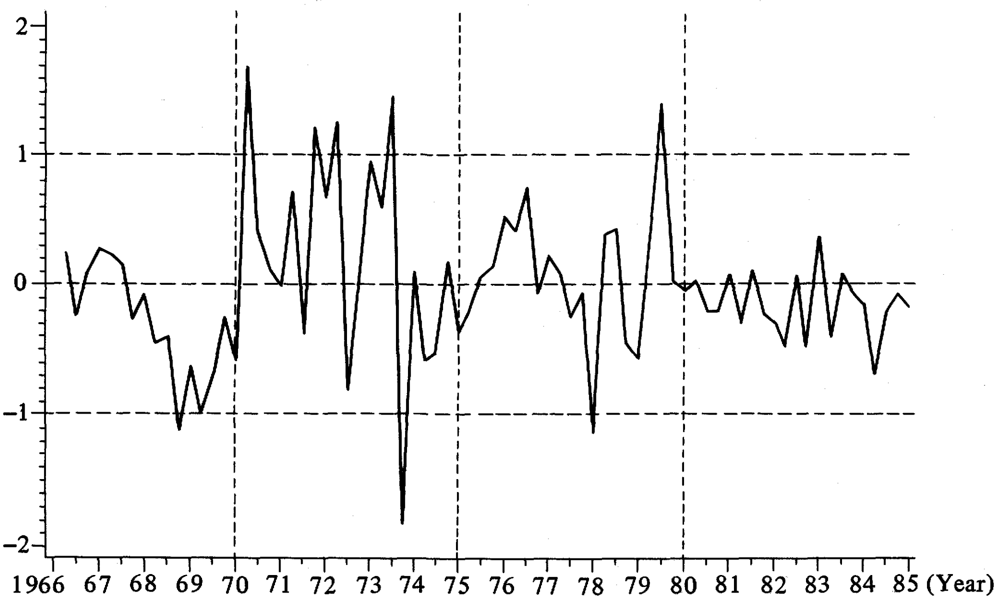
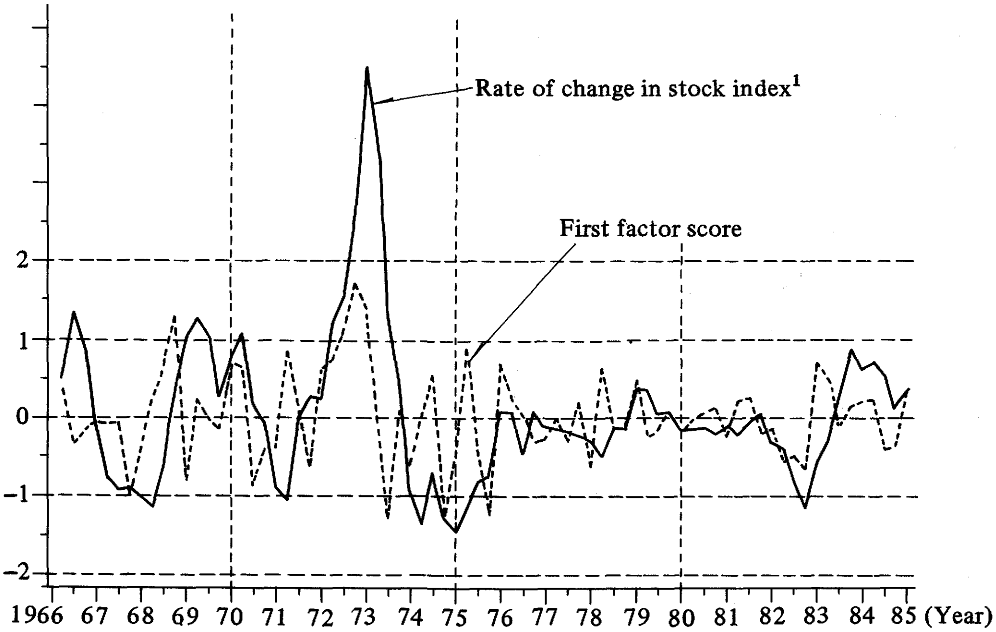


Figure 8. Comparison of Rate of Change in Nikkei-Dow Stock Index and First Factor Score



Note: 1. Monthly change at an annual rate; normalized with mean 0 and variance 1.

at the same time.

Tentatively, we compare the first factor with the Nikkei-Dow Tokyo Stock Index as a proxy for the average market return. Figure 8 shows the similarity between the two series until the 1970s; however, the first factor score diverged from the Stock Index during the 1980s. This latter result coincides with the failure of the market model or single-index CAPM to explain stock market returns. APT with additional factors can explain the asset price generating process which can be adequately represented as multi-factor linear. According to this intuitive discussion, estimating factor scores is useful, making it possible to extend the empirical range of APT.

VI. Summary and Conclusion

Our empirical results support the adequacy of the seven-factor representation for 208 Nikkei-Dow stock returns during the period of more than 20 years. In addition, the tests accepted APT as a plausible model of the stock return process. These results were obtained by factor analysis that was applied to the covariance matrix of asset returns; our results have demonstrated the usefulness of covariance decomposition in empirical studies of asset pricing.

Specifically, the main results of the paper can be summarized as follows: (1) The seven-factor model could explain the variance of 208 individual returns reasonally well during the period of over 20 years; (2) The number of common factors in the model was always seven if the covariances among stock returns were decomposed under Schwarz's Bayesian Criterion; (3) Each estimated factor loading (that is, the coefficient of each common factor) was almost invariant to changes in sample period or composition; (4) More than a half of tests accepted Ross's fundamental theorem of APT, such that the risk premium in expected individual returns resulting from idiosyncratic disturbances was small.

Although only seven factors could explain the twenty-year variance of 208 stock returns remarkably well, two questions remain concerning the seven-factor representation. One is related to SBC and the other is its future stability. Although the criterion for model selection should be unique, it should also have some relation to other criteria. Unfortunately, SBC is the only criterion that allows unique and robust determination of a relatively small number of factors.

Whether the seven-factor representation will continue to have explanatory power in the future remains an open question. The stability of the model requires two conditions. The first is the adequacy of the seven-factor representation itself, and the second is the stability of factor-loading estimates. As we discussed, the model was seen to be robust against different samples. However, the model may not be as robust if the sample is extended into the future and become quite different from the original sample. The factor-loading estimates may still be stable when the sample period is extended for another several years, but they may not be if extended for too many more years.

Finally, we should reemphasize the importance of the return generating process that is implied by APT. In APT, a linear multi-factor return generating process characterizes the mechanism of expectation formation in asset pricing. A set of common factors is equivalent to the market fundamentals which investors use in their evaluation of asset prices. In fact, we can characterize an asset pricing mechanism without computing the relationship between asset prices and observable economic variables. For example, we estimate the score of the fundamentals from the stock return data by using the decomposition formula; thus we can know whether or not the level of the fundamentals has changed. This can provide useful information for investors and policy makers. We have presented one successful application of the covariance decomposition technique as a test of APT. Undoubtedly, however, more work needs to be done to enhance the linear multi-factor model.

Appendix: Brief Explanation of Factor Analysis

The appendix outlines the foundation and the estimating methods of factor analysis. Section 1 presents the basic concept of factor analysis, and Section 2 introduces factor analysis as it is applied to a correlation matrix. Section 3 discusses the solution by maximum likelihood estimation under the assumption of asymptotic normality, and Section 4 the solution by least squares estimation. Section 5 discusses the question of how to determine the right number of factors. Finally, Section 6 introduces estimation of factor scores.

1. Basic Formula of Factor Analysis

Asssumption I: Let unobservable stochastic vectors $\tilde{\delta}\equiv[\tilde{\delta}_1, \ldots, \tilde{\delta}_k]'$ $\tilde{\varepsilon}\equiv[\tilde{\varepsilon}_1, \ldots, \tilde{\varepsilon}_n]'$ have the following properties:

$$E(\tilde{\delta})=O, \tag{A1-1}$$

$$E(\tilde{\delta}\,\tilde{\delta}')=I, \tag{A1-2}$$

$$E(\tilde{\varepsilon})=O, \tag{A1-3}$$

$$E(\tilde{\varepsilon}\,\tilde{\varepsilon}')=\Psi, \tag{A1-4}$$

$$\psi\equiv\begin{bmatrix} \varphi_1 & & O \\ & \ddots & \\ O & & \varphi_n \end{bmatrix},$$

$$E(\tilde{\delta}\,\tilde{\varepsilon}')=E(\tilde{\varepsilon}\,\tilde{\delta}')=O. \tag{A1-5}$$

Definition: For an observable stochastic vector $\tilde{r} \equiv [\tilde{r}_1, \ldots, \tilde{r}_n]'$, there is a linear multi-factor model such that

$$\tilde{r} = \alpha + B\tilde{\delta} + \tilde{\varepsilon}, \tag{A1-6}$$

where $\alpha \equiv [\alpha_1, \ldots, \alpha_n]'$, $B \equiv [\beta_{ij}]$.

Computing the expectation of equation (A1-6), we get

$$E(\tilde{r}) = \alpha. \tag{A1-7}$$

Then, equation (A1-6) is

$$\tilde{r} = E(\tilde{r}) + B\tilde{\delta} + \tilde{\varepsilon} \tag{A1-8}$$

This is the vector representation of the return generating process given by equation(2) in the text.

The covariance matrix of $\tilde{r}$ is defined as

$$\Sigma \equiv E[\{\tilde{r} - E(\tilde{r})\}\{\tilde{r} - E(\tilde{r})\}'].$$

Replacing this with equation (A1-8) and assumption I, we get

$$\begin{aligned}
\Sigma &= E[B\tilde{\delta}+\tilde{\varepsilon})\,(B\tilde{\delta}+\tilde{\varepsilon})'] \\
&= E(B\tilde{\delta}\,\tilde{\delta}'B') + E(\tilde{\varepsilon}\,\tilde{\delta}'B') + E(B\tilde{\delta}\,\tilde{\varepsilon}') + E(\tilde{\varepsilon}\,\tilde{\varepsilon}') \\
&= BIB' + O + O + \Psi, \\
\therefore\quad \Sigma &= BB' + \Psi.
\end{aligned} \tag{A1-9}$$

This is the fundamental equation of factor analysis. Its interpretation is that, as a result of the decomposition of the observable covariance matrix Σ, factor loading (B) and the variance matrix of disturbances (Ψ) are computable in spite of the unobservability of $\tilde{\delta}$ and $\tilde{\varepsilon}$.

2. Factor Analysis Applied to a Correlation Matrix

Let components of a diagonal matrix D be the observable standard deviations of the components of a stochastic vector $\tilde{r}$, that is

$$D \equiv [\mathrm{diag}(\Sigma)]^{\frac{1}{2}} = \begin{bmatrix} \sigma_{r1} & & O \\ & \ddots & \\ O & & \sigma_{rn} \end{bmatrix}.$$

Then, the scale of $\tilde{r}$ can be transformed into $\tilde{r}_s$ such that

$$\tilde{r}_s = D^{-1}\{\tilde{r} - E(\tilde{r})\}, \tag{A2-1}$$

and the factor representation can be rewritten with $\tilde{r}_s$ as

$$\tilde{r}_s = D^{-1}B\tilde{\delta} + D^{-1}\tilde{\varepsilon} \tag{A2-2}$$

or

$$\tilde{r}_s = B_s\tilde{\delta} + \tilde{\varepsilon}_s \tag{A2-3}$$

where

$$B_s \equiv D^{-1}B, \quad \tilde{\varepsilon}_s \equiv D^{-1}\tilde{\varepsilon}.$$

In this case, Assumption I is partly replaced by the following;

$$E(\tilde{\varepsilon}_s) = O \tag{A2-4}$$

$$E(\tilde{\varepsilon}_s\tilde{\varepsilon}_s')=D^{-2}E(\tilde{\varepsilon}\,\tilde{\varepsilon}')=D^{-2}\psi=\psi_s, \qquad \text{(A2-5)}$$
$$E(\tilde{\delta}\,\tilde{\varepsilon}_s')=E(\tilde{\varepsilon}_s\tilde{\delta}')=O. \qquad \text{(A2-6)}$$

Then, the covariance matrix of the standardized stochastic vector $\tilde{r}_s$ is rewritten as

$$\begin{aligned}\Sigma_s &\equiv E[\{\tilde{r}_s-E(\tilde{r}_s)\}\{\tilde{r}_s-E(\tilde{r}_s)\}'] \\ &=E[\{D^{-1}(\tilde{r}_s-E(\tilde{r}_s))\}\{D^{-1}(\tilde{r}_s-E(\tilde{r}_s))\}'] \\ &=D^{-1}E[\{\tilde{r}-E(\tilde{r})\}\{\tilde{r}-E(\tilde{r})\}']D^{-1} \\ &=D^{-1}\Sigma D^{-1}. \end{aligned} \qquad \text{(A2-7)}$$

Equation(A2-7) shows that the covariance matrix Σ_s is equivalent to the correlation matrix when $\tilde{r}_s$ is standardized with mean zero and variance one. Here we get the fundamental equation of factor analysis applied to a correlation matrix by replacing equation (A1-9) with equation (A2-7)

$$\Sigma_s=B_sB_s'+\psi_s. \qquad \text{(A2-8)}$$

3. Conditions Required for Maximum Likelihood Estimation

Let us consider the solution of the fundamental equation when the number of factors k is known. To perform factor analysis, we assume the existence of the inverse matrix Σ_s^{-1} and let a solution set $(\hat{B}_s, \hat{\psi}_s)$ exist. To satisfy this condition, we set

Assumption II: $\Sigma_s > O$ (A3-1)

and, to guarantee a unique solution,

Assumption III: ψ_s of the fundamental equation (A2-8) is unique.

Maximum likelihood estimation is to find an estimate which maximizes the likelihood of the sample correlation matrix S by using T observations of r_s. We can find the maximum likelihood estimator if the factor model satisfies Assumptions I, II and III and the observations are generated from the normally distributed stochastic vector r_s. Subsequently, if factors $\tilde{\delta}$ and disturbances $\tilde{\varepsilon}$ are subject to multivariate normality, maximum likelihood estimation is appropriate.

However, the normality of return cannot be postulated against a linear multi-factor return generating process as in APT. Therefore, it is difficult to define the likelihood of the observation r_s, maximum likelihood factor analysis is not applicable to the APT model.

If we assume a fixed model of factor representation, we can define the conditional density function for the observation r_s. However, we still have the problem of a divergent likelihood function. In fact, the number of unknown parameters is so large that the first order condition for likelihood maximization shows only the point of infection (Anderson and Rubin 1956).

4. Least Squares Factor Analysis

We can adopt least squares estimation even if maximum likelihood estimation cannot be applied to some factor representations. First, consider the fit function which describes the estimation error of the sample correlation matrix. It is defined as

$$L(\Sigma_s, S) \equiv \tfrac{1}{2}\mathrm{tr}\{(S - \Sigma_s)W\}^2 \qquad \text{(A4-1)}$$

where W is a positive definite matrix. There are some candidates for W, such as I, D_s^{-2}. We set

Assumption IV: $W = S^{-1}$. (A4-2)

Under Assumption IV, minimization of the fit function given by equation (A4-1) is called weighted least squares estimation. There are two reasons for adopting weighted least squares estimation. First, under the condition that the diagonals of the population and sample correlation matrices are the same, the weighted least squares estimator is equal to the maximum likelihood estimator. Second, the idea of the fit function is common in the estimation of a factor score.

Now let us consider the weighted least squares estimators $\hat{\Sigma}_s$, $\hat{B}_N$, $\hat{\psi}_s$ which minimize the fit function under Assumptions I through IV,

$$L(\Sigma_s, S) = \tfrac{1}{2}\mathrm{tr}\{(S - \Sigma_s)S^{-1}\}^2. \qquad \text{(A4-3)}$$

The total differential of the fit function is

$$dL = -\mathrm{tr}\{S^{-1}(S - \Sigma_s)S^{-1}d\Sigma_s\}. \qquad \text{(A4-4)}$$

The first order conditions for minimizing the fit function given by equation (A4-3) are

$$\frac{\partial L}{\partial B_s} = -2S^{-1}(S - \Sigma_s)\Sigma_s^{-1}B_s = O, \qquad \text{(A4-5)}$$

$$\frac{\partial L}{\partial \psi_s} = -\mathrm{diag}\{S^{-1}(S - \Sigma_s)S^{-1}\} = O. \qquad \text{(A4-6)}$$

If we replace Assumption II with

Assumption II:′ $S > O$, (A4-7)

the solution of $\hat{B}_s$ is

$$(S - \Sigma_s)\Sigma_s^{-1}\hat{B}_s = O, \qquad \text{(A4-8)}$$

and the solution of $\hat{\Sigma}_s$ is

$$\mathrm{diag}\{S^{-1}(S - \hat{\Sigma}_s)S^{-1}\} = O, \qquad \text{(A4-9)}$$

which is equivalent to the solution of $\hat{\psi}_s$ such that

$$\hat{\psi}_s = \mathrm{diag}(S - \hat{B}_s\hat{B}_s'). \qquad \text{(A4-10)}$$

The above three solution equations are not independent so that it is imposible to estimate $\hat{B}_s$ and $\hat{\psi}_s$ simultaneously. Hence, we follow Joreskog (1967)'s algorithm in order to guarantee convergence to the solution.

First, set $\psi_s=\psi_{s0}$, and get

$$S^*=S^{-\frac{1}{2}}(S-\psi_{s0})S^{-\frac{1}{2}}. \tag{A4-11}$$

Now take the k largest eigenvalues of S^* and put them into a diagonal matrix D^*. If V^* is the orthonormal eigenvalue matrix of D^*, then

$$B_{s0}=S^{-\frac{1}{2}}V^*D^{*\frac{1}{2}}. \tag{A4-12}$$

Substituting this initial solution into equation (A4-10), and repeat the algorithm until the solution becomes convergent.

5. Determining the Right Number of Factors

In practice, the right number of factors is unknown, and n estimates must be computed. Thus we get n least squares estimates for factor loading matrix $\hat{B}_s$. The next step is to select the best estimate among these.

Unless the true factor representation is known, the test for determining the right number is also unknown. Even if the sample variance is asymptotically normally distributed, statistically significant covariances tend to indicate too many factors (Geweke and Singleton 1980).

Usually, several tests are applied to factor models, and the minimum number of factor representation is adopted. All tests require asymptotic normality of factors so that there is no one best test for factor models. In any case, we apply the chi-square test and two information criteria. The chi-square test is based on the fact that the ratio of the maximum likelihoods of the population and sample correlation matrices is asymptotically subject to a chi-square distribution. When $\Sigma_s=\hat{\Sigma}_s$, the maximum value of the logarithmic likelihood of S is

$$L_0=\frac{T-1}{2}\{\log|\hat{\Sigma}|+\mathrm{tr}(\hat{\Sigma}^{-1}S)\}. \tag{A5-1}$$

On the other hand, when $\Sigma_s=S$, the maximum value of the logarithmic likelihood of S is

$$L_1=\frac{T-1}{2}\{\log|S|+\mathrm{tr}(S^{-1}S)\}=\frac{T-1}{2}(\log|S|+n). \tag{A5-2}$$

Then the ratio of both likelihoods, which is equal to the difference of both logarithmic likelihoods, is

$$\chi_0{}^2\equiv\frac{T-1}{2}\cdot 2(L_0-L_1)=(T-1)\{\log|\hat{\Sigma}|-\log|S|+\mathrm{tr}(\hat{\Sigma}^{-1}S)-n\} \tag{A5-3}$$

and subject to a chi-square distribution with freedom of $\frac{1}{2}\{(n-k)^2-(n+k)\}$. If $\chi_0{}^2$ is larger than the value of the chi-square distribution at a given significance level, the null hypothesis that k factors are sufficient to represent the correlation is rejected. Hence, according to this likelihood ratio test, the factor model which meets this condition is considered adequate.

Popular information criteria, which improve the information in the maximum likelihood, are Akaike's Information Criterion (AIC) and Schwarz (1978)'s Bayesian criterion (SBC), which are respectively

$$AIC \equiv L_1 - k, \tag{A5-4}$$

and $$SBC \equiv L_1 - \frac{k}{2} \cdot \log T. \tag{A5-5}$$

The factor representation which minimizes these criterion values are selected. However, it is important to note that there is no objective standard by which a particular specification can be judged correct; the only requirement is that, for the model selected, it must be possible to define the likelihood of the sample correlation matrix.

6. Estimation of Factor Score

The estimators of $\hat{B}_s, \hat{\psi}_s, \hat{\Sigma}_s$ lead to the estimator for factor score vector $\tilde{\delta}$. We adopt the least squares estimator which minimizes the estimation error,

$$L(\delta, \hat{\delta}) = \frac{1}{n} \cdot E\{(\hat{\delta} - \delta)(\hat{\delta} - \delta)'\}. \tag{A6-1}$$

It is not difficult to find that the least squares estimator of $\tilde{\delta}$ is

$$\tilde{\delta} = B'\Sigma_s \cdot r. \tag{A6-2}$$

REFERENCES

Anderson, T. W. and Rubin. H., "Statistical Inference in Factor Analysis", *Proceeding of the Third Berkeley Symposium*, 5, 1956, pp.111-150.

Aoyama, Mamoru, "Risuku no Saihyōka ni tsuite—Wagakuni Kabushiki Shijō ni okeru Jisshō Kenkyū—", *Keizaigaku Kenkyū*, University of Tokyo, 22, 1979, pp.44-52.

———, "Fukakujitsusei to Senkō", in Katsunosuke Moroi and Takaaki Wakasugi, eds., *Gendai Keiei Zaimuron*, 1984, pp.21-41.

Black, F., Jensen, M. and Scholes, M., "The Capital Asset Pricing Model: Some Empirical Tests", in M. Jensen, ed., *Studies in the Theory of Capital Market*, New York: Praeger Publishers, 1972.

Blume, M. E. and Friend, I., "A New Look at the Capital Asset Pricing Model", *Journal of Finance*, 28-1, 1973, pp.19-33.

Brown, S. J. and Weinstein, M. I., "A New Approach to Testing Asset Pricing Models: The Bilinear Paradigm", *Journal of Finance*, 38-3, 1983, pp.711-743.

Chamberlain, G.,"Funds, Factors, and Diversification in Arbitrage Pricing Models", *Econometrica* 51-5, 1983, pp.1305-1323.

Chen, N. F., Roll, R. and Ross, S. A., "Economic Forces and the Stock Market", *Journal of Business*, 59-3, 1986, pp.383-403.

Cho, D. C., "On Testing the Arbitrage Pricing Theory: Inter-battery Factor Analysis", *Journal of Finance*, 39-5, 1984, pp.1485-1502.

———, Elton, E. J. and Gruber, M. J., "On the Robustness of the Roll and Ross Arbitrage Pricing Theory", *Journal of Financial and Quantitative Analysis*, 19-1, 1984, pp.1-28.

Dhrymes, P. J., Friend, I. and Gultekin, N. B., "A Critical Re-examination of the Empirical Evidence on the Arbitrage Pricing Theory", *Journal of Finance*, 39-2, 1984, pp-323-346.

———, Friend, I., Gultekin, M. N. and Gultekin, N. B., "New Tests of the APT and their Implication", *Journal of Finance*, 40-3, 1985, pp.659-674.

Elton, E. J. and Gruber, M. J., *Modern Portofolio Theory and Investment Analysis*, 2nd edition, New York: John Wiley & Sons, 1981.

Fama, E. F.. and Macbeth, J. D., "Risk, Return and Equilibrium: Empirical Tests", *Journal of Political Economy*, 71, 1973, pp.607-636.

Geweke, J. F. and Singleton, K. J., "Interpreting the Likelihood Ratio Statistic in Factor Models when Sample Size is Small", *Journal of the American Statistical Association*, 75, 1980, pp.133-137.

Hamao, Y., "Pōtoforio Riron APT—Kinyu Shisan Unyō no Shakudo ni", *Nihon Keizai Shinbun Keizai Kyōshitsu*, August 22, 1986.

Horimoto, Saburo, "Wagakuni ni okeru Saitei Hyōka Riron (APT) no Kenshō", Hikone Ronsō, Shiga University, 231, 1986, pp.43-54.

Huberman, G., "A Simple Approach to Arbitrage Pricing Theory", *Journal of Economic Theory*, 28, 1982, pp.183-191.

Joreskog, K. G., "Some Contributions to Maximum Likelihood Factor Analysis", *Psychometrika*, 32, 1967, pp.443-482.

———, "Factor Analysis by Least-squares and Maximum Likelihood Methods", in Enslein K., et al., eds., *Statistical Methods for Digital Computers*, New York: John Wiley & Sons, 1977, pp.125-153.

Kobayashi, Takao, "Shijō no Kinkō to Shōken Kakaku", in Katsunosuke Moroi and Takaaki Wakasugi, eds., *Gendai Keiei Zaimuron*, 1984, pp.42-69.

Konya, Fumiko, "Kabushiki Shijō ni okeru Tōshika Kōdō to Shijō Kōritsu", *Keisoku Shitsu Tekunikaru Pēpā*, Japan Securities Research Institute, No.44, 1978, pp.83-94.

Kryzanowski, L. and To, M. C., "General Factor Models and the Structure of Security Returns", *Journal of Financial and Quantitative Analysis*, 18-1, 1983, pp.31-52.

Lawley, D. N. and Maxwell, A. E., *Factor Analysis as a Statistical Method*, New York: Macmillan, 1971.

Maru, Junko and Royama, Shoichi, "Kabushiki Shijō ni okeru Shūeki to Kiken", *Keisoku Shitsu Tekunikaru Pēpā*, Japan Securities Research Institute, No.29, 1974, pp.1-43.

Mossin, Jan, *The Economic Efficiency of Financial Markets*, Toronto: D. C. Heath and Company, 1977.

Okamoto, Tadashi, "Inshi Bunseki no Kiso", Nikka Giren Shuppansha, 1986.

Roll, R., "A Critique of the Asset Pricing Theory's Tests", *Journal of Financial Economics*, 4, 1977, pp.129-176.

Rosenberg, B. and Guy, J., "Prediction of a Beta from Investment Fundamentals", *Financial Analysis Journal*, I, 1976 pp.3-15.

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

———, "Mutual Fund Separation in Financial Theory, the Separating Distribution", *Journal of Economic Theory*, 15, 1978, pp.254-286.

Sakakibara, Shigeki, "CAPM no Saikentō to Kigyō Kibo Kōka", *Kokumin Keizai Zasshi*, 147(5), 1983, pp.88-112.

Sakakibara, Shigeki, *Gendai Zaimu Riron*, Chikuma Shobō, 1986.

Sato, S., "Nippon ni okeru β Risuku no Jisshō Kenkyū", *Keizai Riron*, Wakayama University, 199, 1984, pp.53-88.

Schwarz, G., "Estimating the Dimension of a Model", *Annual Statistics*, 6, 1978, pp.461-464.

Shanken, J., "The Arbitrage Pricing Theory: Is it Testable?", *Journal of Finance*, 37-5, 1982, pp.1129-1140.

———, "Multivariate Proxies and Asset Pricing Relations: Living with the Roll Critique", *Journal of Financial Economics*, 18, 1987, pp.91-110.

Sharpe, W., *Investments*, 3rd edition, New York: Prentice Hall, 1986.

Trzcinka, C., "On the Number of Factors in the Arbitrage Pricing Model", *Journal of Finance*, 41-2, 1986, pp.347-368.