A Reexamination of Homogeneous Stock Grouping in the Context of the APT: An Application of Discriminant Analysis

C.R. Narayanaswamy*
University of Idaho

Investment professionals traditionally classify stocks into groups such as growth, stable, and cyclical stocks and form portfolios on the basis of these classifications. Farrell (1974, 1975) shows how to form such homogeneous groups by analyzing the extramarket returns of stocks. Farrell's grouping methodology is cited widely and frequently applied in practice. Yet few studies have examined whether Farrell's classification of stocks is consistent with asset pricing models. This study validates Farrell's methodology using multiple risk measures that are relevant in the context of the arbitrage pricing theory.

INTRODUCTION

Classifying stocks into homogenous groups has important implications for portfolio management tasks such as portfolio construction, performance measurement, and portfolio revision (Farrell, 1974 and 1975). The stock groups identified by Farrell and the methodology he uses often are cited in finance literature and applied in the investments industry.[1] The prevalence of the technique and its application merits a closer look, particularly from the perspective of an asset pricing model.

Classifications of stocks are relevant for investment purposes only if the stocks in different groups have different risk characteristics. Asset pricing models help quantify and price the risk characteristics of stocks. This paper investigates whether the risk profiles consistent with the arbitrage pricing model yield classifications similar to the groups identified by Farrell.

Classification techniques group similar objects. An important requirement of a classification scheme is that it considers the salient properties of objects. Such considerations may vary. Therefore, the characteristics deemed relevant and important depend on the task at hand; only the relevant characteristics can provide a meaningful profile of an object.

Asset pricing models postulate that the relevant risk of an asset is its nondiversifiable risk. The relevant risk is caused by the asset's sensitivity to pervasive factors. The remaining risk is of little consequence, as it does not affect a security's price. According to the CAPM the relevant risk is the systematic risk related to a common factor (market portfolio). This risk is quantified by the beta coefficient. The arbitrage pricing model is based on the assumption that returns on risky assets are linearly related to multiple factors

* I wish to thank Ravi Jagannathan, Sudi Sudarsanam, and two anonymous referees for their comments and suggestions. Any remaining errors are my responsibility.

[1] For example, A.G. Becker, Inc. used the Farrell methodology to form homogeneous stock groups. This firm was acquired by the SEI Corporation in 1983. SEI continues to form these stock groups and updates them every six months (Cottle, Murray, and Block, 1988).

0747-5535/96/1400-0003 $02.50

and that an asset's risk is reflected in the multifactor beta coefficients. Because asset pricing models quantify the relevant risk, it seems plausible that these quantities can be used as their risk profile.

Although the capital asset pricing model postulates that the relevant risk of a security is reflected in its beta coefficient, two celebrated studies find nonmarket risk to be significant. King (1966) documents the presence of substantial covariation among stock returns attributable to industry factors, whereas Farrell (1974) shows the possibility of the existence of broader than industry factors in addition to the market factor.

King's study consists of 63 stocks belonging to six different industries defined according to standard industrial code. King shows that residuals obtained from applying the market model have substantial covariation for stocks within an industry. This finding leads him to conclude that there are industry-related factors causing significant covariation of stock returns.

The usefulness of classifying stocks on the basis of the standard industrial code, however, has been questioned by some researchers. Cohen and Pogue (1967) show that the performance of portfolio selection models does not improve even after industry indexes are considered. Morgan (1977) also demonstrates that there is little to be gained using industry classification. Elton and Gruber (1970,1971) argue that pseudoindustry classifications based on the correlation structure of security returns are superior to groupings based on standard industry classifications.

Farrell (1974) investigates empirically the presence of broader than industry groups. Farrell identifies stock groups considering the collinearity characteristics of stock returns. He groups stocks that are significantly correlated and are generally independent of stocks in other groups. Farrell chooses a sample of 100 stocks in such a manner that the industry effect is neutralized and removes the effect of the market factor by considering the residuals of the regression of stock returns on the market factor. The correlation of these residuals leads him to conclude that stocks naturally cluster into growth, cyclical, stable, and oil stock groups. Martin and Klemkosky (1976) also show that significant extramarket covariation of asset returns can be attributed to groupings of stocks based on the scheme suggested by Farrell.

Results obtained by Farrell are important for our study. Because the industry effects are minimized through the sample selection process in Farrell's study, the correlations of residuals must be caused by factors that are broader than industry. These factors must be pervasive or economy wide. As the correlations are due to pervasive factors, a better asset pricing model should account for these factors. The risk of a stock as profiled by the asset pricing model should be useful in forming homogenous stock groups.

The arbitrage pricing model is based on the assumption that the return on a risky asset is linearly related to a few pervasive factors. If the return-generating process is specified reasonably well, the risk profile of a stock can be obtained from the regression of the stock's return on these factors. This is accomplished in Berry, Burmeister, and McElroy (1988). They classify stocks into seven groups, form equally weighted portfolios of securities in each of these groups, and regress the returns on these portfolios of stocks on five macroeconomic factors. They use the regression coefficients as the risk profiles of the stock groups. Although Berry, Burmeister, and McElroy (1988) provide some insight into the risk profiles of stocks, they do not show whether these profiles are statistically distinguishable.

Fogler, John, and Tipton (1981) provide an economic meaning to the pseudoindustries identified by Elton and Gruber (1973) by examining the canonical correlations between economic factors and factors representing pseudoindustries. In order to model the return-generating process with economic factors, Fogler, John, and Tipton use a three factor model;

- A stock market index;
- A U.S. government bond index; and
- A corporate bond index

are the three pervasive economic factors. They obtain pseudoindustry factors by extracting the dominant principal components from the returns of the stocks in their sample. These pseudoindustry factors also can be considered as the factors used for pricing risky assets in the arbitrage pricing model. (See Chamberlain and Rothschild, 1983.) By a judicious choice of sample, Fogler, John, and Tipton determine the number of principal components that are significant and suitable for further analysis. They choose a sample of stocks representing all four groups of the Farrell sample. After obtaining the risk profiles of the stocks (by regressing the stock returns on the economic factors), they visually inspect for a pattern in the risk profiles of the stocks. They look for a similar pattern in the correlation coefficients of the stock returns with the dominant principal components. The number of principal components they retain for further analysis is the number that provides the identifiable pattern of stocks consistent with Farrell's classification. Although the Fogler, John, and Tipton study examines the relationship between specified economic factors and factors obtained from principal component analysis, it also shows that Farrell's classifications are consistent with the risk profiles obtained from both specified and factor analytic methods. This conclusion, however, is based on visual inspection.

In contrast, Baer (1993) classifies stocks by applying cluster analysis. He obtains risk profiles of stocks using a five factor model comprising

- Unexpected increase in default risk;
- Unexpected change in the term structure;
- Unexpected deflation;
- Unexpected downward revision of anticipated sales growth; and
- A residual factor.

Baer classifies stocks on the basis of their risk profiles using cluster analysis. These classifications differ considerably from those obtained by Farrell, although Baer uses the same sample of stocks. The differences may be due partly to the statistical methodology applied by Baer. Cluster analysis groups objects based on the Euclidean distance between clusters' mid-points. It ignores the relative importance of the regression coefficients: the stocks' sensitivities to the factors. Therefore, Baer's conclusion that classifying stocks based on multibeta coefficients is inconsistent with Farrell's results may not be correct.

Classifications obtained from multiple risk measures may be sensitive to the statistical methods used. The present study addresses this issue. Instead of using cluster analysis, we use multiple discriminant analysis. Multiple discriminant analysis has been applied extensively in bankruptcy studies for grouping companies by financial ratios. We apply this technique on the stocks' risk measures. In contrast to Fogler, John, and Tipton (1981) whose conclusions are based on visual inspection, we use multiple discriminant analysis for checking the classifications. We show that stock groupings derived from the Farrell methodology are consistent with multiple risk measures if multiple discriminant analysis is used. Such consistency in classification may be useful for security analysis and portfolio management.

METHODOLOGY

THE ARBITRAGE PRICING MODEL

This study uses the arbitrage pricing theory to obtain risk profiles of stocks. The APT is based on the assumption that the return on a security is linearly related to a few pervasive factors. This can be expressed as

$$(1)\ r_i = E_i + b_{i1} f_1 + b_{i2} f_2 \ldots + b_{ik} f_k + e_i$$

where:

r_i	=	The return on security i;
E_i	=	The expected return on security i;
f_j	=	The unexpected change in the value of factor j;
b_{ij}	=	The sensitivity of the return on security i to factor j; and
e_i	=	The residual term, the return idiosyncratic to security i.

(r_i, f_j, and e_i are random variables.) It is assumed that the factors are not correlated among themselves or with the idiosyncratic terms. If the idiosyncratic terms of securities are not correlated, the model is supposed to have a strict factor structure; if not, the model is supposed to have an approximate factor structure. For the strict factor structure Ross (1976) shows that in the absence of arbitrage opportunities, the equilibrium relationship can be expressed as:

$$(2)\ E_i = r_f + b_{i1} \lambda_1 + b_{i2} \lambda_2 \ldots + b_{ik} \lambda_k$$

where:

r_f	=	The risk free return; and
λ_j	=	The market price for one unit of risk of factor j.

Chamberlain and Rothschild (1983) show that the relationship in equation (2) is valid for the approximate factor structure also if large numbers of assets are traded in the market.

STATISTICAL TECHNIQUE USED FOR CLASSIFICATION

The main purpose of this study is to examine whether grouping stocks on the basis of risk characteristics is justified. The same sample considered by Farrell (1974, 1975) is used here also. Classifications of these stocks into the four groups suggested by Farrell are used as the *a priori* group memberships. The chosen statistical technique should help us check whether the groups are homogenous within, but sufficiently differentiated from one another. The most appropriate technique for this purpose is multiple discriminant analysis.[2]

Multiple discriminant analysis helps determine whether relationships exist between the quantitative variables and a classification variable of objects. For example, in the present study, stocks are described by means of their risk profile as reflected by the factor sensitivities (quantitative variables). These stocks also are classified into various *a priori* groups (qualitative variables). Discriminant analysis develops a (discriminant) function of the quantitative variables so that each object (stock) is given a score. These scores are used for classifying the objects. The discriminant function is developed in such a manner that the separation of groups is achieved while preserving the *a priori* membership within the groups.[3]

The effectiveness of multiple discriminant analysis in classifying stocks into respective groups is indicated by the percentage of correct classifications. This measure could be biased when all stocks in a sample are used for deriving the discriminant function because of the tendency of multiple discriminant analysis to overfit the data. Lachenbruch (1967) shows that almost unbiased estimates of the probability of correct classifications can be obtained by developing the discriminant function using all observations except one. The omitted observation is used for checking the classification accuracy of the model. This procedure is repeated for all observations. The proportion of objects correctly classified, Lachenbruch demonstrates, is an almost unbiased estimate of the classification accuracy of the model. This jackknife approach is hereafter referred to as the *Lachenbruch holdout method*. Using all the stocks to develop the discriminant function is referred to as the *original sample method* hereafter.[4]

---

[2] Multiple discriminant analysis was introduced in financial research by Altman (1968) in his study of bankrupt firms. He investigates whether firms in danger of failing can be statistically discriminated on the basis of financial ratios. Altman argues that financial ratios can be used to obtain the relevant profiles of firms to which discriminant analysis can be applied in order to predict imminent bankruptcy. The analysis in this study is similar to Altman's. Instead of using financial ratios as the profile of a firm, this study uses the risk characteristics of stocks obtained by an asset pricing model. Multiple discriminant analysis is applied to the risk profiles in order to verify whether the risk profile of the stocks and their *a priori* classifications are consistent.

[3] Baer (1993) applies cluster analysis to the (multidimensional) risk profiles. The application of cluster analysis to multidimensional risk profile has a drawback. Cluster analysis measures the Euclidean distance between clusters' mid-points. The relative weights given to the factor sensitivities are the same. For example, if the Chen, Roll, and Ross (1986) model is used, the sensitivities to monthly production, change in expected inflation, unanticipated inflation, risk premium, and term structure have the same weights. In reality, however, the market may place different importance on these factors. Further, these factor sensitivities are affected by scaling. In contrast, Farrell (1974, 1975) uses only one measure—correlation of extramarket returns—to classify stocks. Therefore, results from Farrell's study do not have this drawback.

[4] Sudarsanam and Taffler (1985) use the original sample method and Lachenbruch holdout method to assess the classificatory power of the models with reference to U.K. capital markets.

ESTIMATION OF RISK MEASURES

The arbitrage pricing theory is based on the assumption that the relevant risk of an asset is due to the sensitivity of its returns to pervasive or economy-wide factors. These sensitivities—the b_{ij}s in equation (1)—can be obtained by assuming a return-generating process that requires specification of macroeconomic factors as in Chen, Roll, and Ross (1986) or through factor analytic techniques as shown by Roll and Ross (1980). Application of the specified factors approach requires knowledge of what the pervasive economic factors are.

THE SPECIFIED FACTORS APPROACH

To apply the specified factors approach, this study uses two return-generating processes popular in the APT literature: the five factor model proposed by Chen, Roll, and Ross (1986) and the two factor model suggested by Chen, Grundy, and Stambaugh (1990).

THE CHEN, ROLL, AND ROSS MODEL

Chen, Roll, and Ross (1986) model stock returns as a linear function of five macroeconomic variables. They show these variables are significant in explaining expected returns on stocks. The return on stocks according to this model is

$$(3)\ r_{i,t} = a_i + b_{i,MP}MP_t + b_{i,DEI}DEI_t + b_{i,UI}UI_t + b_{i,UPR}UPR_t + b_{i,UTS}UTS_t + e_{i,t}$$

where:

$r_{i,t}$	=	The return on stock i;
MP_t	=	The growth in industrial production;
DEI_t	=	The change in expected inflation;
UI_t	=	The unexpected inflation;
UPR_t	=	The risk premium;
UTS_t	=	The term structure; and
$e_{i,t}$	=	The residual error term.

t is the time subscript. The risk premium is the return on bonds rated Baa and under by Moody's minus the return on long-term government bonds. The term structure is the return on long-term government bonds minus the return on Treasury bills. The regression coefficients $b_{i,MP}$, $b_{i,DEI}$, $b_{i,UI}$, $b_{i,UPR}$, and $b_{i,UTS}$ are the sensitivities of stock i's return to the respective factors. These coefficients (multifactor betas) are used as the risk profile of each stock in subsequent statistical analysis.

THE CHEN, GRUNDY, AND STAMBAUGH MODEL

The second return-generating process considered is the one proposed by Chen, Grundy, and Stambaugh (1990). Chen, Grundy, and Stambaugh consider a two factor model in their empirical analysis of dividend yield effects:

$$(4)\ r_{i,t} = a_i + b_{i,RVW}\ RVW_t + b_{i,PREM}\ PREM_t + e_{i,t}$$

where:

$r_{i,t}$ = The return on stock i in excess of the Treasury bill rate;
RVW_t = The excess return on the value-weighted portfolio of stocks on the New York Stock Exchange in excess of the Treasury bill rate;
$PREM_t$ = The return on a portfolio of bonds rated by Moody's as below Baa in excess of that of the U.S. government bond; and
$e_{i,t}$ = The residual error term.

PREM is the same as risk premium in the Chen, Roll, and Ross model. The difference between the two factor (Chen, Grundy, and Stambaugh) model and the five factor (Chen, Roll, and Ross) model is the other variables. Chen, Roll, and Ross model the return on stocks using (nonequity) macroeconomic variables. The Chen, Grundy, and Stambaugh model replaces monthly production, factors related to inflation, and the term structure by the (equity) variable RVW. Because Chen, Grundy, and Stambaugh find their two factor model to have substantial explanatory power regarding cross-sectional differences in expected returns on stocks, their model seems to be a viable alternative to the five factor model of Chen, Roll, and Ross. Therefore, the performance of the multifactor betas obtained from the Chen, Grundy, and Stambaugh model also is examined in this study.

The two factor return-generating process (Chen, Grundy, and Stambaugh) we use is similar to the return-generating process used by Fogler, John, and Tipton (1981) also. The three factors used by Fogler, John, and Tipton are excess returns on the CRSP value-weighted index, excess returns on U.S. government bonds, and excess returns on Aa quality corporate bonds. The first factor, excess returns on the CRSP value-weighted index, is common to both the Chen, Grundy, and Stambaugh and the Fogler, John, and Tipton models. The second factor (PREM) of the Chen, Grundy, and Stambaugh model is somewhat like the second and third factors in the Fogler, John, and Tipton model.

FACTOR ANALYTIC APPROACHES

As an alternative to specifying the economic factors and estimating the multibeta coefficients, many researchers analyze the covariance matrix of security returns. Chamberlain and Rothschild (1983) show that the maximum likelihood method of factor analysis or principal component analysis can be used for this purpose .

Using the return-generating process described in equation (1), the covariance matrix of security returns can be written as

$$(5)\ \Sigma = BB' + \psi$$

where:

Σ = The $(n \times n)$ covariance matrix of security returns;
B = The $(n \times k)$ matrix of factor sensitivities;
ψ = The $(n \times n)$ covariance matrix of the idiosyncratic returns; and
n = The number of securities.

Chamberlain and Rothschild (1983) show that, in a large asset market (that is, as $n \rightarrow \infty$),

(6) $P = B\,L$

where:

P	=	The matrix of the first k eigenvectors of Σ; and
L	=	A $(k \times k)$ orthogonal transformation matrix.

Chamberlain and Rothschild's results imply, for small universes of securities, maximum likelihood factor analysis is preferable as it attempts to separate the ψ matrix, whereas principal component analysis does not. Connor and Korajczyk (1988) and Shukla and Trzcinka (1990), however, show that principal component analysis and the maximum likelihood factor analysis provide comparable results. Therefore, in this study maximum likelihood factor analysis and principal component analysis are applied to estimate the factor sensitivities.

One of the problems encountered in applying factor analytic methods to estimate the multibetas is in specifying the number of factors. Although the APT makes fewer assumptions compared to the CAPM, the APT does not specify the number of pervasive factors. In empirical studies, however, most often we find specifications of five factors. Therefore, in this study the five factor model is assumed for estimating factor sensitivities through the factor analytic approaches.

THE DATA

SPECIFIED FACTORS

Following Chen, Roll, and Ross, the five economic state variables used in the first return-generating process are monthly growth in industrial production, change in expected inflation, unexpected inflation, risk premium, and term structure. Data on the Treasury bill rate, rate on long-term government bonds, and returns on bonds rated Baa and under are obtained from Ibbotson (1987).

Consistent with Chen, Roll, and Ross, the monthly growth rate in production in month t is defined as the natural logarithm of industrial production in month t minus the natural logarithm of industrial production in month t - 1. As suggested in Chen, Roll, and Ross, the monthly growth rate in production is led a month to make this variable contemporaneous with other series. Data for industrial production are obtained from the *Survey of Current Business* (1969) and the *Federal Reserve Bulletin* (1970). These publications have statistics of the monthly industrial production with 1957 as the base year. For 1961 through 1969—the time period chosen for the present study—the statistics on industrial production are compiled using 1957 as the base year. Gittings (1991) shows that economic time series are highly susceptible to rounding errors caused by periodic revisions of the base. Such revisions can distort the pattern and timing of changes in the index numbers that are preserved in the original data. For this reason, we use the original data for statistical analysis.

The time series for unexpected inflation and changes in expected inflation are constructed using the procedures adopted in Chen, Roll, and Ross. Unexpected inflation is defined as actual inflation in excess of expected inflation. Actual inflation for the time period is computed from the Consumer Price Index figures obtained from the *Survey of Current Business* (1969) and the *Federal Reserve Bulletin* (1970). Expected inflation must be estimated, however.

Fama and Gibbons (1984) compare several categories of models (time series models, interest rate models, and the Livingston surveys) that forecast inflation. They conclude that interest rate models are superior to time series models and the Livingston surveys. The interest rate models they consider require forecasts of the real rate of interest. They forecast the real rate using several methods. In one method they use the 12 month moving average of the realized real rates of interest as an estimate of the expected real interest rate for the next month. Fama and Gibbons demonstrate that this series is comparable to the estimates obtained from more sophisticated models. Therefore, we use the 12 month moving average for estimating the expected real rates of interest. The estimates of real rates of interest are subtracted from the monthly nominal rate on Treasury bills to construct the series on expected inflation. The actual inflation rate minus the expected inflation serves as the estimate of the unexpected inflation rate.

DATA ON STOCKS

Farrell (1974, 1975) carefully chooses the sample in his study so that the sample meets several important criteria. The firms whose stocks were chosen are all large (in the *Fortune* 500 list); the stocks are traded on the New York Stock Exchange; and the sample is chosen in such a manner that the industry effects are neutralized. All the requirements imposed by Farrell are equally applicable to this study. The same sample of stocks and the same time period considered by Farrell are used in this study. We consider only 85 of the 100 stocks used by Farrell for which monthly returns for the entire relevant time period (January 1961 to December 1969) could be obtained.

There are other reasons to use the same sample and time period; first, results of the Farrell study have been available to researchers and investment analysts for several years. Second, the *a priori* classifications of stocks are necessary for developing the discriminant function. These classifications have been done by Farrell. Further, other researchers—Baer (1993) and Fogler, John, and Tipton (1981)—also have used the Farrell sample in investigating stock groupings using the APT.

EMPIRICAL RESULTS

Our results are summarized in Table 1 and Table 2. Both tables report the consistency of *a priori* classifications of stocks and classifications of multiple discriminant analysis on the risk profiles of stocks.[5] The classificatory power of multiple discriminant analysis is tested by two methods. In the first (original sample method), all 85 stocks in the sam-

[5] The DISCRIM procedure of SAS, Version 6, is used for multiple discriminant analysis.

Table 1-Classification Summary (Original Sample)

Method	Growth	Stable	Percent of Stocks Classified Cyclical	Oil	Total
Growth					
A	76.19	14.29	4.76	4.76	100
B	76.19	9.52	9.52	4.76	100
C	85.71	14.29	0.00	0.00	100
D	100.00	0.00	0.00	0.00	100
Stable					
A	8.70	78.26	8.70	4.35	100
B	4.35	30.43	21.74	43.48	100
C	0.00	100.00	0.00	0.00	100
D	0.00	100.00	0.00	0.00	100
Cyclical					
A	6.06	30.30	60.61	3.03	100
B	6.06	15.15	66.67	12.12	100
C	3.03	0.00	96.97	0.00	100
D	3.03	0.00	96.97	0.00	100
Oil					
A	0.00	0.00	0.00	100.00	100
B	0.00	25.00	0.00	75.00	100
C	0.00	0.00	0.00	100.00	100
D	0.00	0.00	0.00	100.00	100

Method A is based on the five factor model suggested by Chen, Roll, and Ross (1986). Method B uses the two factor model of Chen, Grundy, and Stambaugh (1990). Method C uses principal component analysis. Method D uses maximum likelihood factor analysis. Figures in each row are the percent of stocks, classified *a priori* as belonging to a particular group, assigned to the various groups based on results from multiple discriminant analysis

ple are used to develop the discriminant function. In the second (Lachenbruch holdout method), one stock is removed from the sample for developing the discriminant function. The stock that is removed later is used as the holdout sample and classified using the discriminant function. This procedure is repeated for all 85 stocks to check the classificatory power of multiple discriminant analysis.

The *a priori* classification of each stock in Tables 1 and 2 is based on the results obtained by Farrell (1974, 1975). Figures in each row indicate the assignment of stocks to the various groups based on the results from multiple discriminant analysis. Perfect consistency between *a priori* classification and the later assignment is indicated if all the stocks from each initial group are assigned to the same group. Method A represents the five factor model of Chen, Roll, and Ross. Method B is the two factor model of Chen, Grundy, and Stambaugh. Method C represents principal component analysis, and Method D uses maximum likelihood factor analysis.

Table 1 indicates that in the original sample method the risk profiles obtained by the five factor model of Chen, Roll, and Ross generally outperform those of Chen, Grundy, and Stambaugh. The percent of correct classifications by the Chen, Roll, and Ross model is 76.19 percent for growth stocks, 78.26 percent for stable stocks, 60.61 percent for cyclical stocks, and 100 percent for oil stocks. The Chen, Roll, and Ross model does better than the Chen, Grundy, and Stambaugh model in identifying stable and oil stocks. It performs as well as the Chen, Grundy, and Stambaugh model in classifying growth

Table 2-Jackknifed Classification Summary (Lachenbruch Sample)

Method	Percent of Stocks Classified				
	Growth	Stable	Cyclical	Oil	Total
Growth					
A	57.14	19.05	19.05	4.76	100
B	71.42	14.29	14.29	0.00	100
C	85.71	14.29	0.00	0.00	100
D	90.48	9.52	0.00	0.00	100
Stable					
A	13.04	52.17	30.43	4.76	100
B	4.35	30.43	21.74	43.48	100
C	13.04	86.96	0.00	0.00	100
D	8.70	86.96	4.35	0.00	100
Cyclical					
A	18.18	30.30	48.48	3.03	100
B	9.09	15.15	63.64	12.12	100
C	9.09	0.00	90.91	0.00	100
D	6.06	0.00	93.94	0.00	100
Oil					
A	25.00	12.50	12.50	50.00	100
B	0.00	25.00	0.00	75.00	100
C	0.00	12.50	50.00	37.50	100
D	0.00	0.00	25.00	75.00	100

Method A is based on the five factor model suggested by Chen, Roll, and Ross (1986). Method B uses the two factor model of Chen, Grundy, and Stambaugh (1990). Method C uses principal component analysis. Method D uses maximum likelihood factor analysis. Figures in each row are the percent of stocks, classified *a priori* as belonging to a particular group, assigned to various groups based on the results from multiple discriminant analysis

stocks, but is inferior in classifying cyclical stocks. Classifications are best when risk profiles obtained from maximum likelihood factor analysis and principal component analysis are used. The percentage of correct classification is almost 100 percent when these two methods are used.

When the Lachenbruch holdout technique is applied, however, results are less impressive. These results, as shown in Table 2, indicate that the classificatory power based on the profiles of stocks obtained from the application of the Chen, Roll, and Ross model is more sensitive to sample selection than is the Chen, Grundy, and Stambaugh model. The proportion of correct classification of stocks drops from 76.19 percent to 57.14 percent for growth stocks, 78.26 percent to 52.17 percent for stable stocks, 60.61 percent to 48.48 percent for cyclical stocks, and 100 percent to 50 percent for oil stocks. A slight decrease in classificatory power can be observed for the Chen, Grundy, and Stambaugh model also. The performance of factor analytic models is much better than the specified factors for the Lachenbruch holdout technique. For example, when maximum likelihood factor analysis is applied, the overall performance declines about 8 percent. For principal component analysis the drop is about 14 percent.

The odds of assigning a stock to a group at random (as there are four groups) is 25 percent. The overall correct classifications in the original sample method for the Chen, Roll, and Ross, Chen, Grundy, and Stambaugh, maximum likelihood factor analysis, and principal component analysis are 73 percent, 60 percent, 99 percent, and 99 percent, respectively. In the Lachenbruch holdout method the corresponding figures are 51.82 percent for the Chen, Roll, and Ross model, 57.65 percent for the Chen, Grundy, and Stambaugh model, 91 percent for maximum likelihood factor analysis, and 85 percent for principal component analysis. These results indicate that the risk profiles of stocks consistent with the APT are useful for classifying stocks into homogeneous groups. The growth, stable, cyclical, and oil groups may proxy for factors in the APT. The multifactor betas obtained from the application of factor analytic methods are better suited for such classifications. The Farrell methodology and the factor analytic methods are closely related, as they use the correlation matrix of returns. This could be a possible reason for their superior performance. The multifactor betas obtained from the Chen, Grundy, and Stambaugh model appear to be more reliable than the results of the Chen, Roll, and Ross model.

CONCLUSION

Farrell (1974, 1975) suggests an objective method for identifying stocks as growth, stable, or cyclical stocks. He also shows how mutual fund performance can be grossly distorted using only market-related risk (beta) and indicates how performance measurement can be improved using indexes formed on the classification of stocks.

The method suggested by Farrell for identifying stocks as cyclical, income, or growth stocks now is used widely in the industry. Stock indexes based on such classifications have found widespread practical applications, particularly for portfolio construction and performance evaluation. Classification of stocks is relevant for investment purposes only if the stocks in different groups have different risk characteristics. Fogler, John, and Tipton (1981) and Baer (1993) investigate whether Farrell's classifications are supported with multiple risk measures based on the arbitrage pricing model. In contrast to these studies, we demonstrate—using an objective technique—that the risk profiles of stocks that are relevant in the context of the arbitrage pricing theory provide a taxonomy that validates Farrell's grouping methodology.

REFERENCES

1. Altman, Edward, "Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy," *Journal of Finance*, 23 (September 1968), pp. 589-609.

2. Baer, James D., "An Empirical Investigation of Risk Classes: Are Common Proxies Valid?" *Quarterly Review of Economics and Finance* (Spring 1993), pp. 33-51.

3. Berry, Michael A., Edwin Burmeister, and Marjorie B. McElroy, "Sorting Out Risks Using Known APT Factors," *Financial Analysts Journal* (March-April 1988), pp. 29-42.

4. Chamberlain, Gary, and Michael Rothschild, "Arbitrage and Mean-Variance Analysis on Large Asset Markets," *Econometrica*, 51 (September 1983), pp. 1281-1301.

5. Chen, Nai-Fu, Richard R. Roll, and Stephen A. Ross, "Economic Factors and the APT," *Journal of Business*, 59 (July 1986), pp. 383-403.

6. Chen, Nai-Fu, Bruce Grundy, and Robert F. Stambaugh, "Changing Risk, Changing Risk Premiums, and Dividend Yield Effects," *Journal of Business*, 63, no. 1, part 2 (January 1990), pp. S51-S70.

7. Cohen, Kalman, and Jerry Pogue, "An Empirical Evaluation of Alternative Portfolio Selection Models," *Journal of Business*, 40 (April 1967), pp. 166-93.

8. Connor, Gregory, and R.A. Korajczyk, "Risk and Return in an Equilibrium APT: Application of a New Test Methodology," *Journal of Financial Economics*, 21 (September 1988), pp. 255-289.

9. Cottle, Sidney, Roger F. Murray, and Frank E. Block with the collaboration of Martin L. Leibowitz, *Graham and Dodd's Security Analysis*, fifth edition (New York: McGraw-Hill Book Company, 1988), pp. 88-89.

10. Elton, Edwin J., and Martin J. Gruber, "Estimating the Dependent Structure of Share Prices—Implications for Portfolio Selection," *Journal of Finance*, 28 (December 1973), pp. 1203-1232.

11. Elton, Edwin J., and Martin J. Gruber, "Improved Forecasting Through the Design of Homogenous Groups," *Journal of Business*, 44 (October 1971), pp. 432-450.

12. Elton, Edwin J., and Martin J. Gruber, "Homogenous Groups and the Testing of Economic Hypothesis," *Journal of Financial and Quantitative Analysis*, 5 (1970), pp. 581-602.

13. Fama, Eugene F., and Michael R. Gibbons, "A Comparison of Inflation Forecasts," *Journal of Monetary Economics*, 13 (May 1984), pp. 327-348.

14. Farrell, James L., George Pinches, and James Scott, "Growth Stock Fundamentals: Implications for Portfolio Management," (1981). Unpublished paper quoted in *Guide to Portfolio Management* by James L. Farrell (New York: McGraw-Hill Book Company, 1983), pp. 211 and 220.

15. Farrell, James L., "Homogeneous Stock Groupings," *Financial Analysts Journal*, 31 (1975), pp. 50-62.

16. Farrell, James L., "Analyzing Covariation of Returns to Determine Homogenous Stock Groupings," *Journal of Business*, 47 (April 1974), pp. 186-207.

17. *Federal Reserve Bulletin* (1970).

18. Fogler, Russell H., Kose John, and James Tipton, "Three Factors, Interest Rate Differentials and Stock Groups," *Journal of Finance*, 36 (May 1981), pp. 323-335.

19. Ibbotson Associates, *Stocks, Bonds, Bills, and Inflation 1987 Year Book*, (Chicago: Ibbotson Associates: 1987).

20. Gittings, Thomas A., "Rounding Errors and Index Numbers," *Economic Perspectives*, Federal Reserve Bank of Chicago (May/June 1991), pp. 2-10.

21. Ingersoll, Jonathan E, "Some Results in the Theory of Arbitrage Pricing," *Journal of Finance*, 39 (September 1984), pp. 1021-1039.

22. King, Benjamin, "Market and Industry Factors in Stock Price Behavior," *Journal of Business*, 39 (January 1966), pp. 139-191.

23. Lachenbruch, P.A., "An Almost Unbiased Method of Obtaining Confidence Intervals for the Probability of Misclassifications in Discriminant Analysis," *Biometrics*, 23 (December 1967), pp. 639-645.

24. Martin, John and Robert Klemkosky, "The Effect of Homogenous Stock Groupings on Portfolio Risk," *Journal of Business*, 49 (July 1976), pp. 339-349.

25. Morgan, I, G., "Grouping Procedures for Portfolio Formation," *Journal of Finance*, 32 (December 1977), pp. 1759-1765.

26. Roll, Richard, and Stephen A. Ross, "An Arbitrage Pricing Theory Approach to Strategic Portfolio Planning," *Financial Analysts Journal*, 40 (May-June 1984), pp. 14-26.

27. Roll, Richard, and Stephen A. Ross, "An Empirical Investigation of the Arbitrage Pricing Theory," *Journal of Finance*, 35 (December 1980), pp. 1073-1104.

28. Ross, Stephen A., "The Arbitrage Pricing Theory of Capital Asset Prices," *Journal of Economic Theory*, 13 (December 1976), pp. 341-360.

29. Ross, Stephen A., "Mutual Fund Separation in Financial Theory—Separating Distributions," *Journal of Economic Theory*, 17 (April 1978), pp. 254-286.

30. *SAS/STAT User's Guide*, Version 6, Fourth Edition, Volume 1, (Cary, NC: SAS Institute Inc., 1985)

31. Shukla, Ravi, and C. Trzcinka, "Sequential Tests of the Arbitrage Pricing Theory: A Comparison of Principal Components and Maximum Likelihood Factors," *Journal of Finance*, 45 (December 1990), pp. 1541-1564.

32. Sudarsanam, P.S., and R.J. Taffler, "Industrial Classification in UK Capital Markets: A Test of Economic Homogeneity," *Applied Economics*, 17 (April 1985), pp. 291-308.

33. *Survey of Current Business* (U.S. Department of Commerce, 1969).