AN ALTERNATIVE TO CONVENTIONAL INDUSTRY DEFINITIONS IN MODELING EARNINGS GENERATION PROCESSES

By

WILLIAM KEMP CARTER

Bachelor of Science in Business Administration
University of Southern Mississippi
Hattiesburg, Mississippi
1973

Master of Science
University of Southern Mississippi
Hattiesburg, Mississippi
1974

Submitted to the Faculty of the Graduate College
of the Oklahoma State University
in partial fulfillment of the requirements
for the Degree of
DOCTOR OF PHILOSOPHY
December, 1977

AN ALTERNATIVE TO CONVENTIONAL INDUSTRY DEFINITIONS IN MODELING EARNINGS GENERATION PROCESSES

Thesis Approved:

Larry B. Christensen
Thesis Adviser

James R. Boatsman

John D. Rea

James E. Shamblin

Norman N. Durham
Dean of the Graduate College

PREFACE

This study consists of an examination of a non-industry-based method of classifying firms into groups for the purpose of accounting for inter-firm performance similarities. The primary objective is to determine whether a grouping scheme based on leverage, liquidity, payout, growth, and size might outperform traditional industries in the search for common factors in firms' performance.

The author wishes to express his appreciation to his dissertation committee chairman, Dr. Lanny G. Chasteen, and to Dr. James R. Boatsman, for their assistance throughout this research. Appreciation is also expressed to the other committee members, Dr. John D. Rea and Dr. James R. Shamblin.

Special thanks are due to the Ernst and Ernst Foundation, whose generous financial support has made this research possible.

Finally, gratitude is expressed to my wife, Imelda, for her encouragement and many sacrifices.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER I

INTRODUCTION

Research concerned with time-series properties of earnings numbers and estimates of the parameters of these time series models frequently is conducted with the assumption that earnings numbers of firms may be regarded as series of observations on independent random variables. To the extent that cross-sectional dependencies in firms' earnings exist, such an assumption is inappropriate. Therefore, these analytical tests performed upon firms' earnings number series depend for their validity upon, inter alia, the extent to which the assumption of inter-firm independence is justified. As these tests are currently one subject of empirical research in accounting, a technique for removing existing cross-sectional dependencies from the earnings number series of firms is desirable.

As common factors are abstracted from firms' earnings, series of residual earnings numbers are produced which more closely satisfy the ideal of cross-sectional independence. Recent research in accounting and finance has investigated the potential of market and industry factors to account for these commonalities in earnings (and in security returns).

In another area of research as well, there exists a need to abstract common factors from the earnings series of a sample of firms. Efforts to characterize market expectations of earnings require identification of an "unexpected earnings" construct, which is frequently defined as that portion of earnings not attributable to the market effect. These estimated series of unexpected earnings are then examined to determine their degree of association with unexpected security returns, defined analogously. The objectives of such analysis have been to assess the information content of reported earnings numbers via their association with security returns (4); to establish the "usefulness" of quarterly EPS data (16); and to determine which of several earnings definitions is more closely associated with security returns (10) (11).

In these studies, an API is calculated by assigning each firm to one of two groups depending on the sign of the firm's unexpected earnings, with the assumption that each firm's assignment is independent of the assignments of other firms. Since the unexpected earnings values are estimated by abstracting only the market effect from reported earnings numbers, any remaining inter-firm commonalities render the independence assumption invalid.

This area of research, then, stands to benefit from the identification of non-market-wide earnings associations. Magee (34) has demonstrated empirically that failure to remove industry effects present in "unexpected" earnings can have a noticeable effect on the results of conventional API analysis. Magee (34) replicated part of the Beaver and Dukes (10) analysis after abstracting both market and industry effects in reported earnings, and found a dramatic decrease in significance of the difference Beaver and Dukes observed.

Earnings numbers of firms are expected to exhibit cross-sectional dependencies due to extant economy-wide influences such as monetary policy and interest rates. Further, earnings series of firms in the same industry are expected to exhibit additional dependencies due to factors which are common to these firms, but not common to all firms in the market. The source of these intra-industry associations, in theory, is participation in common input and output markets.

Ideally, common factors should be identified and abstracted from firms' earnings until perfectly uncorrelated series of earnings residuals result. The interpretation of these residuals would be that they are due to unique characteristics of each firm. One measure of success in removing common factors is to test the extent to which this ideal has been achieved. To assess the usefulness of market and industry factors in explaining dependencies among firms' earnings series, it is necessary to abstract these factors from the series, and examine how closely the residual earnings numbers approach the ideal of inter-firm independence.

An alternative test of usefulness of market and industry factors in explaining commonalities in firms' earnings series is the regression of a firm's earnings onto market and industry indices (34) (15) (14). When this analysis is performed for a sample of firms, the multiple R^2 averaged over firms is interpreted as that portion of variability in sample firms' earnings which is explained by market and industry factors. The difference between this average R^2 and that obtained by using the market index as the sole independent variable is attributed to industry factors.

Need for the Research

Upon inspection, some flaws in the "industry effect" argument become apparent. First, industry membership, and even industry definition, are not objectively observable phenomena. Even if a classification scheme such as the Standard Industrial Code (SIC) were deemed acceptable, the question of how broadly industries are to be defined has no obvious answer. It seems plausible that the similarities among firms in an industry would be greater if the industry were defined more narrowly. Are not firms which manufacture only micro-wave ovens expected to be more similar to each other than are all those firms which manufacture electrical appliances of any kind? Depending on the breadth of industry definition chosen, then, the strength of industry effects in firms' earnings may vary considerably. There is evidence of this phenomenon in the literature (34) (22).

Second, even if an acceptable breadth of industry definition could be agreed upon, the question remains whether any such classification scheme can be meaningful. The more basic question of the domain of industry definition has not been resolved. Given that product similarity determines industry membership, must the products be similar (substitutes) in production, or in consumption, or both? The object of interest, which is whether firms participate in the same input and output markets, embraces both. Requiring both would result, however, in essentially single-product industries, containing firms which produce items for the same use, manufactured from the same raw materials. These criteria have been relaxed by the SIC classifications, which base industry membership on substitutability in production or in consumption, rather than both. However, within a single SIC 3-digit industry, the classifications are not consistent. For example, apparel items made from plastics or textiles, such as raincoats, are classified with plastic shower curtains. Within this grouping are found products which are substitutes in consumption but not in production--plastic raincoats and cloth raincoats--others which are substitutes in production but not in consumption--plastic raincoats and shower curtains--and others which are substitutes in neither consumption nor production--cloth raincoats and plastic shower curtains. These examples are not meant to criticize the SIC classifications, but to illustrate how product diversity renders any such scheme practically incapable of achieving the objective of classifying firms together which are "alike" in some respect.

In fact, there is serious doubt as to whether such a classification scheme is meaningful at all. Bock (13) states that:

> . . . a detailed analysis of the LB (line of business) categories made up of single 4-digit SIC codes would be likely to show that few, if any, could be directly correlated with markets whose boundaries are not open to serious debate (p. 8).

Benston (12) states the case even more forcefully:

> . . . the products included in the FTC (Federal Trade Commission) designated lines of business, which are based on SIC categories, generally are not substitutes in demand. Hence, they do not relate to markets and are of little value for many of the benefits claimed . . . There also is doubt that the categories contain products that are substitutes in production, which obviates other benefits, such as industry studies . . . (p. 63).

Finally, any product-similarity-based definition of industry results in exclusion of those firms whose diversification makes them impossible to classify in one industry. Another aspect of the same problem concerns those multi-product firms that are classified. Frequently, a firm whose shipments in one industry exceed some minimum portion of its total shipments is classified as a member of that industry, provided it has greater shipments in no other industry. The objective of industry classification hardly seems achieved by such a solution.

These flaws in traditional industry classification are inconsistent with the expectation that industry factors could account for nontrivial commonalities in firms' earnings. They further indicate that the explanatory power attributable to industry factors should vary inversely with the breadth of the industry definition chosen, as found by Magee (34).

Purpose of the Research

The performance of business firms is certainly affected by a variety of factors, some unique to particular firms and some held in common with others. For a classification scheme to produce groups of firms which hold potential for accounting for earnings commonalities, then, the scheme must produce intra-group similarity and inter-group dissimilarity with respect to some of those underlying earnings-influencing factors. The unanimity of past research in choosing traditional industry classifications as the grouping scheme implies an assumption that product-similarity-based industries constitute the best possible way of representing inter-firm similarities with respect to earnings-influencing factors. It is not at all clear why this should be so, especially in light of the many flaws which hinder traditional classification in the achievement of its objective.

Traditional industry classification can be very generally described as one method of segmenting a sample of firms into a set of mutually exclusive but not necessarily collectively exhaustive groups. Product similarity is the sole criterion for inclusion of two firms in the same group, and the required level of similarity may vary according to the analyst's wishes. Viewed in this way, industry classification is simply a clustering method, where industries are clusters, firms are objects, the sole clustering variable is product type, and the inter-object difference measure takes on only two values--a pair of objects or clusters are either "similar" or "dissimilar".

Since some underlying firm characteristics must exist which produce inter-firm earnings associations, then firms that are most like with respect to these characteristics should exhibit the strongest earnings associations. But rather than assume that traditional industry classification produces the greatest intra-group similarity and inter-group dissimilarity with respect to these unknown earnings-influencing characteristics, it is desirable to investigate whether some other basis of classification might prove more useful. Since industry classification seeks to represent the effects of firms' competition in the same input and output markets, it is difficult to imagine a grouping scheme broader in intended scope. However, in light of the many flaws of the traditional scheme and its unimpressive results to date, one such effort is undertaken in this study. This approach consists of identifying some easily measured variables to be used as measures of, or surrogates for, underlying earnings-influencing characteristics of firms. Similarity with respect to these variables will be used as the criterion for forming groups of firms, where the traditional scheme uses product similarity. The use of the variables chosen is warranted by the theoretical support given in the current literature for their potential to influence firms' performance. The usefulness of this grouping scheme will be compared with that of traditional industry classification in terms of ability to account for commonalities in earnings number series of a sample of firms.

Literature Review

If security prices result from actions of market agents whose concern is with cash consequences of their actions, then to the extent that reported earnings portend cash consequences, theoretical support exists for a link between security prices and reported earnings. Empirical evidence of this link is provided by Ball and Brown (4). Recent research works relevant to this study are therefore divisible into two groups: those concerning security returns and those concerning reported earnings.

Studies of Security Returns

In his definitive study of market and industry factors in stock price changes, King (31) found that approximately one-half of a security's price variation was explained by the market effect and an additional one-tenth by industry effects, on the average. However, "his sample and his empirical techniques were chosen in order to maximize the portion of this interdependence [among price changes of securities] which could be explained . . ." (38, pp. 695-696) by market and industry factors. King's sample of 63 firms from six SIC two-digit industries

> . . . reflected King's desire to explain as much as possible of the dependence structure among a group of stock price relatives on the basis of industry factors and accordingly included only stocks representing well defined, homogeneous industry groups. Unfortunately it is doubtful that conclusions concerning the significance of industry factors based on such a sample should be extrapolated to a universe of stocks many of which do not belong to any single, well-defined industry group (38, p. 697).

Myers (38) replicated King's study, and on the basis of his findings issued a call for further research:

> Thus, despite some evidence that certain components tended to be associated more closely with certain industry groups than with others, there was substantial evidence that some components of security price changes were virtually independent of industry classifications. This suggests the possibility that some characteristics of companies other than their industry classification could be associated with similar patterns of price change behavior (p. 704).

Meyers findings are consistent with two conclusions. First, industry effects may not be as well-defined as King's analysis seemed to indicate. Second, King's choice of six particular industries may have resulted in an observed industry effect somewhat stronger than that of industries in general.

Gaumnitz (24) used a methodology similar to King's to cluster 140 firms based on their similarity of security returns. He found clusters of firms having little resemblance to traditional industry groupings. This result was obtained regardless of whether the market effect was first abstracted from the security return series. Gaumnitz concluded that some industries produced noticeable effects, but a majority did not.

Fertuck (22) compared the relative strengths of SIC one-, two-, and three-digit industry effects in security returns. He also formed clusters of firms based on past covariance of residuals to determine whether such pseudoindustries would exhibit similar return patterns in a future period. Concerning the effect of breadth of industry definition on the strength of observed effects, Fertuck found the expected result--that the narrower the definition of industry, the stronger the observed effect. SIC one- and two-digit industry effects accounted for less than three percent of the variance in security returns. The strength of SIC three-digit industries varied from as low as 1.4 percent to as high as

11.5 percent. By way of comparison, the market effect accounted for between 25 and 30 percent of the variance in security returns. The clustering of firms based on covariance of past residuals performed worse than traditional industry groupings. However, the one instance in which three-digit industries explained an average of more than three percent of variance in returns was in the sample containing only those industries exhibiting past patterns of high intra-industry covariance of residuals. Fertuck (22) concluded that

> . . . for studies using the Market Model, it is necessary to be very careful when deciding whether to use an industry index to remove systematic movements. In some industries the industry effect is trivial and can be safely ignored. In others, it can be as large as a third of the market effect (p. 847).

Aber (1) examined the performance of several multi-index models of security returns in eliminating dependencies in the return residuals of a sample of 75 stocks. For comparative purposes, one model was tested utilizing a market index as the sole independent variable. Three other types of indices were also used: SIC two-digit industry indices, indices based on the performance of six stock groups, such as "growth" and "conglomerate"; and four macroeconomic indices, such as the Consumer Price Index. One multi-index model was constructed using the market and industry indices; another with the market and stock group indices; and the last with market, stock group, and macroeconomic indices. These three multi-index models were compared with the single index model in terms of the extent to which return residuals from each model were uncorrelated across stocks.

Using a Chi-square test for the absence of correlations in each model's residuals, Aber found that all three multi-index forms were significant at the one percent level, while the single-index model was not significant at the five percent level. His results demonstrate that significant commovements in security returns escape the market index, but are subject to measurement by other indices.

Numerous researchers have abstracted a market effect from security price changes to obtain series of "unexpected" price changes. May (36), Baskin (7), Beaver (9), Beaver and Dukes (10), Beaver and Dukes (11), Ball and Brown (5), Ball (6), Kaplan and Roll (30) and Foster (21) are some examples. Based on the relative insignificance of industry effects as observed by King (31), none of these studies sought to remove other common factors.

Studies of Reported Earnings

To investigate whether earnings numbers might exhibit market-wide and industry-wide associations similar to those found in security returns, Brown and Ball (15) applied SIC two-digit classifications to a sample of 316 firms using six different definitions of earnings. The strength of market and industry factors was measured as the average of R^2 values from the 316 least-squares regressions. Each firm's earnings was regressed onto market and industry indices of earnings, once for each of the six definitions of earnings used. The market index was calculated as the average value of the earnings numbers of all the firms. To avoid multicollinearity of the two independent variables, the industry index was defined as that portion of industry-wide average earnings which was not explained by the market factor. By calculating average R^2 values with the market index as the sole independent variable, and again with market and industry indices both used as independent variables, the explanatory powers of the two factors were measured. Brown and Ball found that an average of 35 to 40 percent of earnings variability was accounted for by the market effect, and an additional 10 to 15 percent by industry effects.

Gonedes (25) examined the empirical properties of market index models using two definitions of earnings and two index construction schemes. In addition to the usual average-earnings method of index construction, he tested a second scheme in which each firm's earnings number is weighted by the size of the firm's common stockholder equity. The two indexing methods then measure the relative importance of number of firms versus size of firms in capturing market-wide earnings commonalities. In addition, for each definition of earnings used, Gonedes defined another earnings measure as consisting of first differences in the series, as opposed to levels. Gonedes' analysis indicated that the first-difference form of the market model, with an equity-weighted index, is the most statistically valid model of earnings number generation. This model outperformed (in terms of predictive ability and satisfaction of statistical assumptions) the other forms of the market model tested, as well as alternative (non market) models including a martingale and several "naive" models such as linear extrapolations of past observations. However, upon examination of residuals, he found evidence of some remaining cross-sectional correlation.

Magee (34) hypothesized that the observed weakness of industry effects relative to the market effect is due to the broad definition usually given to industry, and that industry effects might therefore be better assessed by use of a narrower classification scheme. He chose SIC four-digit industries in his study, and, based on Gonedes' (25) results, used the equity-weighted index construction scheme. Magee divided the firms in each industry into two groups, one for calculation of the equity-weighted industry index and another for use in the earnings regressions. Those in the index-construction groups of all industries, plus some firms not classified into industries at all, were used for calculation of the market index. Magee found the market effect accounted for an average of 18 percent of variability in 165 sample firms' earnings, and industry effects explained an additional 16 percent. It should be noted that while Magee observed a relatively strong industry effect, he also used a rather "strong" definition of industry. His results are therefore consistent with *a priori* expectations of the relation between the narrowness of industry definition and the strength of the industry effect. Magee concluded that the traditional assumption of uncorrelated residuals from the market model is very questionable.

Brealey (14) chose a sample of 217 firms representing 20 industries. Although he neglected to report the basis on which he selected these industries, his list of industry names appears to represent a mixture of SIC three-, four-, and five-digit industries. Brealey found that an average of 21 percent of movement in earnings could be accounted for by the market factor, and another 21 percent by industry effects.

At least two studies involving alternative grouping schemes have been conducted. Elton and Gruber (19) calculated 15 annual growth rates of earnings per share, 1948-1963, for each of 180 firms. These 15 growth rates were used as variables in principal components analysis, and 11 of the principal components were chosen to serve as clustering variables. The pseudo-industries that resulted from the clustering were compared with traditional industries in terms of usefulness in predicting earnings per share changes.

The EPS prediction method consisted of fitting an OLS regression within each industry or pseudo-industry, whose dependent variable was the 1960-1961 EPS change, and whose independent variables were 23 measures of "the type and size of sources of funds, measures of uses of funds, measures of profitability, measures of historical growth rates, and measures of liquidity" (19, p. 82). Theoretical support for use of the particular variables is not reported by Elton and Gruber.

The OLS regression was fitted cross-sectionally for each industry and pseudo-industry, with each member firm providing one observation on each variable. The resulting regression equation was used with 1963 values of the 23 independent variables to predict the 1963-1964 EPS changes of the firms in the industry or pseudo-industry from which the equation was calculated. For comparative purposes, Elton and Gruber (19) also calculated EPS forecasts using a mechanical extrapolation technique which had been found in an earlier study to be indistinguishable in performance from the forecasts of security analysts.

The mechanical method outperformed traditional industries at the .01 level of significance. The alternative grouping scheme outperformed the mechanical method at the .05 level. Elton and Gruber concluded that the alternative grouping scheme is more useful for preparing forecasts of EPS than are traditional SIC four-digit classifications, and is also more useful than the mechanical technique which performs as well as security analysts.

An alternative conclusion is that the <u>forecasting technique</u> used within the groups and industries performs better in the groups than in the industries. Since the groups were formed on the basis of EPS growth rates, this would seem quite possible. In essence, the findings could have resulted from a bias on the part of the forecasting technique used. One additional qualification needs mention concerning the identities of some of the 23 variables used. Not only are no arguments offered in support of the particular variables used, but two of them are themselves measures of the change in EPS: the five-year average growth in EPS and the five-year standard deviation of growth in EPS. Since the variable to be predicted is the growth in EPS, the use of these two in its prediction would seem to render the analysis somewhat circular.

Williams and Goodman (45) addressed the question of whether a grouping of firms based on their financial characteristics would be similar to industry classifications, in an effort to show that firms with similar operating characteristics (industry membership) also have similar financial characteristics. An implicit assumption made by Williams and Goodman is that traditional industry classification is itself meaningful, that is, that industry membership measures some similarity of operational characteristics.

The financial characteristics used in the analysis were 14 in number, and, like Elton and Gruber (19), Williams and Goodman (45) offered no theoretical justification for their inclusion. A discriminant function based on the 14 financial variables was able to classify an average of 72 percent of the sample firms into the correct SIC industry in each of the five years tested. Worthy of note is the fact that the sample of firms used in the discriminant analysis represented only five SIC industries, so that a random classification would correctly identify an average of at least 20 percent of the firms.

Summary

The foregoing survey of recent research points to several conclusions. First, there is ample empirical evidence that the observed strength of industry effects in firms' earnings and in security returns varies directly with the narrowness of the definition chosen for the industries. As mentioned earlier, this is entirely consistent with *a priori* reasoning. Second, there is a conspicuous lack of unanimity among researchers as to the appropriate degree of aggregation for defining an industry, which is consistent with the earlier observation that industry membership is not an objectively observable phenomenon. Third, the relative weakness of industry effects (both in firms' earnings, the domain of the current study, and in security returns) has led in more than one instance to the question of whether some other grouping scheme may prove useful; observations to this effect are found in Meyers (37), Fertuck (22), and Brown and Ball (15). This is the research question addressed in the remainder of the current study.

Overview of Subsequent Chapters

The theoretical relationships which provide the basis for this study are examined in Chapter II. Chapter III describes the reserach methodology used and the reasoning behind the choices of particular analytical techniques. The results of the analysis are reported in Chapter IV. Chapter V contains a summary of the research results and contributions, discussion of the limitations of the study, and recommendations for further research.

CHAPTER II

CONCEPTUAL BASIS FOR THE RESEARCH

The usual linear model of security returns was first introduced by Sharpe (42) as a manageable simplification of the Markowitz (35) approach to portfolio selection. In this "linear" or "market" model, returns on different securities are assumed to be related to each other only through their relationship to an index of overall market performance. This is the simplifying assumption made by Sharpe, which obviates the need to calculate the covariance of every possible pair of securities in the portfolio. This assumption is equivalent to stipulating that if the market effect were abstracted from each security return series, the residual series of each security's returns would be uncorrelated with the residual series of any other security. The relation between the return series of a given security and the market index of returns is given by Sharpe as follows, where i denotes the particular security and t denotes the time period:

$$R_{i,t} = A_i + B_i I_t + C_{i,t}. \qquad (1)$$

A and B are the intercept and slope parameters of the linear relation between the market index, I, and the random variable, R, whose distribution determines the possible returns on the security. C represents a random element, or distribution term, in the series of security returns. The Sharpe formulation also makes the following restrictions:

$E(C_i) = 0$, and therefore $E(R_i) = A_i + B_i(E(I))$;

$E(C_iC_j) = 0$ for all $i \neq j$.

Since there is both theoretical and empirical support for the supposition that the generation of security returns is related to some concept of firms' earnings, a corresponding model is posited for earnings, where E is the random variable whose distribution determines the possible earnings numbers. The ' denotes that the variable or constant relates to firm earnings rather than to security returns; otherwise, all symbols are used as they were above:

$$E_{i,t} = A'_i + B'_i I'_t + C'_{i,t} \tag{2}$$

again with $E(C'_i) = 0$ and $E(E'_i C'_j) = 0$ for all $i \neq j$.

An assumption of both models is that the random element in the series is serially uncorrelated, and further is cross-sectionally uncorrelated with the corresponding random element in the series of any other firm, as stated symbolically above. This has led to the use of market models to derive estimates of the unexpected components of series of security returns and firms' earnings. Confining attention to the model of firms' earnings given in equation (2), E_{it} can be defined as any one of several measures of performance, such as reported earnings as a percentage of total assets or as a percentage of total invested capital. The market index of performance, I', would then be estimated by some average of E_{it} values calculated across a large sample of firms, with the particular sample representing the market. The parameters of the relation between these two variables are subject to estimation by ordinary least squares (OLS). The regression model can be represented as:

$$E_{i,t} = a_i + b_i M_t + e_{i,t} \tag{3}$$

where M is the estimate of market performance I', a and b are the OLS estimates of the parameters A' and B', and e is the regression error or residual. The regression residual e is then taken as an estimate of the random element C'. When the assumptions of the OLS regression model are met, the e series of a given firm will in fact be serially uncorrelated and have zero expectation.

However, the use of the residual series as a measure of the firm's unexpected earnings depends as heavily on another assumption of the market model, which is that the earnings series of a group of firms are related to each other only through their relation to the market, i.e., that the random element in the earnings series of each firm is not correlated with that of other firms. To the extent that this assumption is invalid, the market model of firms' earnings can be used to calculate firms' unexpected earnings only if existing cross-sectional correlations among the e terms do no violence to the test being performed.

The basis for such additional cross-sectional correlation exists in any earnings commonality among a group of firms which is not captured by the measure of market performance. Since the market index of performance captures the greatest possible earnings commonality shared by all firms, any additional commonality present in the earnings series of some (but not all) firms will result in earnings residuals of these firms exhibiting the same commonality in the form of remaining cross-sectional correlation. The presence of such a commonality in addition to the market factor is presumably measurable through some characteristic of firms which produces correlations among their earnings series. This weakness in the market model can therefore be described as an omitted variable. If the particular characteristic of firms could be measured, or if the firms possessing this characteristic could be grouped to permit inclusion of the group's earnings index as a second independent variable in the model, the resulting residual series would be devoid of the earnings commonality attributable to the new variable.

This essentially describes the basis for choosing industry effects as the cause of remaining earnings correlations, and industry indices of performance as the additional variable needed in the model. A group of firms participating in the same input and output markets should exhibit earnings commonalities through the effects that caprices of supply and demand have on the firms' performance. But if not all firms participate in these markets, the market index of performance cannot capture this commonality--the market index captures only that greatest possible commonality shared by all firms. If industries do consist of groups of firms participating in common input and output markets, then use of industry indices of performance could remove these commonalities from firms' earnings just as use of the market index removes the market effect.

Chapter I discussed the weaknesses in SIC industry classifications. Therefore, it is not surprising that research has failed to discover consistently sizeable commonalities in the earnings of firms in the same industry in addition to the commonality attributable to the market effect. This raises the question of whether some other grouping of firms might perform better, and whether the small commonalities attributed to industry effects might be due to chance. Indeed, if each firm's earnings component not attributable to the market effect is correlated to some degree with the corresponding component of many other firms, then a choice of several firms to form an "industry" may produce a group exhibiting some commonality due to chance.

The usual explanation of why firms in the same industry will exhibit earnings commonalities consists largely of an oral tradition. As stated above, firms' participation in common markets is the gist of this argument. Upon inspection, this premise appears to be a summary of many influences, some of which are (1) a common degree of financial risk (leverage) compatible with the particular degree of business risk existing in the industry; (2) a common mix of asset types best suited to the production activity; (3) competition for a finite supply of investors' contributed capital; (4) the effects of industry-wide growth or decline; and (5) the effects of firm size, i.e., the minimum scale of operations necessary for efficient operation in the industry.

The imprecision of using industry membership to surrogate these inter-firm similarities is revealed upon inspection of the phenomena named above: these effects are not unique to intra-industry associations. For example, the influence on performance caused by a particular degree of leverage, (1), is not only shared by firms in one industry which have that degree of leverage, but is also common to any other firm with that leverage ratio. If the intent is to measure inter-firm similarity of leverage, then it is not necessary to assume that common industry membership surrogates a common degree of leverage--the leverage ratio can be measured directly. With respect to the supposed similarity of asset mix, (2), this too can be measured more directly using reported financial data than by assuming firms in the same industry have the same asset mix.

Competition for a finite supply of investor capital, (3), occurs not only between firms in the same industry, but among all firms. The investors' choice of one firm's stock over another's is based on the risk and return characteristics of the two securities, which are more directly measurable via characteristics such as the market beta and the firm's payout ratio. The effects of growth and decline of an industry, (4), are also not necessarily common to all firms in the industry. A failing firm in a growing industry could hardly be expected to enjoy the effects of industry growth as could a growing firm in the same industry. Therefore, rather than assume all firms in an industry share alike in the effects of industry growth or decline, the growth of each firm can be measured directly. The effects of firm sizes, (5), on performance need not be assumed common to all firms in a particular industry, since size too can be measured directly for individual firms.

Because of the weaknesses of traditional industry classification, it is hypothesized that an alternative grouping scheme may perform better in explaining variability among firms' earnings series and in producing residual series devoid of cross-sectional correlation. This alternative scheme is based on the five phenomena discussed above, which are among those which traditional industry classification purports to represent. These five have been chosen primarily because they can be represented by five simple, familiar financial variables--leverage, the measure of financial risk; liquidity, one characteristic of asset mix; payout, a determinant of investors' choice; the change in total assets, a measure of growth; and total assets, a measure of size.

Due to the nature of the market effect in firms' earnings, i.e., that it is calculated to measure that greatest possible commonality in the earnings of all firms, then any characteristic of firms which influences earnings and is not common to all firms is capable of producing earnings commonalities which will not be captured by the market index. If traditional industries are in fact a poor surrogate for groups of firms which participate in common markets, then alternative classification schemes based on inter-firm similarities with respect to variables like the five discussed above may be useful in characterizing earnings commonalities which escape the market index, since such variables surrogate the effects traditionally attributed to firms' participation in common markets. In short, any measurable characteristic of firms which may theoretically influence earnings, and which differs across groups of firms, is a candidate for use as a criterion for forming groups in an effort to explain and remove correlations not captured by the market effect, as traditional industries are purported to do.

In this study, five characteristics of firms have been chosen as the criteria to be used in grouping firms as an alternative to traditional industry grouping. These five were chosen on the basis of their simplicity and familiarity, as well as their potential to surrogate the effects of firms' participation in common markets. The object of this effort is to illustrate the ease with which meaningful grouping schemes may be developed, and to compare the performance of traditional industry grouping with this alternative scheme.

In addition, a third method of grouping will be examined. It is of interest to determine whether the modest performance of industry classification in explaining earnings variability and in producing uncorrelated residuals is much larger than that attributable to chance. Therefore, groups of firms will be formed by random selection, and the strength of earnings commonalities present in these random groupings will be compared with that of traditional industry groups and with that of the alternative grouping scheme proposed in this study.

The five characteristics chosen as the criteria for forming groups of firms are leverage, liquidity, payout, growth, and size. Each variable's inclusion in the analysis is warranted by the theoretical support given in the literature for its ability to influence firms' earnings. This support is offered in the final section of this chapter. Where possible, the direction of the theoretical relation between earnings and each variable is developed, providing a prediction which will be tested in the preliminary stage of the analysis. The measure of firms' performance chosen for use as the dependent variable is presented in the next section.

The Performance Measure

The dependent variable $E_{i,t}$ used as a measure of firms' performance is defined here as an earnings measure scaled by the size of invested capital. The purpose of this scaling is to remove one effect that firm size is expected to have on earnings. Large firms must on average produce greater earnings than small firms, or else there would be serious economic incentives for investors to avoid forming large firms. Since this relation between size and earnings may overshadow other effects of interest, the performance measure is defined as dollars earned per dollar invested. Common stockholders are arbitrarily selected as the investing agents, so the numerator of the earnings measure is defined as Net Income minus preferred dividends, and the denominator as Common Stockholder Equity. Through this scaling of the performance measure, it should be possible to determine whether size affects performance in another way, namely, whether returns per dollar of invested capital vary with the amount invested. The possibility of such a relation will be examined by use of firm size as one of the independent variables.

Refer to the model of firm earnings given by equation (2). The earnings number series of two firms, i and j, are represented by $A'_i + B'_i I'_t + C'_{i,t}$ and $A'_j + B'_j I'_t + C'_{j,t}$. Some conclusions are possible concerning the correlation of the two series. Since A'_i and A'_j are constant terms, they contribute nothing to the correlation. The second term in each expression consists of a constant slope multiplied by the index--the same index for both firms. These terms are therefore perfectly correlated. The third terms are theoretically perfectly uncorrelated across firms. The degree of correlation between the two number series depends, then, on whether variations in the second term or variations in the third term dominate in determining the total variation in the expression through time. If, for example, the variations in the second term dominate, it could be due to large changes in the market index through time, or to relatively large slope parameters, or a combination of both. The variations in the third term could dominate to produce relatively uncorrelated earnings series only if the temporal variation of the random element of each series is very large compared to the changes in the second term. In short, an observed correlation between the earnings number series of the two firms may reflect principally their relation through the market index, which is entirely consistent with expectations. The object of greater interest is whether the random elements in the series are truly uncorrelated across firms, and evidence of this may be hidden by the dominance of the second term's variations in determining the total variation in each series through time.

Some advantage may be gained by performing the same analysis with the dependent variable defined as changes in, rather than the levels of, the performance measure. The series of values of the dependent variable for the two firms would then be given by:

$$A'_i + B'_i I_{t+1} + C'_{i,t+1} - (A'_i + B'_i I_t + C'_{i,t})$$
$$= (A'_i - A'_i) + B'_i(I_{t+1} - I_t) + (C'_{i,t+1} - C'_{i,t})$$

and similarly for firm j:

$$(A'_j - A'_j) + B'_j(I_{t+1} - I_t) + (C'_{j,t+1} - C'_{j,t}).$$

For each expression, the first term in parentheses is zero; the second is the slope parameter times the change in the market index; and the third is the change in the random element. The degree of correlation between the two series again depends on which of the latter two terms' variations dominate in determining the variation of the series.

Since each random element is serially uncorrelated and has zero expectation, temporal variations in its changes are likely to be as great as the temporal variations in its levels. Variations in the second term, however, depend on the relative size of changes in the market index as compared with levels of the index. The relative importance of variations in the second and third terms in determining the total variation of the series, and therefore in determining the correlation between the series of the two firms, is potentially different if the dependent variable is defined as changes in the performance measure rather than levels. If changes in the market index are relatively small compared to values of the index, which would seem to be a possibility, then variation in the random element of each series will be more evident when changes in performance are used as the dependent variable's values. This hypothesis provides a testable statement: correlations among the earnings series of firms are expected to be greater than correlations among the earnings-changes series of firms.

To provide for a test of this relationship, two forms of the dependent variable $E_{i,t}$ will be used in the analysis. The first, defined initially in this section, is the level of (Net Income minus preferred dividends)/(Common Stockholder Equity). The second is defined as the first differences in the values of this series.

These two forms of performance measures are both found frequently in the literature. Gonedes (25) has concluded that the first-difference form is slightly superior in describing a statistically accurate model of firms' earnings for isolating the market effect. We will attempt to replicate his results by using both forms, levels and changes, and will extend his work in that we investigate both market and "sub-market" effects in earnings.

The Independent Variables

This section contains discussion of the five characteristics chosen as the bases for forming groups of firms as an alternative to traditional industry grouping. Theoretical support is given for the relationship between firms' performance and each of the five characteristics, or independent variables. It is this potential of each variable to influence firms' performance that warrants use of the variables in forming meaningful groups of firms. Where possible, the direction of the relation between performance and each variable is also predicted; these predictions will be tested in an early stage of the analysis.

Leverage

Provided the rate of return on assets exceeds the cost of obtaining debt financing for those assets, then the use of debt increases the rate of return earned by investors in equity. Of the total return earned by the firm, proportionately more is enjoyed by owners than by creditors per dollar invested by each group, since the "excess" return on debt-financed assets accrues to owners.

However, the presence of debt carries with it the chance that in a period of poor business conditions, the rate of return on assets may fall so low that not only may there be very little return, if any, to owners, but the entire dollar return on all assets may be insufficient to cover the fixed return demanded by creditors. Had the firm chosen to obtain no debt financing, returns would still be low, but the firm could survive if owners were content to wait for better returns in the future. This alternative may not be available if substantial amounts of debt are present. If the firm is unable to meet the fixed amounts of debt service from a source other than current earnings, liquidation may be the only alternative.

When conditions are not so severe that liquidation is probable, but severe enough that the rate of return on assets falls below the return due to creditors, then of the total return earned by the firm, proportionately more is enjoyed by creditors than by owners, per dollar invested. Thus, the advantage produced by debt financing in a period of high returns on assets becomes a disadvantage in periods of low returns. The effects of both prosperity and depression on the returns enjoyed by owners are magnified by the presence of fixed commitment financing. This magnification of both good and bad effects can be described as simply an increase in the variance of returns to owners, or risk. Thus, the relative proportion of debt to equity is one component of the total riskiness of a firm, namely financial risk.

Under the risk-premium hypothesis, owners will demand an additional return to compensate them for bearing the additional risk. This argument is consistent with an expectation that firms with a higher degree of financial leverage must be producing relatively higher returns for owners, i.e., that leverage will be observed to have a positive relation with firm performance, as performance is defined here.

An opposing argument can be formulated by reference to the other component of a firm's total risk, namely business risk. If the operations undertaken by a firm are inherently risky, owners will demand a return premium to compensate for this form of risk as well. If the possible return from an undertaking is not sufficient to compensate for its inherent riskiness, it will not be undertaken at all. However, the return may be just sufficient to compensate for business risk, but not for any additional (financial) risk. For a firm to survive in such an environment, it must reduce the total risk borne by its owners. Only one component of risk is so controllable--financial risk. It is altogether possible that the return to owners may be quite high, reflecting the large business risk premium, while little or no debt financing is present. This argument is therefore consistent with the expectation for negative relation between leverage and the performance measure.

These opposing arguments render the relation between leverage and performance an indeterminate one. There is empirical evidence of this theoretical indeterminancy. Gale (23) conducted a study of the relation between market share and returns to equity, in which one control variable was defined as the ratio of equity to total capital, which is equal to one minus leverage as defined in this study. Based on the risk-premium hypothesis, he expected to find this variable positively related to return to equity: less debt signals high business risk, and therefore high profits for risk-average owners. The results showed that the relation was indeed positive and significant. In terms of the definition of leverage used in this study, his findings support an inverse relation between performance and leverage. In contrast, Baker (3) investigated the relation of leverage to profitability, using the same leverage definition as did Gale. He, unlike Gale, found the relation to be negative and significant.

Wippern (46) examined the relation between financial structure and the value of the firm, using a rather complex composite measure of leverage. He found that the use of leverage does enhance the wealth of shareholders, implying, like Baker, a positive relation between performance and leverage. In contrast, several studies have reached conclusions consistent with those of Gale. Sullivan (43) conducted a study of the relationships among market power, profitability, and leverage, and found that larger, more profitable firms tended to use less debt in their capital structure. Hall and Weiss (26) similarly found a negative relation between profitability and leverage, as did Arditti (2).

Due to the apparent theoretical and empirical ambiguity surrounding the relation between performance and leverage, the direction of the relation will not be predicted in this study. Leverage will be measured as the ratio of total debt to total assets, which is the definition employed most frequently in recent literature.

Liquidity

The portion of a firm's total invested capital which is held in the form of liquid assets is an indication of the firm's capacity to adapt to changing business conditions. As liquid assets are, by definition, more convertible into any other type, a more liquid firm is more readily able to bring about desired adjustments in its total portfolio of assets. This should serve to reduce the overall riskiness of the firm.

The risk premium hypothesis therefore applies to liquidity as it does to leverage, discussed above. For brevity, the argument is not repeated here. The same reasoning leads to the expectation that liquidity and firm performance will exhibit a positive relation--greater business risk requires greater liquidity to lower the total riskiness of the firm, and therefore the high (business) risk-high profit firms will tend to have greater liquidity.

However, as was the case for leverage, an opposing argument exists. The factors of production responsible for the greater portion of firms' performance are largely illiquid--machinery, buildings, natural resources, and inventories. If liquid assets are less profitable per dollar invested, this would suggest that the expected relation between performance and liquidity is negative.

As with leverage, a theoretical ambiguity exists concerning the direction of the liquidity-performance relation, so no prediction shall be offered. Liquidity will be measured as the ratio of current assets to total assets.

Liquidity is frequently used to refer to the ability of a firm or individual to pay existing debts as they mature, a slightly different construct than that intended by the use of the term in this study. A more precise term for use in the present study may be "non-capital-intensity", but for brevity and convenience, this wording is undesirable.

Payout

The theoretical relationship between a firm's performance and its tendency to distribute earnings in the form of dividends has been frequently addressed in analytical and empirical literature. Miller and Modigliani (39) argue that dividend policy should have no effect on the rate of return required by shareholders. If so, no relation is expected between payout and performance. Miller and Modigliani argue that if a low-payout firm's owners desire more dividends, they need only sell some of their shares. If a high-payout firm's owners would prefer retention of earnings by the firm, they need only reinvest their dividends in additional shares.

Opponents of this argument maintain that since the variability in a dividend stream is far less than the potential variability in the stream of wealth increments due to share price appreciation, the total variability of the owners' income is reduced whenever more dividends are paid. Since risk-averse owners would prefer this decrease in variability, the total return they demand from the firm is adjusted downward.

This would permit a less profitable firm to meet owners' required return through higher payout, leading to an expectation for a negative relation between performance and payout.

The capital gains treatment afforded share price appreciation in the past clearly gives an incentive to owners to prefer retention of earnings over dividends, cet. par., since dividends are taxed as ordinary income. The proponents of the negative payout-return relationship argue that the effect of dividends in reducing the variability of the total return outweighs this tax effect, so that the total return demanded by owners will still decrease as payout is increased, preserving the negative relation of payout and performance.

A third argument not often mentioned in the literature is that the tax treatment of dividends versus capital gains leads to a shareholder preference for retention. A relatively unprofitable firm could then reduce the return its owners demand by filling their preference for earnings retention. This would lead to an expectation that low payout would occur in low performance firms. A highly profitable firm would not have this incentive for earnings retention, since its high earnings alone could satisfy the required return. However, this fails to explain why a high performance firm would then choose to pay more dividends, which is necessary to complete a positive relation between performance and payout. This argument is therefore rejected.

Arditti (2) sought to relate required return to several variables, among them the dividend payout rate. He found a significant negative relation, which supports the argument of the second school of thought outlined above. McKibben (37) utilized an expected negative relation between payout and return in successfully identifying a group of firms with consistently superior security returns over a 12-year period.

The relation expected to be observed between payout and performance in this study is negative. Payout will be measured as the ratio of dividends on common stock to income available for common (Net Income minus Preferred Dividends).

Growth

When growth is measured as the change in total assets, it represents the attraction of new amounts of capital to the firm. This can be achieved through additional debt or equity investment, or through another form of owners' investment, earnings retention.

If capital does in fact move in response to the expected return to be earned, a firm which is succeeding in attracting new capital must be offering higher returns than some competing uses of capital. This creates an expectation that growth is positively related to the performance of the firm.

A competing argument is that higher return is not the only explanation for a firm's success in attracting new capital. The attraction of the firm as an investment for new capital depends not only on the expected return, but upon risk as well. This produces no necessary relation between growth and performance, since a low-risk, low-profit firm could meet investors' demand as well as a higher-risk, higher-profit firm, and therefore could experience growth. However, in the current study, growth, like the other independent variables, will be measured over a considerable period of years. For a firm to produce a high value of the growth variable, it must experience a sustained rate of growth over this time. Is is maintained here that while the risk-return argument could explain growth in a low-profit firm in the short run, the ability of a low-risk, low-profit firm to continue to find and undertake additional low-risk projects over the long run is suspect.

It is hypothesized that the growth variable used in this study measures the firm's record of long-run sustained success in attracting new capital, and that its expected relation with firms' performance is positive. Growth will be measured as the change in the size variable, i.e., the change in the natural logarithm of total assets.

It is unclear what effect mergers, acquisitions, and divestments of subsidiaries will have on the growth-profit relation. These types of changes in the firm's total assets do not necessarily represent the acquisition (or loss) of "new" capital, so no conclusion is possible, in general, concerning whether such growth is a signal of superior performance. In recognition of these possibilities for spurious observations of "growth", firms whose identities changed significantly due to merger, acquisition, or divestiture of a segment shall be excluded from the study.

Size

As stated above in the definition of the performance measure used in this study, the size of a firm has one direct effect on the total earnings it produces. Large firms must produce a larger dollar amount of earnings than do smaller firms, or else the owners (and managers) of large firms would have a powerful incentive to segment a large firm into two or more small ones. The ability of large firms to produce a greater absolute amount of earnings is made apparent, therefore, by the continued existence of large firms.

It is because of this obvious relation between size and absolute dollar earnings that the performance measure was defined as earnings scaled by the size of invested capital. Nothing would be gained by demonstrating the obvious relation, and its presence could have obscured other effects of interest. The performance measure used here should remove automatically the effects of size on absolute earnings and make it possible to examine the effects of size on the rate of earnings per dollar invested. This is the size-performance relation of interest.

It is conceivable on several grounds that large firms may produce a greater return per dollar invested than do small firms. If the operations of the large firm and the small firm are identical except for size, the large firm has potential advantages provided by quantity discounts on its larger purchases of raw materials. Some services the small firm must seek in the market, such as insurance, can be provided internally by the large firm; thus the large firm may produce savings equal to the profit that would have accrued to the outside provider of these services, but without the additional investment in capital that the service company must make. Continuing the example of insurance services for illustration, the costs of adversary relationship between the insured and the insurer that could arise in the case of a disputed claim, which include attorneys' fees and court costs, would simply not arise when the large firm buys no outside insurance services.

In addition to these economies of scale, there is another potential profit advantage accruing to only large firms. Small profitable projects may be undertaken by small or large firms. Large profitable projects may be undertaken by only large firms. Large firms will therefore find that they have available not only the investment alternatives that are open to small firms, but also others. If some of the most profitable projects are among those requiring large investment, and if capital moves in response to profit differences, then relatively high profitability must exist in any activity in which large firms are operating. If small projects offered equally high rates of return, large firms would have no incentive to undertake large-scale investments. Baumol (8) has stated this relationship concisely:

> . . . at least up to a point, increased money capital will not only increase the total profits of the firm, but because it puts the firm in a higher echelon of imperfectly competing capital groups, it may very well also increase its earnings per dollar of investment even in the long run, after all appropriate capital movements are completed (p. 34).

Using this notion of the size of a firm, i.e., that of the largest scale of investment alternatives open to the firm, the relevant measure of size must be based on total capital regardless of source. Total owners' equity does not accurately measure the maximum scale of investment possible, for additional capital from any source increases this maximum possible scale. Total sales or some other measure of activity is not appropriate either, for such measures may indicate only that the firm operates near the final stages of the production process.

There are reasons to expect that the relation of size and performance is curvilinear, requiring some adjustment to the use of total capital (assets) as the size measure. For one, diseconomies of size may begin to set in beyond some point. Presumably firms will not operate beyond the point where diseconomies outweigh economies--higher rates of return would be earned by scaling down. But before this point is reached, the early effects of diseconomies should begin to slow the rate at which additional size increases the returns per dollar. Another reason for the increases in profit rates to slow beyond some point is the effect of public criticism. Very large firms are certain to be more in the public eye, and therefore may have incentives to avoid reporting the highest profit rates of all firms. Owners and managers may both prefer to take some of their rewards in other forms--sacrifice some return for lower risks, for example.

This curvilinear relation is more likely to be accurately measured using a transformation of total assets. The natural logarithm has been chosen for use in this study in recognition of its successful use in other recent research. Size will therefore be measured as the natural logarithm of total assets, and its predicted relation with the performance measure is positive.

Hall and Weiss (26) tested the size-profit relation, using the log transformation adopted here, and found the expected positive sign. This result was unchanged whether profit was defined as return to owners' equity or as return on total assets. Hall and Weiss strengthened this finding by stating evidence that their analysis should tend to understate the size-profit relation, if anything.

Summary

This chapter has outlined the bases for expecting firms' earnings to exhibit cross-sectional dependencies, and the reasons why the conventional single-index model of earnings fails to account for all of these comovements. In light of the flaws which plague traditional industry classifications, as discussed in the first chapter, it is desirable to investigate whether some other grouping scheme might perform better in accounting for these dependencies which escape the single (market) index model.

The grouping scheme proposed as an alternative to traditional industries is based on five familiar financial characteristics of firms. Their use in developing a meaningful method of analysis is warranted by theoretical and empirical support for a relation between firms' performance and each one of the five characteristics. This support has been presented in the preceeding section.

The computational techniques used in forming groups of firms based on these five characteristics, the tests and interpretations to be conducted, and the criteria for sample selection are the topics of Chapter III.

CHAPTER III

METHODOLOGY

In Chapter II, the linear model of firms' earnings based on the market index of earnings was given as equation (2):

$$E_{i,t} = A'_i + B'_i I'_t + C'_{i,t} \qquad (2)$$

which can be estimated by the OLS regression equation (3):

$$E_{i,t} = a_i + b_i M_t + e_{i,t}. \qquad (3)$$

A two-index model of earnings, such as the traditional model based on market and industry factors, can be similarly estimated by:

$$E_{i,t} = a'_i + b_{m,i} M_t + b_{g,i} G_{i,t} + e'_{i,t} \qquad (4)$$

where E and M are as before, a' and e' are the intercept and residual terms of the two-index form, b_m and b_g are the slope coefficients of the two indices, and G is the second index, traditionally an index of industry performance.

Equation (4) shall be used as a general representation of OLS estimates of the two-index models used in this study, but with the second index, G, not restricted to represent industry performance. G shall instead represent the index of performance of a group of firms, with the nature of the group defined in three different ways: traditional industries, groups formed at random, and groups formed on the basis of similarity with respect to the five financial variables given in Chapter II. This study consists essentially of comparing the performances of the three forms of group index in explaining the variations in firms' performance (earnings) and in producing series of residuals which are uncorrelated across firms.

In using these two-index regression models, it is desirable to avoid multicollinearity of the two indices used as independent variables. This can be achieved by the method found in Brown and Ball (15). Before estimating equation (4), the second index will be redefined as that portion of the group index of performance which is not explained by the first index. The two indices will be calculated as some average of performance over the several firms. The group index thus calculated will then be regressed onto the market index, and the residuals resulting from this regression will constitute the redefined group index. These equations are presented symbolically in the summary at the end of this chapter.

One other adjustment is also desirable. If the indices are calculated as some average of earnings across firms, then as each firm's earnings series is regressed onto the indices, it is to an extent being regressed onto itself. This spurious relationship can be eliminated from the analysis by eliminating from the calculation of the indices the one particular firm under consideration. The market index will therefore be some average of earnings across all firms except the one firm which is to be related to the index. This will require the calculation of a unique market index for each firm in the analysis, but is an easy adjustment to make. Similarly, in calculating the group index for each firm, all firms in the group except the one firm under consideration will be used in deriving the group index. This will result in a unique group index for each firm in the group.

The main thrust of the analysis will consist of comparing the values of R^2 from the regressions using the three different forms of group index, and comparing the extent to which each form produces uncorrelated residuals. Before this step is reached, however, the sample of firms must be selected, the variables measured, preliminary tests will be performed, and the firms will then be divided into groups. These steps are the topics of the following sections.

Sample Selection

Firms will be selected from the Compustat Annual Industrial File. This restriction in selection is imposed to achieve the significant advantages in data collection afforded by electronic data retrieval. The potential for this limitation to introduce sample selection biases that do violence to the generality of the study is examined, with other limitations of the analysis, in the concluding chapter.

The selection criterion which will eliminate the greatest number of Compustat firms is the necessity for a common fiscal year-end. In the search for commonalities in earnings series of firms which is central to this study, it is essential that the reported earnings numbers relate to the same time periods for all firms. If significant temporal variations are present in the performance of several firms, staggered reporting periods used by the firms will produce earnings series in which this commonality is masked. To minimize the number of firms eliminated by this criterion, firms will be chosen with a fiscal year common to the greatest number of firms. December 31 is by far the most popular fiscal year-end date, so all firms with any other fiscal year-end are excluded from the sample.

A second important selection criterion is that sufficient financial data of each firm must be reported on the Compustat tape to permit calculation of the five independent variables and the two performance measures for the entire 20-year period spanned by the tape (1956-1975), thus providing the greatest possible number of years' observation for use in the analysis. A third criterion is that the sample must exclude firms experiencing merger, quasi-reorganization, acquisition of a subsidiary, or devestiture of a subsidiary during the period for which data are reported. The necessity for this restriction was examined in the development of the growth-performance relation in Chapter II.

One final selection criterion is also necessary. To permit the dependent and independent variables to be measured over a period of years in a meaningful way, i.e., to prevent nominal changes in accounting and reporting from dominating the observed value changes of a variable, it is necessary to exclude those firms whose Compustat data are inconsistent due to reporting changes.

Measurement of the Five Financial Variables

The function of the five financial variables is to serve as bases for forming groups of firms as an alternative to traditional industry groupings. A clustering algorithm, discussed later in this chapter, will be used to form the groups. In the context of clustering in general, the firms will be viewed as objects, and firms with the greatest similarity will be grouped together. Similarity will be measured by the financial variables--the closer are the values of the variables of two firms, the more similar are the two firms.

To serve as input to this clustering procedure, it is necessary that each variable take on one value for each firm. Therefore, the variables will not be measured for each firm for each year, but rather only once for each firm over the entire period of years on which the grouping is to be based.

The performance measures, or dependent variables $E_{k.i,t}$, do not enter into the clustering procedure at all. They will be used only in the final, central step of the analysis--the calculation of market and group indices of performance, and regression of each firm's performance measure series onto the appropriate indices.

The definition of the performance measures do, however, affect the number of years' data available for the analysis, and for that reason must be discussed briefly here. The first form of the performance measure, $E_{1.i,t}$, was defined as an accounting rate of return on invested capital--Net Income minus preferred dividends, divided by common stockholders' equity. Since the numerator of this ratio measures a flow spanning a year in time, while the denominator is a stock measured as of one point in time, the question arises whether the denominator should be measured as of the end of the year or the beginning. Assuming that there is some time lag involved in committing newly-acquired capital to the earning process, and that the desired measure of invested capital is the average amount involved in the earning process, then the beginning-of-year measure is preferred.

This precludes calculating the performance measure in the first year for which data are available, since the beginning-of-year value of common stockholder equity is not available. The first value given is as of the end of 1956, which is to be used in calculating the value of the performance measure for 1957. This leaves 19 years available for use, out of the original 20.

Since the second performance measure to be tested, $E_{2.i,t}$, consists of year-to-year changes in the levels of the first measure, it can be calculated beginning in the second year for which the first measure is available. This form of the dependent variable can therefore be calculated beginning in 1958, leaving 18 years' observations to be used in the analysis, 1958 through 1975.

This time period shall be divided in half to permit the groups of firms formed in the first half to be tested not only as to their ability to account for earnings commonalities in this same time period, but to be tested similarly in the second half. These two testing opportunities represent tests of *ex post* and *ex ante* predictions of earnings commonalities. This technique will also make it possible to examine the stability of membership in the groups formed, since another set of groups can be formed using second-half data and its membership compared with that of the original groups.

These decisions result in a nine-year period of data for use in forming the original set of groups; with this established, it is possible to present the formulas to be used in calculating the values of the five financial variables for each firm, denoted $FV1_i$ (i=1,...,5). Recall that one value of each variable will be computed for each firm for the entire nine-year period, not one value for each year. The Compustat data items used in these formulas are listed in Table I (all tables are presented in Appendix A). Table II presents the formulas in terms of the data item numbers of Table I.

Several alternatives are available for the computation of the financial variables. In the case of a variable measured as a ratio, for example, the nine-year value could conceivably be the average of the nine annual ratios, or the ratio of the average numerator and the average denominator. The nine-year measure of size could be calculated as the logarithm of the average value of total assets, or as the average of the logarithms of the total asset values. In short, when a transformation or arithmetic operation is to be performed in calculating the variable, and some form of nine-year averaging is also necessary, it is not obvious whether the nine-year averaging should occur before or after the arithmetic or transform operation. It can make a difference if one of the nine annual observations is an extreme value. The decision made in this study was that the effect of an extreme value should be minimized. This can be accomplished in the case of the log transform by extracting the log before averaging, and in the case of a ratio by averaging numerator and denominator before dividing. The computational formulas in Table II reflect these decisions.

Values of the subscript t in Table II denote years for which the data are reported on the Compustat tape. Since the definitions adopted for the performance measures eliminate the first two years from the analysis, the first year of the nine-year measurement period is year 3. Beginning-of-year values of balance sheet items are used for the financial variables rather than end-of-year values, which explains the appearance of the subscript for year 2 in the formulas. For example, the nine-year measure of leverage for years 3 through 11 is calculated from balance sheet amounts reported at the ends of years 2 through 10. Income Statement items and dividends, since they are reported at the end of the year to which they relate, carry subscripts 3 through 11. Data numbers in the formulas of Table II refer to the Compustat Item numbers given in Table I.

It was mentioned above that a second set of groups will be formed from the second nine years' data for comparison with the groups formed originally. The formulas for computing the values of these variables, denoted $FV2_i$ (i=1,...,5), to be used in forming this second set of groups, are derived by simply advancing every subscript shown in Table II by nine years. Since this computational change is quite simple, the list of formulas need not be repeated for the second nine years. The formation of this second set of groups, and the uses to be made of it, are discussed in a later section of this chapter.

Preliminary Tests

The five variables to be used as the bases for forming groups of firms as an alternative to traditional industries were developed in Chapter II, and the formulas for their calculation were presented in the preceding section of this chapter. The theoretical relationship between firms' performance and each of the five variables was outlined in Chapter II, for it is these relationships which create a potential for the resulting groups of firms to exhibit the desired performance similarities.

Correlations Between Financial Variables and Performance Measures

Based on the arguments given in Chapter II, predictions were made concerning the sign of the relation between performance and three of the five variables. These predictions will be tested once the values of the variables are calculated for each firm. The object of the test is whether the predicted relation can in fact be observed across firms; for example, whether firms with higher payout rates do exhibit poorer performance.

The test chosen is the Spearman correlation coefficient. The Pearson coefficient, its parametric counterpart, is not appropriate due to the assumptions its use would require concerning the distributions of the variables. A Spearman coefficient will be calculated between each of the two performance measures and each of the five financial variables. The 10 resulting coefficients will be compared in terms of sign and strength with the predictions of Chapter II. Where no prediction is made in Chapter II, a Spearman coefficient close to zero will be taken as support of the indeterminate relation. Where predictions were made, a non-zero coefficient with a sign in agreement with the predicted sign will be taken as supportive.

In conducting the Spearman test, each firm will provide one value of each financial variable and one value of each performance measure. The coefficient will therefore represent the strength of their relationships across firms, since each firm provides one observation. Values of the financial variables are provided by the formulas of Table II. The two performance measures must also be calculated as single nine-year values for each firm. These nine-year measures will be used only in this Spearman test; in the regression of each firm's performance series onto market and group indices, which constitutes the central test of the study, each firm's series of annual values of performance measures will be used.

The formulas for calculating the nine-year values of performance measures are presented in Table III (see Appendix A). The same decisions on the effects of extreme annual values that were faced in calculating the financial variables are faced here; they are resolved in the same fashion. Table III also presents formulas for the annual values of the performance measures as developed in Chapter II; these values will be used only in the later portions of the analysis for regression onto the market and group indices. Since annual values will be calculated for each firm, the subscript denoting year is given in its general form as t. For the nine-year measures, only one value is calculated for each firm, so its t subscript takes on values corresponding to the first 11 years of Compustat data. Like the formulas of Table II, those of Table III are stated in terms of the Compustat data items listed in Table I.

Analysis of Variance

It is the hypothesis of this study that groups formed from a sample of firms on the basis of similarity with respect to the five financial variables may exhibit earnings commonalities resulting from the potential of the five variables to influence earnings. Implicit in this reliance on grouping is an assumption that the five variables may not only influence performance directly, but that they may interact to produce further effects. Otherwise, regressing earnings on six independent variables--market index and five financial variables--would stand to reveal these earnings influences most directly. But if the five variables do produce interactive effects, then groupings of firms based on the five, with calculation of group indices of earnings, will not only stand to capture the interactive effects, but will do it in a way comparable to traditional industry groupings.

This expectation of interactive effects can be tested. The mean value of each variable calculated across all firms will be used to rank each firm's value of that variable as either high or low. The five variables will then be used as treatments in an analysis of variance, so the significance of interactive effects can be determined. The hypothesis to be tested is whether the interactive effects of the levels (either high or low) of the five variables produce differences in the dependent variable (the performance measure) larger than can be attributed to chance.

These two preliminary tests will be conducted to insure that the five chosen variables are not included at whim. The theoretical support for each variable's use given in Chapter II is the first step in establishing the reasonableness of the variable's posited relation with firms' performance. The two preliminary tests described in this section will either support or contradict the arguments given, and, if supportive, will more firmly establish the predicted relationships via empirical evidence.

Formation of the Groups

The central question addressed in this study concerns the relative performance of the three forms of the G index in the two-index model of firms' earnings. Each form of the G index simply represents a different way of forming groups of firms. For brevity, it is desirable to assign each method of grouping a name. Therefore, groups consisting of traditional industries will be referred to as industries; groups formed by random selection of firms will be called random groups or simply "random"; and the groups formed on the basis of similarity with respect to the five financial variables, since they will be formed via a clustering algorithm, will be called clusters. This terminology is perhaps imprecise, since all three types of groups could be described as clusters, but it will suffice for convenience of exposition.

Many choices are possible in determining the number of firms to constitute a group and the number of groups to be formed by each method. Clusters and random groups are versatile--virtually any number of groups of any size can be formed. Industry grouping does not possess this versatility. Compustat three-digit industries, which correspond very closely to SIC three-digit industries, are chosen for used in this study; however, they provide a limited number of possible groups with a limited number of firms per group. Varying the size and number of groups from one grouping method to another may have undesirable effects on the observed performance of the three methods. If, for example, 10 industries are selected with 15 to 20 firms in each, and are compared to 20 clusters of 8 to 10 firms each, how much of an observed difference in the performance of the group indices in accounting for earnings variability is due to the difference in average group size, how much is due to the difference in number of groups, and how much is due to the effect of real interest?

To avoid this indeterminancy in interpreting results of the analysis, all three methods will be restricted to forming the same number of groups, with the same numbers of firms in each group. Since the industries selected for use establish an inflexible number and size of groups, the other two methods will be adjusted accordingly.

It is necessary to impose a further restriction on the formation of groups, that of a minimum number of firms per group. Some floor on group size is necessary to distinguish group effects in performance from the unique characteristics of each firm. This problem was addressed in Brown and Ball (15), where a minimum of 10 firms per industry was imposed. To achieve comparability of results between this study and that of Brown and Ball, the same minimum is chosen here for all three grouping methods.

From the available sample of firms, three-digit industries will be selected which contain at least 10 firms in the sample. The resulting number and size of industries will determine the number and size of clusters and random groups to be formed. For example, if the sample provides four industries of 12 firms each, and four of 10 firms each, then the clustering algorithm will be restricted to forming four clusters of 12 and four clusters of 10. Random groups will be formed in exactly the same pattern.

With this method established for selecting industries and for determining the number and size of clusters and random groups, the specifics of the random selection and the clustering will now be delineated.

Random Groups

The available sample of firms chosen from those listed on the Compustat file will be assigned serial numbers beginning with 001 in the order they appear in the file. The highest serial number assigned will be equal to the number of firms in the sample. To achieve randomness, a second listing of the same serial numbers will be made. The order of the numbers in this list will be determined by the order in which they appear in the first three digits of the random number table given in Appendix A-B of Canada (17). The starting point in the random number table will be determined by the first entry of the first column. Its first digit will determine the row, and its second digit the column, at which to begin the search for the three-digit serial numbers assigned to the sample firms. The search will continue until enough serial numbers are selected, without repetition, to equal the number of firms contained in the selected three-digit industries. To form the first random group, the first n serial numbers in the second (random) list will be selected, where n is the number of firms in the largest three-digit industry selected. These n serial numbers identify the n firms to form the first random group. The next m serial numbers will identify the second random group, where m is the number of firms contained in the second-largest three-digit industry. In this fashion, random groups will be formed which correspond exactly in number and size to the industries.

Clusters

The groups of firms proposed in this study as an alternative to traditional industries will be formed on the basis of firms' similarity with respect to the five chosen financial variables. Differences between two firms' values of the variables will be used as the measure of dissimilarity or distance between the two firms, and will thus determine whether the two firms will belong to one of the groups.

It is conceivable that some of the five variables may measure in common a particular earnings-influencing characteristic of firms. This is equivalent to saying that there are underlying constructs which, when shared by several firms, produce commonalities in the earnings of these firms. It is hypothesized in this study that the five chosen financial variables are manifestations of some of these underlying constructs, but not necessarily in one-to-one relationships. Two or more of the five variables may to some extent be measuring the same earnings-influencing construct. The presence of correlations among the five variables would be consistent with this supposition.

To the extent that the five financial variables are correlated with each other, the possibility exists that they are measuring the same construct. A clustering of firms based on the five variables would assign implicitly heavier weights to such a construct, since it would essentially enter the computation of inter-firm dissimilarity more than once. The objective is to cluster the firms based on underlying and possibly unknown earnings-influencing constructs. These characteristics are believed to be manifested in the joint information contained in the five financial variables.

The appropriate technique to permit clustering based on the underlying constructs, but without redundant measurement that may be present in the five variables, is factor analysis. The five financial variables therefore will be factor analyzed, each firm providing one set of observations on the five variables. The initial set of factors will be extracted by the method of principal components. These factors will then be rotated by the varimax method to produce a set of rotated factors. The principal components could conceivably be used as the clustering variables, since, by using only the principal components with eigenvalues greater than one, parsimony of data would be achieved, thus reducing the number of variables used in the clustering algorithm. The factor rotation will be performed, however, to achieve another of the advantages of factor analysis--simple structure. The rotated factors will each be highly correlated with relatively few of the five financial variables, unlike the principal components, which would be much more difficult to interpret in terms of the five financial variables. The value of simple structure consists of this ease in interpretation of the rotated factors.

For each firm, rotated factor scores (RFS) will be calculated for those factors with eigenvalues greater than one, and these RFS will be used as the inputs to the clustering algorithm. Only those rotated factors with eigenvalues greater than one will be used, as this is the traditional criterion for achieving parsimony (27).

With the RFS used as inputs to the clustering algorithm (one score on each rotated factor calculated for each firm in the sample), the only restriction on the formation of clusters is that the number of clusters formed and the numbers of firms in each must correspond to the number and size of groups formed by the other two methods. If, for example, eight industries are available in the sample, four of 10 and four of 12 firms, then four clusters of 10 and four of 12 will be formed to achieve the best comparability of results.

Continuing the example, the objective is to produce eight groups out of a sample of, say, 400 firms, such that the similarity exhibited by firms within each group is greater than that of any other possible set of groups of the same size. In terms of the number of possible groupings to be examined in an exhaustive search, the clustering task is quite formidable.

Four groups of 10 and four of 12 firms comprise a total of 88 firms. But the clustering procedure must not only select the "best" 88 out of the sample of 400, although this alone would require examination of 400!/(400-88)! possible solutions. The algorithm must also examine all possible ways of dividing the 88 into four clusters of 10 and four of 12. This second requirement would furthermore affect the choice of the "best" 88 in the first step. The "best" 88 are not the 88 most like each other, since it is not desired to produce one group of 88 similar firms. The "best" 88 are those which can be divided into the "best" four groups of 10 and four of 12 firms. If an algorithm were available for defining the billions of possible solutions, the computation time required to select the one with the most similar groups of firms would be prohibitive--similarity or dissimilarity would have to be measured, based on values of the RFS, for each one of the possible ways of dividing the 88 into eight groups, and all this for each possible 88 out of the 400.

An algorithm for formation of optimal clusters is therefore impractical. This problem is not new to cluster analysis. Discussion of the infeasibility of optimal clustering is found, for example, in Hartigan (28), Young and Calvert (47), and Everitt (20).

An acceptable solution is provided by hierarchical clustering. The n objects to be clustered, firms in this case, are viewed initially as n clusters, each containing only one object. At each step in the clustering, those two clusters most similar to each other are combined into one new cluster. After m steps, n - m clusters remain, at least one of which contains more than one object. After n - 1 steps, only one cluster remains, containing all n objects. A measure of aggregate within-cluster dissimilarity, or error, may be calculated at each step, revealing the increased heterogeniety of clusters as larger and larger clusters are formed. Such a measure equals zero before the first step, since the initial single-object clusters possess no intra-cluster dissimilarity. The maximum value of the error is attained after the final step in which a single cluster of all n objects is formed. Abrupt increases in the error measure indicate natural clustering stages.

Such an error measure can also provide a criterion for deciding which pair of clusters should be combined at each step, out of all possible pairs existing at that time. That choice which causes the least possible increase in the error measure can be taken as indication of the two most similar clusters. The error measure may thus provide the clustering criterion, as well as signaling the analyst that a desired level of aggregation may have been reached.

A hierarchical clustering algorithm will be used to form the clusters of firms in this study, using the RFS as the clustering variables. Inter-firm differences in the RFS will be taken as the measure of inter-firm dissimilarity. Beginning with the entire available sample of firms, the algorithm will proceed until a cluster the desired size or larger is formed. If, for example, the largest industry contains 12 firms, the first cluster formed containing at least 12 firms will signal the desired level of clustering. From this cluster will be chosen exactly 12 firms which are most alike with respect to their RFS. The first cluster will consist of these 12 firms.

The remaining n - 12 firms will then be resubmitted to the clustering algorithm, which will repeat the analysis beginning with n - 12 single-firm clusters. Once a cluster is formed at least as large as the secondary-largest industry, the firms in this cluster will be used to select the exact number needed for the second cluster. In this fashion, the number and size of clusters can be tailored to exactly correspond to the number and size of industries and random groups.

The HGROUP algorithm, which is given in Veldman (44), was chosen to serve as the hierarchical procedure in this study. At each clustering stage, the HGROUP program names the two clusters combined at that stage, the number of objects in each of the two clusters, and the serial (input) numbers of the objects. These characteristics of HGROUP make it well suited for the purposes of this study. The number of firms in the new cluster formed at each stage will indicate when a cluster of the desired minimum size has been formed. Identification of the firms in this cluster will facilitate selection of the exact number of firms needed.

HGROUP calculates an error term at each stage, consisting of the within-cluster variance totaled over all clusters existing at that stage. The two clusters to be combined at the next stage are chosen to produce the least possible increase in this within-cluster variance, or error. With the RFS provided as input to the algorithm, the clustering criterion can be stated as minimization of the value of the following expression, where the bar over RFS denotes the mean value of a RFS calculated over only those firms belonging to a particular cluster:

$$\text{Error} = \sum_{c=1}^{C} \sum_{f=1}^{F} \sum_{i=1}^{I} (RFS_{cfi} - \overline{RFS}_{ci})^2 .$$

The particular cluster is denoted cluster c, and the number of clusters that will remain after two are combined is denoted C; a firm is denoted f, and the number of firms in the particular cluster is denoted F, where F will vary from one cluster to another; a RFS is denoted by i, and the number of RFS used in the algorithm is denoted I (the value of I may be as few as one or as many as four, depending on the number of rotated factors which are found to have eigenvalues greater than one in the factor analysis).

Once a cluster is formed containing the minimum desired number of firms, the mean value of each RFS for this cluster will be calculated. Each member's Euclidean distance from the cluster mean will be computed, and the most extreme outliers will be eliminated from the cluster until the exact number of firms needed remains.

The supposition that this procedure for forming clusters will result in reasonable approximations to optimal clustering is based on the large number of firms in the available sample from which clusters will be formed, and on the behavior of the HGROUP error term at early stages of clustering hierarchy. If the sample contained just enough firms to form the needed clusters, so that all firms would have to be placed in one of them, this use of a hierarchical procedure would likely produce poor approximations of optimal clusters. But this begs the question, since in such a small-sample case the optimal clustering could be found by examining all possible segmentations of the sample. It was the sample size's prohibition of optimal clustering that led to the choice of the hierarchical approximation.

With a sample of several hundred firms available, it is plausible that the first cluster formed with at least the dozen or so members needed to match the size of the largest industry will in fact be a close approximation to an optimal cluster. This is supported by the behavior of the error term at early steps in the hierarchy. Infinitesimal increases in the error term consistently occur in the first few steps of clustering from a large sample, and the number of steps through which this phenomenon continues grows as the size of the sample grows. This is consistent with priors concerning the distributions of large samples. In a sample of 10 objects, it is likely that some two are quite similar; in a sample of 50, it is virtually certain. In a sample of 100, a reasonably compact cluster of 10 or 12 should frequently occur; in a sample of 300 or 400 objects (like that of this study) it is virtually certain.

The large sample of firms available and the behavior of the error measure at early clustering stages are relied upon to produce reasonable approximations at an optimal clustering. Since the hierarchical algorithm will be used only until one group of the desired size is formed, only early clustering stages will ever be used. The number of firms available to the algorithm in forming the next cluster will be less by the number in the last cluster formed, but the desired cluster size will also decrease as each cluster is formed, since the largest clusters will be formed first.

An alternative method of clustering firms has received attention in the literature, and deserves mention here. This is the method used by King (31) and Meyers (38) consisting of a factor analysis of firms' security return series, with the firms viewed as the variables to be factor analyzed. Several firms with high loadings on the same factor are identified as belonging to a cluster. In this fashion, factor analysis serves as a clustering technique, and is called Q-factor analysis.

Since the first factor explains as much as possible of the variability present in all the firms' return series, it is an estimate of the market effect. Succeeding factors identify clusters or industries upon inspection of the matrix of factor loadings, as already mentioned. Values of factor scores can be used as the market index (the score on the first factor) and cluster or industry indices (the scores on the second and succeeding factors).

Although from a statistical standpoint this technique is tailor-made for the purposes of the research undertaken here, it has been rejected. The principal reason for its rejection is the ad hoc nature of such an analysis. It appears circular to form groups of firms based explicitly on inter-firm similarities observed in a performance measure, and then to use these groups in an effort to explain the very same similarities. This analysis might suceed in accounting for the performance similarities, but it would do so by explicitly considering these similarities in the first place. The reasons for the success of such an analysis would have no conceptual basis--it could only be said that the groups of firms that best account for performance similarities are those groups which exhibit the greatest such similarities.

Livingston (33) has documented other undesirable qualities of Q-factor analysis. First, he found that factor analysis is so sample sensitive that different samples from the same population produce such different estimates of market and industry effects in security returns that the results of the analysis lack generality. Second, he illustrated that different varieties of factor analysis yield significantly different estimates of industry effects when applied to the same data. Third, he generated return series composed of seven orthogonal market and industry factors for a sample of artificial firms, and then performed a principal components analysis. There were negative correlations between firms' residual series where zero correlations had been deliberately "built in" the original data. The second through seventh components identified by the analysis were combinations of the "true" components, and the seventh was found to be insignificant.

Livingston (33) further illustrated that regression onto a broadly-defined market index, like that used in this study, avoids the disadvantages of the Q-factor technique, and gives better estimates of the correlations between firms' residual series.

Calculation of Market and Group Indices

The market index for each firm will be calculated as an average of the dependent variable's values over all other firms in the sample, as stated in the introduction to this chapter. This avoids the spurious relationship that would result if the firm were included in the calculation of its own index.

The averaging method chosen is that based on common stockholder equity of each firm. Rather than a simple mean of the dependent variable over all other firms, each firm's value will be weighted by the ratio of its common stockholder equity to the total common equity of all the firms used in the calculation. This index construction scheme can be interpreted as emphasizing the relative sizes, rather than the number, of firms in the market. Its use in this study is based on the finding of Gonedes (25) that it produced the most valid model of firms' earnings numbers.

Since two forms of dependent variable are used, levels and changes, two forms of the market index will be calculated, one based on each form of the dependent variable. Therefore, in regressions where the dependent variable is levels of earnings, the market index will be the equity-weighted average of all the other firms' levels of earnings; where the dependent variable is changes in the earnings measure, the market index will be the equity-weighted average of all other firms' changes in earnings.

Calculation of the group index will consist of two steps, as stated in the introduction to this chapter. The second step removes from the group index the influence of the market index, so that the two indices used in the regression will not exhibit multi-collinearity.

First, a "raw" group index will be calculated for each firm. Just as the market index is an average of the dependent variable over all other firms in the sample, the raw group index will be an average of the dependent variable over all other firms in the group to which this particular firm belongs. Little evidence is available concerning the relative usefulness of alternative index construction schemes for groups of firms. Therefore, two schemes will be tested here. In one, the group index for a particular firm will consist of the simple arithmetic mean of the dependent variable's values over all other firms in the group. In the second scheme, each of the other firms' values will be weighted by common stockholder equity. These two forms of group index will be referred to as the simple and weighted indices. Their performances will be compared in this study.

Having calculated the raw group index for each firm in the group (whether the group is an industry, random group, or cluster) in both simple and weighted forms, the group indices will be redefined as that portion of the raw indices not explained by the market index. Each firm's simple group index will be regressed onto the firm's market index, and the resulting residuals will be used as the final form of the simple group index. The firm's weighted group index will also be regressed onto the firm's market index, and the residuals from this regression will constitute the final form of the firm's weighted group index. These regressions are described in equation form in the concluding summary of this chapter.

Regression of Performance Measures onto Indices

Seven sets of regression equations will be fitted, with quartile R^2 values calculated from each set. Clusters formed using $FV1_i$, the first nine-year values of the five financial variables, will determine the firms and group indices for the first set. The second set of equations will be fit for the firms in the three-digit industries, with the industries determining the values of the group index. Random groups will define the third set of equations. These three sets of regression equations will be fitted using the first nine years' values of the dependent variables $E_{k.i,t}$ and, of course, the corresponding years' values of the market and group indices.

The groups of firms used in the first three sets of equations will also be used to fit equations to the second nine years' values of the dependent variables and corresponding values of the firms' market and group indices. In this fashion, the clusters of firms formed from the first nine years' values of the financial variables will be compared to industries and random groups not only in terms of their performance in accounting for commonalities in earnings of the same nine-year period, but also in the succeeding nine years. These two time periods tested constitute ex post and ex ante tests of the relative performance of the three methods of forming groups of firms.

The seventh set of regression equations to be fitted will utilize new clusters of firms formed on the basis of $FV2_i$, the second nine-year period's values of the five financial variables. These equations will be fitted to the values of dependent variables and corresponding indices of the second nine-year period, constituting an ex post test of cluster performance in the second nine years. A corresponding test of industries would be identical to the second nine-year equations already described above for industries, so no new equations need be fitted for industries. A corresponding test of the random groups' second nine-year performance need not be conducted either, since the second nine-year test described above for original random groups will serve both functions simultaneously.

Each of the seven sets of equations will be fitted in two forms: one using the market index as the sole independent variable, and one using both market and group indices as independent variables. The quartile values of R^2 over all firms will then be used to estimate the relative strengths of cluster vs. industry vs. random effects in firms' earnings. The increase in the R^2 when the group index is used in the equation, compared to when the market index alone is used, measures the portion of the variability in the dependent variable accounted for by the group index, but not correlated with the market index. By calculating this difference for each type of grouping scheme used, the performance of industries relative to clusters and random groups can be evaluated.

For each firm in a group (whether the group is an industry, a cluster, or a random group), six regression equations will therefore be fitted: one with market index alone, one with market and simple group indices, one with market and weighted group indices, and these same three with each form of dependent variable (levels and changes). Thus, if the industries contain a total of 100 firms in the sample, then each of the seven sets of regression equations will contain 600 regressions to be fitted, producing a total of 4200 regressions.

Tests of Residuals

The final step in the analysis will consist of examination of residuals from the regressions described in the preceeding section. As stated there, seven data sets will be used: industries, random groups, and the first set of clusters will each be used in both nine-year periods, plus the second set of clusters in the second period. Within each data set, six forms of regression equation will be fitted for each form: one with market index only, one with market and simple group indices, one with market and weighted group indices, and these three for each of the two forms of performance measure, levels and changes. These seven data sets and six forms of equation produce 42 combinations. With n firms in each data set, the 42 combinations will result in 42 sets of residuals, each set containing n series of nine annual residual values each. In addition, the original earnings measures, levels and changes, are also available for each of the seven original data sets, which comprises an additional 14 sets of values, each containing n series of nine annual observations each.

For each set of n series of values, a correlation coefficient will be calculated between each possible pair of series. With n series available, n(n-1)/2 distinct pairs exist. These correlation coefficients may be conveniently represented as the above-diagonal elements of an n by n matrix. The element in the i^{th} column and j^{th} row is the correlation coefficient between the series of values of the i^{th} and j^{th} firms. A total of 56 such matrices of correlation coefficients will be calculated, 42 from the regression residuals and 14 from the original earnings values.

The purpose of the examination is to determine the extent to which each matrix of coefficients differs from an identity matrix. As stated in Chapter I, an ideal result would be to produce perfectly uncorrelated residual series for a sample of firms, in which case the correlation matrix would be exactly equal to an identity matrix. However, the presence of remaining correlations among even small numbers of firms in the sample will prevent the achievement of this ideal. Therefore, it is certain that none of the observed correlation matrices will be exactly equal to identity matrices. The relevant question then becomes whether one type of residual, such as that resulting from use of clustered groups of firms, exhibits a correlation matrix which is not <u>significantly</u>

different from an identity matrix, while the correlation matrices corresponding to industry groupings, for example, are significantly different from identity matrices.

Use of the 14 matrices calculated from the original earnings measures in addition to the matrices calculated from regression residuals will produce some evidence on whether abstracting only the market effect achieves substantial reduction in inter-firm correlation. Comparisons will also be possible between the two forms of performance measure and the two index construction schemes, as well as the three methods of forming groups of firms.

Bartlett's sphericity test, presented in Cooley and Lohnes (18), provides a chi-square statistic for testing the hypothesis of zero inter-firm correlation at a given confidence level. The hypothesis of independence can be rejected if the calculated statistic is greater than the value of chi-square for n(n-1)/2 degrees of freedom at the desired confidence level. The calculated statistic is given by

$$-((n-1) - (2p+5)/6)\log_e D$$

where p is the number of observations over which each correlation coefficient is calculated, nine in this study, and D is the determinant of the correlation matrix. Lawley (32) offers an alternative formula for computation of the statistic, in which $\log_e D$ is replaced by the sum of squares of correlation coefficients above the diagonal.

In addition to the chi-square test of inter-firm independence, another descriptive statistic will be calculated for each of the 56 correlation matrices. Since all values will be between one and minus one, the mean of each matrix's values will approach zero as successive commonalities are removed from the residual series. For the original earnings measures, correlations are expected to be predominately positive, since not even the market effect will have been removed. However, a mean correlation of zero could result from many extreme positive values offset by many extreme negative values, in which case the zero mean is a very inaccurate representation of the overall degree of correlation present among the firms' residual series. Therefore, the mean absolute value of each matrix's above-diagonal values will be calculated as the second descriptive statistic.

A final characterization of each correlation matrix will consist of a frequency distribution of its elements. Since all elements are between one and minus one, intervals of one-tenth will provide 20 classes into which the elements of each matrix will fall. The number of correlation coefficients in each of the 20 intervals will be tabulated for each of the 56 correlation matrices, using only the above-diagonal elements.

Summary

From the Compustat Annual Industrial File will be selected all those firms which meet the criteria delineated in the section of this chapter on sample selection. For each of these firms, the following values will be calculated:

i) the five financial variables $FV1_i$ (i=1,...,5), measured over the first nine years of the available 18--not one value of each variable for each year, but one for the nine-year period.

ii) the five financial variables measured over the second nine-year period, represented by $FV2_i$ (i=1,...,5).

iii) levels and changes in performance measured over the first nine-year period, to be used only in calculating the Spearman coefficients and in performing the analysis of variance.

iv) annual values of the levels and changes in performance, $E_{1.i,t}$ and $E_{2.i,t}$, for each of the 18 years.

Formulas for the calculation of i) and ii) are presented in Table II; for iii) and iv), in Table III.

This study's main thrust is directed toward the hypothesized relations between firm performance and the five financial variables, FV_i, which can be presented in a simple equation form: performance = $f(FV_i)$. The methodology consists of the following steps:

1. Spearman rank correlation coefficients will be calculated across firms to test the hypothesized relation between performance and each of the five financial variables $FV1_i$. Since two forms of performance measure are used, as presented above in iii), 10 coefficients will be calculated. Comparison with the predicted relations between performance and the financial variables will constitute a test of these theoretical relationships.

2. Each of the two nine-year measures of performance will be used as the dependent variable in an analysis of variance. The five financial variables will constitute the experimental treatments, each assigned a value of either high or low for each firm. The mean of each variable over all firms will determine whether a firm's value of that variable is designated high or low. Significant interactions of the five treatments will support their use in forming groups of firms to explain performance commonalities.

3. All three-digit industries will be selected which contain at least 10 sample firms. These groups will constitute the traditional industries in both the first and second nine-year periods. The number of industries thus selected, and the numbers of firms they contain, will determine the number and size of groups to be formed by the other two methods.
4. Firms will be selected at random to form groups corresponding exactly in number and size to the industries. These will serve as the random groups in both nine-year periods.
5. Clusters will be formed on the basis of similarity with respect to scores on the rotated factors of the $FV1_i$, corresponding exactly in number and size to the industries and random groups. A second set of clusters will be formed using scores on the rotated factors of the $FV2_i$. It is likely that some of the same firms which make up the industries will also be contained in random groups, or in clusters, or even in both.
6. Eighteen annual values of the market index will be calculated for each firm. Since two forms of performance measure E_k are used, a market index for each firm must be calculated for each E_k. The indices' values will be the equity-weighted averages of E_k calculated over all other firms in the sample. The groups of firms will have no influence on the values of these market indices. Where CSE_j denotes the common stockholder equity of the j^{th} firm, the market index for each of the two E_k for firm i is given by

$$\frac{\sum_{j \neq i} (E_{k,j})(CSE_j)}{\sum_{j \neq i} CSE_j}$$

In this and the following index formulas, the time subscript is suppressed, since each term in all formulas refers to any year t.

7. For each firm placed in a group by any of the three grouping methods, "raw" group indices will be calculated. If a particular firm is a member of, say, an industry and a random group, two entirely different group indices will be calculated; in other words, a firm's group index is not determined by the firm, but by the group to which the firm belongs. Again, distinct indices must be calculated for the two distinct E_k. Also, both simple average and equity-weighted average indices will be used, to test the relative performance of the two averaging schemes. Simple index values for firm i are given by the following, where J is the number of firms in firm i's group, and j denotes only those firms in that group:

$$\frac{\sum_{j \neq i} E_{k.j}}{J - 1}$$

Values of the weighted index are given by:

$$\frac{\sum_{j \neq i} (E_{k.j})(CSE_j)}{\sum_{j \neq i} CSE_j}$$

8. The "raw" group indices calculated in step 7 will be used as dependent variables in regressions with the corresponding market index as the independent variable. Denoting the raw indices from step 7 as GR, and the market index of the firm as M, the equation to be fitted is

$$GR_t = a + bM_t + G_t \tag{5}$$

The equation will be fitted once for each firm, for each of the four firms of GR (two forms of E_k times two index construction schemes--simple and weighted). The residuals, G, will be used as the group index's values in fitting the final regression equation given in the introduction to this chapter as equation (4), and repeated in a more general form below as equation (6).

9. Seven sets of regression equations will be fitted. Each set will correspond to one grouping of firms in either the first or last nine years of the 18-year period. Industries, random groups, and the clusters formed from FV1 are available in both nine-year periods, providing six of the seven data sets. The seventh is provided by the clusters formed from FV2. Within each set, the same number of equations will be fitted, since the number of firms was held constant across the three grouping methods. For each firm, six regression equations will be used: one with the market index as the sole independent variable, one with market and simple group indices, and one with market and weighted group indices, for a total of three; and each of these three for each form of E_k--levels and changes. Quartile R^2 values taken across each set of firms will be used as measures

of the relative usefulness of the two index construction schemes, the two forms of E_k, and the three methods of forming groups of firms. The equations to be fitted for each firm are presented below as a general form of equation (4). The value of k determines the form of performance measure used, levels or changes. Observations are taken across time, t, for each firm i. The x subscript on the G index indicates that one of three methods of grouping is used in calculating the group index--cluster, industry, or random; the s subscript denotes the two index construction schemes, simple and weighted. The general form of the first four equations to be fitted for each firm is

$$E_{k.i,t} = a'_{i,t} + b_{m,k,i}M_{k,i,t} + b_{g,k,i}G_{x.k,s,i,t} + e'_{k,i,t} \qquad (6)$$

Since values of market and group indices are unique to each firm, they are now subscripted with the i, unlike their form in equation (4). All terms are subscripted with a k to denote the form of performance measure E_k, also unlike equation (4).

For each firm, equation (6) represents four of the six regression equations to be fitted--two values of k times two values of s. The fifth and sixth forms use only the market index, and are given below for k = 1 and k = 2.

$$E_{k.i,t} = a''_{i,k} + b_{k,i}M_{k,i,t} + e''_{k,i,t} \qquad (7)$$

10. For each of the 42 sets of regression residuals (seven data sets times six forms of regression equation), a matrix of

inter-firm correlation coefficients will be calculated. An additional 14 such matrices will be calculated from the values of the original earnings measures (seven data sets times two earnings measures--levels and changes). For each of the 56 matrices, the mean absolute value and a frequency distribution will be calculated using above-diagonal elements. Bartlett's sphericity test will be used to assess the significance of each matrix's difference from an identity matrix.

CHAPTER IV

RESULTS

The selection criteria stated in Chapter III produced a sample of 359 firms. Appendix B consists of a list of these firms, identified by Compustat industry number, industry name, company number, and company name. They are also assigned serial numbers 1 through 359, by which they will be identified throughout the remainder of this text.

Preliminary Tests

Values of the $FV1_i$ and the two nine-year performance measures were calculated for each of the 359 firms according to the formulas of Tables II and III (see Appendix A). A Spearman rank correlation coefficient was calculated across firms between each of the two performance measures and each of the five $FV1_i$. Table IV of Appendix A presents the 10 Spearman coefficients in two columns, one for the correlation with the level of performance, and one for that with the change in performance. The third column gives the sign of the predicted relationship developed in Chapter II.

Overall, the coefficients of Table IV are in support of the predicted relationships. The two financial variables whose relationships with performance were considered indeterminate, leverage and liquidity, have Spearman coefficients relatively close to zero. The negative relation predicted for payout and the positive relation for growth are well supported by the sign and magnitude of coefficients. The coefficients for size have opposite signs from those predicted, but are extremely close to zero for both forms of performance measure.

For the second preliminary test, the mean value of $FV1_i$ was calculated across the 359 firms. Each firm's value of each variable was classified as high or low in relation to the mean. The five variables were then used as treatments, each with high and low levels, in an analysis of variance of the nine-year measures of performance. F statistics were calculated for the ratio of each interaction sum of squares to the residual, or error, sum of squares. The result of interest is the probability that a greater F could be observed by chance. Eleven interaction terms were associated with probabilities of .10 or less, which provides evidence that significant performance influences result from interactions of the $FV1_i$, as hypothesized in Chapter III. These 11 interaction terms and the associated probabilities of greater F values are presented in Table V of Appendix A.

The Groups

Industries

Eight three-digit industries containing 10 or more firms each were found in the sample of 359 firms; they contain a total of 107 firms. These industries are assigned the symbols I1 through I8. Table VI of Appendix A lists the industry numbers, industry names, number of firms contained in each, and the serial numbers of the member firms. All the other firms of the original 359 belong to industries with fewer than 10 members. Serial numbers refer to the order in which the firms are listed in Appendix B.

Random Groups

Firms were selected by serial number, using a table of random numbers given in Canada (17), to form groups corresponding exactly in size to the industries. The first 18 serial numbers found in the table identified the first group, the next 17 identified the second group, and so on. The eight random groups thus formed were assigned symbols R1 through R8. Table VII lists the number of firms in each group, and the serial number of each member firm.

Clusters

The first step in formation of the clusters consisted of a factor analysis of the five $FV1_i$ to obtain rotated factor scores for input to the clustering algorithm. Two rotated factors were found to have eigenvalues greater than one, and were therefore chosen for use in the clustering. These two rotated factors are denoted $RF1_i$ because they are formed from the $FV1_i$; the rotated factors formed later from the $FV2_i$ will be denoted $RF2_i$. The $RF1_i$ consist of the following linear combinations of the $FV1_i$:

$$RF1_1 = 0.43893FV1_1 - 0.04001FV1_2 - 0.48260FV1_3 + 0.43587FV1_4 + 0.09291FV1_5$$

$$RF1_2 = 0.20428FV1_1 - 0.52003FV1_2 - 0.02241FV1_3 - 0.13733FV1_4 + 0.66412FV1_5$$

The factor matrix is presented in Table VIII Of Appendix A. Inspection of the loadings reveals convenient interpretations of the $RF1_i$. Since $RF1_1$ is highly positively correlated with leverage and growth, and highly negatively correlated with payout, it could be described as an aggressiveness factor--high leverage, fast growth, and a tendency toward earnings retention. $RF1_2$ is highly positively correlated with size and highly negatively correlated with liquidity, and is therefore rather difficult to interpret. Interpretation of factors is, of course, largely subjective, and forunately it is not important for the purposes of this research.

$RF1_1$ and $RF1_2$ were used as the clustering variables in the HGROUP algorithm. The first cluster formed of 18 or more firms terminated the hierarchy. This process was repeated with the remaining 341 (359 minus 18) firms to find 17 firms for the second cluster. The eight necessary clusters were identified in this fashion, and were assigned symbols C1.1 through C1.8. The integer 1 common to all the symbols signifies that the clusters were formed from the $RF1_i$. The second set of clusters formed later in the analysis from the $RF2_i$ will be named C2.1 through C2.8. Table IX of Appendix A lists the numbers of firms in each cluster and the serial numbers of the member firms. Serial numbers refer to the order in which the 359 sample firms are listed in Appendix B.

The clusters C1.i are interpretable in terms of the $RF1_i$. Inspection of $RF1_i$ values of the firms in the eight clusters reveals the intra-group similarity on which the clusters were formed. Table X of Appendix A presents each cluster's values of the $RF1_i$, characterized as very high, high, moderate, low, and very low.

The $FV2_i$ were factor analyzed to provide rotated factors for formation of the second set of clusters, C2.i. Three rotated factors of the $FV2_i$ had Eigenvalues greater than one. They will be denoted $RF2_i$, and consist of the following linear combinations of the $FV2_i$:

$$RF2_1 = 0.01032FV2_1 + 0.64462FV2_2 + 0.05462FV2_3 + 0.02987FV2_4 - 0.58275FV2_5$$

$$RF2_2 = 0.00587FV2_1 + 0.12216FV2_2 + 0.64702FV2_3 - 0.56273FV2_4 + 0.10870FV2_5$$

$$RF2_3 = 0.76586FV2_1 + 0.25320FV2_2 + 0.19741FV2_3 + 0.27238FV2_4 + 0.29907FV2_5$$

The factor matrix for the $RF2_i$ is presented in Table XI of Appendix A. $RF2_1$ is highly negatively correlated with size and highly positively correlated with liquidity, but has very small loadings associated with the other three variables. It can therefore be described as simply a smallness-liquidity factor. $RF2_2$'s correlations label it as a high payout-nongrowth factor; and $RF2_3$ is essentially a measure of leverage. Comparison of the factor matrix for $RF2_i$ with that for $RF1_i$ reveals the instability of factor analysis results, as mentioned in Chapter III.

The entire clustering procedure was repeated using the $RF2_i$ as the clustering variables. The eight clusters that resulted are denoted C2.1 through C2.8. The number of firms in each of the C2.i and the serial numbers of the member firms are given in Table XII of Appendix A.

Like the first set of clusters, each of the C2.i is interpretable in terms of the $RF2_i$ upon inspection of the factor scores of its member firms. Table XIII of Appendix A gives each cluster's values of the $RF2_i$, characterized as very high, high, moderate, low, and very low.

This second set of clusters was formed for two purposes, to permit an *ex post* test of the strength of any earnings commonalities in the second nine years, and to reveal the degree of stability in the membership of clusters formed across different time periods. The first of these purposes will involve an additional set of 107 regressions, and tests of correlations in the resulting residuals. The second of these purposes is addressed by inspection of the serial numbers of firms in each of the two sets of clusters to determine whether several firms clustered together in one time period were also in one cluster in the other period.

There is evidence of a small but noticeable degree of stability in cluster membership. In each set, the eighth cluster contains three chemical firms out of its total of 10 firms. Two of these three chemical firms are common to both clusters. Of the other seven firms in each of these two clusters, one firm is common to both clusters. This can be seen in Tables IX and XII of Appendix A. The fact that these were the last clusters formed in each set appears to be coincidental. Firm number 81 appears in the list of serial numbers of both clusters. Firms 91 and 94, the two chemical firms, also appear in both. Two other chemical firms, 98 and 131, appear one in each cluster. One additional similarity is present in that firm 171 in cluster C1.8, and firms 165 and 169 in cluster C2.8, are all producers of cement. Thus, at least one of the eight clusters formed in the first nine-year period has retained its identity well enough to be recognizable as one of the eight clusters formed in the second nine-year period.

Another phenomenon in the cluster membership is worthy of note, although its discovery was not an explicit purpose of the clustering.

Two of the clusters formed in the second period, C2.1 and C2.7, bear strong resemblances to industries. Of the 18 firms in C2.1, 10 are closely related to metals, especially steel. Firms 1 and 3 are classified by Compustat as miscellaneous metals; firm 161, as metal and glass containers; firms 177, 178, 179, 180, and 181 are steel firms; firms 192 and 193 are classified as primary smelting and refining. In C2.7, the first six serial numbers out of the 10 identify oil firms. In these two instances, it is plausible that there is some strong intra-industry similarity with respect to clustering variables used in this study.

Regression Results

For each firm in any group, whether an industry, a random group, or a cluster, the market index was calculated as the equity-weighted average of the performance measures taken over all of the other 358 firms. Separate indices were derived for each one of the two forms of annual performance measure, levels and changes.

The group index for each firm was calculated over the other firms in its group. Both simple and equity-weighted indices were calculated for each form of performance measure. By fitting equation (5) for each firm, the portions of these "raw" group indices correlated with the market index were removed.

Four forms of equation (6) and two forms of equation (7) were then fitted for each firm in each set of groups. In all regressions, nine annual observations were used. The 642 equations (107 firms, six equations each) were fitted for a total of seven sets of data: the groups Ri, Ii, and C1.i in each of the two nine-year periods, plus the C2.i in the second nine-year period. For each of the three sets of 107 regression equations fitted in the first nine-year period, Table XIV of Appendix A presents quartile R^2 values separately for the regressions fitted with only the market index, with market and simple group indices, and with market and weighted group indices; these three sets of R^2 values are presented for each of the two performance measures, levels and changes, producing a total of six sets of quartile R^2 values for each of the three data sets used in the first nine-year period.

Inspection of Table XIV reveals very little difference, if any, between the effects of cluster indices versus industry indices in increasing the observed R^2 value over that associated with the market index alone. Both, however, perform noticeably better than the index based on random groups. In the first nine-year period, then, the *ex post* test results do not support the research hypothesis of this study--that a grouping of firms based on several simple financial variables is more meaningful, in terms of revealing earnings commonalities, than traditional industry classification.

Table XV of Appendix A presents quartile values for each of the four data sets used in the second nine-year period. Again, as in Table XIV, each set of quartile values describes the distribution of the 107 R^2 values found for the 107 firms in each set.

Results shown in Table XV for the second nine-year period are not as straightforward as those in the first nine years. The first three data sets of Table XV, which represent the *ex ante* tests, reveal that clusters performed only about as well as the random groups, while industries performed noticeably better. But for the *ex post* second period clusters, C2.i, this is true only for the "levels" measure of performance. For the "changes" measure, the clusters C2.i performed as well as the industries, and consistently better than the random groups.

A conclusion consistent with this evidence is that the clustering procedure performs as well as industry classification in *ex post* tests, but not when the clusters formed in one period are tested in a later period, i.e., not in an *ex ante* test. This is likely due to a tendency of firms' financial variables, including the ones on which the clustering is based, to change substantially over a period of several years, such that the clusters formed in one period no longer possess intra-cluster similarity of financial characteristics in a later period. If the earnings commonalities observed in the first period are in fact related to the financial similarity, then in the second period the commonalities will disappear to the extent that the financial similarities disappear.

Incidentally, the second-period test of industries was not truly an *ex ante* test of industry earnings commonalities. The firms were chosen on the basis of having constant industry membership in the entire 18 year period, so the industry commonalities in earnings observed in the second nine-year period are due not to the firms' industry membership in the first nine years, but to their industry membership in the second nine years as well. This makes the second-period test of industry commonalities somewhat *ex post*, thus creating a bias in favor of stronger industry commonalities. A true *ex ante* test of industry commonalities would involve identifying industries in one period and testing for industry commonalities in earnings in a later period, when some firms will have become members of new industries.

No consistent differences have been found between the two index construction schemes, simple and weighted, nor between the two measures of earnings, levels and changes. Quartile R^2 values have been calculated to permit comparisons between this and earlier studies. On the whole, the evidence presented here on industry effects is in agreement with that offered by other authors, such as that of Brown and Ball (15), Brealey (14), and Magee (34), in terms of the relative magnitudes of market and industry effects. This result is not surprising, in view of the similarity of the samples used in these studies and in the present one.

An overall conclusion supported by the results presented in this section is that in terms of the relative usefulness of industries, clusters, and random groups in explaining earnings variations, industries and clusters perform about equally well, and both outperform random groups. The comparative disadvantage of the random groups, however, is rather small.

The results thus far do not support the research hypothesis of this study, which is that a grouping scheme based on similarity of financial characteristics should perform better than traditional industry groupings in explaining earnings commonalities across firms. An additional and equally significant finding is that the strength traditionally attributed to industry factors in earnings is not a great deal larger than that attributable to chance.

Residuals

The 42 sets of regression equations produced 42 sets of residuals, each comprising nine annual residuals on 107 firms. Within each set of residuals, a correlation coefficient was calculated for each possible pair of firms. The result is 42 correlation matrices measuring 107 by 107 in size. Each matrix is symmetric, of course, and has a value of one for every diagonal element. Similar matrices were also calculated for the original performance measures used as the dependent variables in the regressions, producing an additional 14 correlation matrices (seven data sets times two performance measures).

For each of the 56 matrices, the mean absolute value and a frequency distribution were calculated using above-diagonal elements. Mean absolute values are presented in Table XVI of Appendix A. As expected, the mean absolute values for the original performance measures are quite high, especially for the levels measure.

No consistent differences exist in Table XVI statistics between the two performance measures, levels and changes, except that the mean absolute value for changes is consistently closer to zero than that for levels, as expected. No differences can be found among the three grouping schemes, except that industry members exhibit lower correlations in the first nine years.

The same lack of differences was found in the frequency distributions for the 54 correlation matrices. Very high positive correlations were by far the most frequent values in the matrices calculated for the original levels measure, and to a lesser extent, for the original changes measure. As successive factors were abstracted from the variables, the residuals' correlation matrices' greatest frequencies moved consistently closer to zero. This was true for both levels and changes, for both index construction schemes, and for all three grouping methods.

For these reasons, and due to the great volume of data contained in the 56 frequency distributions, they are not presented here.

The final test of residuals is Bartlett's sphericity test of residuals' correlation matrices, the same matrices whose mean absolute values are given in Table XVI. For each matrix, a statistic was to be calculated according to the formula given in Chapter III based on the determinant of the matrix. However, all 56 matrices had determinants extremely close to zero, due probably to linear dependence among some of the 107 columns or 107 rows. As a result, Lawley's (32) approximation to the Barlett statistic was used. The resulting statistic for each matrix was to be compared to the chi-square value for 107(107-1)/2 degrees of freedom at the desired confidence level.

A formula for approximating chi-square values at extremely large numbers of degrees of freedom is given by Kane (29), based on degrees of freedom, v, and the standard normal deviate, Z_a, at the desired confidence level a. The formula is

$$\text{chi-square} = (1/2)(Z_a + (2v-1)^{1/2}).$$

Using 107(107-1)1/2 = 5671 degrees of freedom, the chi-square value at a .01 significance level is approximately 5921. At a .05 significance level, the value is approximately 5847. The hypothesis of independence among residuals must be rejected at the chosen significance level if the statistic calculated for a given correlation matrix is greater than the chi-square statistic. The statistics calculated by Lawley's approximation were above 20,000 for all 56 matrices, requiring rejection of the hypothesis of inter-firm independence of residuals in every case.

Results of the tests of residuals are therefore inconclusive. Further research might be directed toward these phenomena in a study using a larger number of observations on each firm.

Summary

Results of the preliminary tests were strongly in support of the hypothesized relationships of Chapter II. Signs and magnitudes of the Spearman coefficients between the performance measures and the five financial variables agreed with predictions. In the second test, several significant interaction terms were found among the five financial variables in influencing the levels of the nine-year summary measures of performance.

Factor analysis of the five financial variables produced a set of rotated factors in each nine-year period. The clusters formed on the basis of similarity with respect to rotated factor scores were interpretable in terms of typical scores on each rotated factor.

Results of the regression analysis were not in support of the research hypothesis of this study. Clusters performed only about as well as industries in accounting for earnings commonalities, and both generally outperformed the random group, although not by a wide margin.

Tests of residuals proved inconclusive. No consistent differences could be observed in correlation coefficients calculated among the residuals of the various sets of firms. Further, the hypothesis of zero inter-firm correlation of residuals was rejected for every matrix of correlation coefficients tested.

CHAPTER V

CONCLUSION

This final chapter consists of two sections. The first enumerates several limitations of the research methodology which affect the generality of reported results. The second describes the implications of the study's findings for empirical research in accounting.

Limitations

Restriction of the study to Compustat firms must introduce sampling biases which render the results not generalizable to all firms in general. Firms listed on the Compustat File are among the oldest and largest in the economy, and it is plausible that the observed relationships may not be common to smaller and younger firms. The necessity for reported data spanning a large number of years, however, explicitly requires that the sample be restricted to older firms. The unfortunate result is that inferences can be made only to an unknown population of which Compustat firms constitute a truly representative sample. Therefore, the next-best alternative is chosen: since Compustat firms, being among the oldest and largest in the economy, are of considerable interest as a population, then they, rather than firms in general, are viewed as the population of interest for this study. Generalizations of results to other populations will not be attempted.

A second limitation of this study concerns its role as a challenge to the use of traditional industry classification. The method of grouping firms that was proposed in this study is not intended as a new definition of industries. This grouping method is proposed as an alternative to traditional industries only in searching for inter-firm earnings commonalities, and not in the various other uses commonly made of industry analysis.

A potentially serious statistical flaw is the use of only nine annual observations in fitting each regression equation and in computing correlation coefficients for the resulting residuals. The likelihood of large sampling error is high with such small numbers of observations. However, the differences consistently observed between the performance of the random groups versus the industries and clusters is some evidence that meaningful effects are present--of the two grouping methods for which there were high expectations of performance, industries and clusters, one or both outperformed the random groups in all regression models tested. It is possible, though, that the inconclusive results of the tests of residual correlation matrices are due largely to the small number of observations available.

A final and very general limitation deserves mention. In any study involving a search for variables which are purported to be determinants of some factor of interest, it is difficult to demonstrate conclusively that no relevant variables have been overlooked. To demonstrate that an irrelevant variable has been included would be a much easier task, since the number of variables chosen places a comfortable limit on the search. But to prove that no other relevant variables exist would require examination of innumerable possibilities. In the event that a relevant variable is ommitted from the list of five financial variables chosen in this study, an explainable portion of variability in firms' earnings has gone unexplained. Therefore, the analytical power of a grouping scheme based on financial characteristics of firms may in fact be greater than that indicated by the results of this study.

Implications

In this study, an alternative scheme for grouping firms was proposed. The alternative scheme is based on firms' similarity with respect to five simple, familiar financial variables: leverage, liquidity, payout, growth, and size. Groups of firms formed on the basis of these five variables performed about as well as did Compustat three-digit industries (which correspond closely to SIC three-digit classifications) in accounting for earnings commonalities, and both schemes outperformed a grouping of firms based on chance.

It was the research hypothesis of this study that such an alternative grouping scheme would outperform traditional industries by better representing the effects of firms' participation in common markets. Since the shortcomings of traditional industries in this regard are well documented, the alternative scheme was expected to exhibit superior performance.

Results indicate that although the "participation in common markets" argument may be invalid with regard to traditional industries, industry groupings are nevertheless useful in capturing non-market-wide performance similarities. A conclusion consistent with this finding is that although industries may be poor surrogates for groups of firms participating in common markets, industry groupings are still meaningful in representing earnings commonalities for some other reason.

SELECTED BIBLIOGRAPHY

(1) Aber, John W. "Industry Effects and Multivariate Stock Price Behavior." Journal of Financial and Quantitative Analysis, 11 (November, 1976), pp. 617-624.

(2) Arditti, Fred D. "Risk and the Required Return on Equity." The Journal of Finance, 22 (March, 1967), pp. 19-36.

(3) Baker, Samuel H. "Risk, Leverage and Profitability: An Industry Analysis." The Review of Economics and Statistics, 55 (November, 1973), pp. 503-507.

(4) Ball, Ray and Philip Brown. "An Empirical Evaluation of Accounting Income Numbers." The Journal of Accounting Research, 6 (Autumn, 1968), pp. 159-178.

(5) Ball, Ray and Philip Brown. "Portfolio Theory and Accounting." The Journal of Accounting Research, 7 (Autumn, 1969), pp. 300-323.

(6) Ball, Ray. "Changes in Accounting Techniques and Stock Prices." Empirical Research in Accounting: Selected Studies, 1972. Supplement, The Journal of Accounting Research, 10 (1972), pp. 1-38.

(7) Baskin, Elba F. "The Communicative Effectiveness of Consistency Exceptions." The Accounting Review, 47 (January, 1972), pp. 38-51.

(8) Baumol, W. J. Business Behavior, Value and Growth. New York: MacMillan, 1959.

(9) Beaver, William H. "The Information Content of Annual Earnings Announcements." Empirical Research in Accounting: Selected Studies, 1968. Supplement, The Journal of Accounting Research, 5 (1967), pp. 67-92.

(10) Beaver, William H. and Roland E. Dukes. "Interperiod Tax Allocation, Earnings Expectations, and the Behavior of Security Prices." The Accounting Review, 47 (April, 1972), pp. 320-332.

(11) Beaver, William H. and Roland E. Dukes. "Interperiod Tax Allocation and δ-Depreciation Methods: Some Empirical Results." The Accounting Review, 48 (July, 1973), pp. 549-559.

(12) Benston, George J. "The Federal Trade Commission's Line of Business Program: A Benefit-Cost Analysis." Unpublished paper, 1976.

(13) Bock, Betty. "Affidavit of Betty Bock with Respect to Proposed 1974 FTC Form LB." _United States of America Before Federal Trade Commission in the Matter of Proposed 1974 FTC Form LB._ In Benston, George J., "The Federal Trade Commission's Line of Business Program: A Benefit-Cost Analysis." Unpublished paper, October, 1976.

(14) Brealey, Richard A. _An Introduction to Risk and Return from Common Stocks._ Cambridge: MIT Press, 1969.

(15) Brown, Philip and Ray Ball. "Some Preliminary Findings on the Association Between the Earnings of a Firm, Its Industry, and the Economy." _Empirical Research in Accounting_: _Selected Studies, 1967._ Supplement, _The Journal of Accounting Research_, 5 (1967), pp. 55-77.

(16) Brown, Philip and John W. Kennelly. "The Information Content of Quarterly Earnings: An Extension and Some Further Evidence." _The Journal of Business_, 45 (July, 1972), pp. 403-415.

(17) Canada, John R. _Intermediate Economic Analysis for Management and Engineering._ Englewood Cliffs, New Jersey: Prentice-Hall, 1971.

(18) Cooley, William W. and Paul R. Lohnes. _Multivariate Data Analysis._ New York: John Wiley and Sons, 1971.

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

(20) Everitt, Brian. _Cluster Analysis._ New York: John Wiley and Sons, 1975.

(21) Foster, George. "Accounting Earnings and Stock Prices of Insurance Companies." _The Accounting Review_, 50 (October, 1975), pp. 686-698.

(22) Fertuck, Leonard. "A Test of Industry Indices Based on SIC Codes." _Journal of Financial and Quantitative Analysis_, 10 (December, 1975), pp. 837-848.

(23) Gale, Bradley T. "Market Share and Rate of Return." _The Review of Economics and Statistics_, 54 (November, 1972), pp. 412-423.

(24) Gaumnitz, Jack E. "The Influence of Industry Factors in Stock Price Movements." Paper presented at the Southern Finance Association Meeting, October, 1970. In Reilly, Frank K. and Eugene F. Drzycimski, "Alternate Industry Performance and Risk." Journal of Financial and Quantitative Analysis, 9 (June, 1972), pp. 423-446.

(25) Gonedes, Nicholas J. "Properties of Accounting Numbers: Models and Tests." The Journal of Accounting Research, 11 (Autumn, 1973), pp. 212-237.

(26) Hall, Marshall and Leonard Weiss. "Firm Size and Profitability." The Review of Economics and Statistics, 49 (August, 1967), pp. 319-331.

(27) Harmon, Harry H. Modern Factor Analysis. Chicago: University of Chicago Press, 1960.

(28) Hartigan, John A. Clustering Algorithms. New York: John Wiley and Sons, 1975.

(29) Kane, Edward J. Economic Statistics and Econometrics. New York: Harper and Row, 1968. In Aber, John W., "Industry Effects and Multivariate Stock Price Behavior." Journal of Financial and Quantitative Analysis, 11 (November, 1976), pp. 617-624.

(30) Kaplan, Robert S. and Richard Roll. "Investor Evaluation of Accounting Information: Some Empirical Evidence." The Journal of Business, 45 (April, 1972), pp. 225-257.

(31) King, Benjamin F. "Market and Industry Factors in Stock Price Behavior." The Journal of Business, 39 (January, 1966), pp. 139-190.

(32) Lawley, D. N. "The Estimation of Factor Loadings by the Method of Maximum Likelihood." Proceedings of the Royal Society of Edinburgh, 40 (1940), pp. 64-82. In Morrison, Donald F., Multivariate Statistical Methods. New York: McGraw-Hill, 1967.

(33) Livingston, Miles. "Industry Movements of Common Stocks." The Journal of Finance, 32 (June, 1977), pp. 861-874.

(34) Magee, Robert P. "Industry-Wide Commonalities in Earnings." The Journal of Accounting Research, 12 (Autumn, 1974), pp. 270-287.

(35) Markowitz, Harry M. Portfolio Selection: Efficient Diversification of Investment. New York: John Wiley and Sons, 1959.

(36) May, Robert C. "The Influence of Quarterly Earnings Announcements on Investor Decisions as Reflected in Common Stock Price Changes." Empirical Research in Accounting: Selected Studies, 1971. Supplement, The Journal of Accounting Research, 9 (1971), pp. 119-163.

(37) McKibben, Walt. "Econometric Forecasting of Common Stock Investment Returns: A New Methodology Using Fundamental Operating Data." The Journal of Finance, 27 (May, 1972), pp. 371-380.

(38) Meyers, Stephen L. "A Re-examination of Market and Industry Factors in Stock Price Behavior." The Journal of Finance, 28 (June, 1973), pp. 695-705.

(39) Miller, M. H. and F. Modigliani. "Dividend Policy, Growth, and the Valuation of Shares." The Journal of Business, 34 (October, 1961), pp. 411-433.

(40) Morrison, Donald F. Multivariate Statistical Methods. New York: McGraw-Hill, 1967.

(41) Reilly, Frank K. and Eugene F. Drzycimski. "Alternate Industry Performance and Risk." Journal of Financial and Quantitative Analysis, 9 (June, 1974), pp. 423-446.

(42) Sharpe, William F. "A Simplified Model for Portfolio Analysis." Management Science, 9 (January, 1963), pp. 277-293.

(43) Sullivan, Timothy G. "Market Power, Profitability and Financial Leverage." The Journal of Finance, 29 (December, 1974), pp. 1407-1414.

(44) Veldman, Donald J. Fortran Programming for the Behavioral Sciences. New York: Holt, Rinehart, Winston, 1967.

(45) Williams, W. H. and M. L. Goodman. "A Statistical Grouping of Corporations by Their Financial Characteristics." Journal of Financial and Quantitative Analysis, 6 (September, 1971), pp. 1095-1104.

(46) Wippern, Ronald F. "Financial Structure and the Value of the Firm." The Journal of Finance, 21 (December, 1966), pp. 615-633.

(47) Young, Tzay Y. and Thomas W. Calvert. Classification, Estimation, and Pattern Recognition. New York: American Elsevier Publishing Company, Inc., 1974.

APPENDIXES

APPENDIX A

TABLES

TABLE I

LIST OF COMPUSTAT DATA ITEMS USED

Data Item Number	Name
4	Current Assets
5	Current Liabilities
6	Total Assets
9	Long Term Debt
11	Common Equity
18	Net Income
19	Preferred Dividends
21	Common Dividends

TABLE II

COMPUTATION OF VARIABLES $FV1_i$

Symbol	Variable Name	Formula
$FV1_1$	Leverage	$\sum_{t=2}^{10} (\text{Data } 5_t + \text{Data } 9_t) \;/\; \sum_{t=2}^{10} \text{Data } 6_t$
$FV1_2$	Liquidity	$\sum_{t=2}^{10} \text{Data } 4_t \;/\; \sum_{t=2}^{10} \text{Data } 6_t$
$FV1_3$	Payout	$\sum_{t=3}^{11} \text{Data } 21_t \;/\; \sum_{t=3}^{11} (\text{Data } 18_t - \text{Data } 19_t)$
$FV1_4$	Growth	$\log(\text{Data } 6_{10}) - \log(\text{Data } 6_2)$
$FV1_5$	Size	$\sum_{t=2}^{10} \log(\text{Data } 6_t) \;/\; 9$

TABLE III

COMPUTATION OF PERFORMANCE MEASURES

Performance Measures	Formula
Annual Measures $E_{k.i,t}$:	
k=1, Levels	(Data 18_t - Data 19_t)/Data 11_{t-1}
k=2, Changes	(Data 18_t - Data 19_t)/Data 11_{t-1}
	- (Data 18_{t-1} - Data 19_{t-1})/Data 11_{t-2}
Nine-Year Measures:	
Level	$\left(\sum_{t=3}^{11} \text{Data } 18_t - \sum_{t=3}^{11} \text{Data } 19\right) / \sum_{t=3}^{11} \text{Data } 11_{t-1}$
Change	(Data 18_{11} - Data 19_{11})/Data 11_{10}
	- (Data 18_3 - Data 19_3)/Data 11_2

TABLE IV

SPEARMAN COEFFICIENTS FOR $FV1_i$ AND NINE-YEAR PERFORMANCE MEASURES

Financial Variables $FV1_i$		Nine-Year Performance Measure		Predicted Sign
Symbol	Name	Level	Change	
$FV1_1$	Leverage	.030	.146	None
$FV1_2$	Liquidity	.222	.025	None
$FV1_3$	Payout	-.158	-.295	-
$FV1_4$	Growth	.504	.164	+
$FV1_5$	Size	-.076	-.085	-

TABLE V

SIGNIFICANT INTERACTION TERMS

Dependent Variable	Interaction	F	Level of Significance ≥
Nine-year Level	size-growth-payout	5.326	.02
"	leverage-growth-liquidity	4.136	.04
"	payout-liquidity	5.227	.02
"	size-payout-liquidity	3.712	.05
"	size-leverage-payout-liquidity	6.486	.01
Nine-year Change	payout-liquidity	2.965	.08
"	size-liquidity	3.148	.07
"	leverage-liquidity	3.503	.06
"	size-leverage-liquidity	5.908	.01
"	growth-liquidity	3.002	.08
"	leverage-growth-liquidity	2.922	.08

TABLE VI

MEMBERSHIP OF THE INDUSTRIES

Industry Symbol	Compustat Industry Name	Compustat Industry Number	Number of Members	Serial Numbers of Members
I1	Chemicals	280	18	87 through 104
I2	Autos and Parts	371	17	270 through 286
I3	Drugs	283	15	105 through 119
I4	Oil	291	14	134 through 147
I5	Steel	331	12	177 through 188
I6	Beverages	208	11	35 through 45
I7	Electronics	367	10	260 through 269
I8	Air Transport	451	10	318 through 327

TABLE VII

MEMBERSHIP OF THE RANDOM GROUPS

Group Symbol	Number of Members	Serial Numbers of Members
R1	18	97,189,95,172,158,99,357,266,212,153,195, 67,352,131,140,31,227,209
R2	17	165,299,145,278,108,18,135,10,5,93,35,116, 349,288,90,43,29
R3	15	147,48,160,123,284,63,192,107,305,91,120, 261,3,28,206
R4	14	333,128,322,273,179,327,77,13,188,328,294, 339,52,308
R5	12	218,127,16,303,334,183,171,1,42,296,338,45
R6	11	149,223,257,32,194,62,49,260,40,276,214
R7	10	44,196,350,258,243,60,161,26,346,14
R8	10	282,47,89,7,80,263,69,12,118,248

TABLE VIII

FACTOR MATRIX FOR $RF1_i$

Variable Name	Variable Symbol	Loading on $RF1_1$	Loading on $RF1_2$
Leverage	$FV1_1$	0.671	0.156
Liquidity	$FV1_2$	0.070	-0.685
Payout	$FV1_3$	-0.790	0.096
Growth	$FV1_4$	0.755	-0.298
Size	$FV1_5$	-0.020	0.864

TABLE IX

MEMBERSHIP OF CLUSTERS C1.i

Cluster Symbol	Number of Members	Serial Numbers of Members
C1.1	18	21,27,48,49,54,64,105,106,116,174,176, 184,199,201,221,255,256,286
C1.2	17	25,40,73,74,75,88,98,136,141,158,164,165, 167,169,170,177,247
C1.3	15	38,56,70,86,93,97,109,211,214,219,242, 285,306,334,335
C1.4	14	5,65,108,111,125,133,175,203,217,248,267, 295,302,355
C1.5	12	37,85,99,132,150,207,222,224,232,258,281, 346
C1.6	11	41,55,70,112,130,154,228,231,251,303,310
C1.7	10	69,148,153,190,218,236,239,289,308,333
C1.8	10	35,72,81,91,94,120,131,171,274,293

TABLE X

INTERPRETATIONS OF Cl.i CLUSTERS IN TERMS OF RFl_i

Cluster Symbol	Level of RFl_1	Level of RFl_2
Cl.1	low	low
Cl.2	very low	very high
Cl.3	high	moderate
Cl.4	very low	very low
Cl.5	low	very low
Cl.6	moderate	low
Cl.7	high	high
Cl.8	moderate	high

TABLE XI

FACTOR MATRIX FOR $RF2_i$

Variable Name	Variable Symbol	Loading on $RF2_1$	Loading on $RF2_2$	Loading on $RF2_3$
Leverage	$FV2_1$	-0.003	-0.047	0.899
Liquidity	$FV2_2$	0.838	0.111	0.277
Payout	$FV2_3$	0.030	0.839	0.198
Growth	$FV2_4$	0.066	-0.757	0.348
Size	$FV2_5$	-0.782	0.162	0.359

TABLE XII

MEMBERSHIP OF CLUSTERS C2.i

Cluster Symbol	Number of Members	Serial Numbers of Members
C2.1	18	1,3,68,71,73,74,87,92,134,158,161,177, 178,179,180,181,192,193
C2.2	17	38,33,34,50,55,63,79,99,117,217,229,234, 244,253,267,299,302
C2.3	15	24,112,113,115,125,204,219,259,269,273, 285,287,289,333,359
C2.4	14	20,52,61,150,154,202,223,237,241,251,277, 297,298,339
C2.5	12	18,22,85,109,218,220,255,279,280,313,344, 358
C2.6	11	25,30,40,77,121,152,199,200,216,334,343
C2.7	10	138,142,144,145,146,147,182,195,294,332
C2.8	10	23,49,81,91,94,98,165,169,278,305

TABLE XIII

INTERPRETATIONS OF C2.i CLUSTERS IN TERMS OF $RF2_i$

Cluster Symbol	Level of $RF2_1$	Level of $RF2_2$	Level of $RF2_3$
C2.1	low	high	low
C2.2	very high	moderate	very low
C2.3	high	very low	very high
C2.4	very high	moderate	high
C2.5	moderate	low	high
C2.6	moderate	very high	high
C2.7	very low	moderate	moderate
C2.8	moderate	moderate	moderate

TABLE XIV

QUARTILE R^2 VALUES IN THE FIRST NINE-YEAR PERIOD

Data Set	Equation Form: / Quartile	Levels: Market Index Only	Levels: Market & Simple Group Indices	Levels: Market & Weighted Group Indices	Changes: Market Index Only	Changes: Market & Simple Group Indices	Changes: Market & Weighted Group Indices
Cl.i	1st	.144	.268	.262	.084	.177	.178
	2nd	.311	.584	.561	.253	.444	.422
	3rd	.607	.798	.780	.504	.619	.624
	4th	.943	.966	.965	.893	.905	.905
Ii	1st	.150	.367	.329	.055	.179	.206
	2nd	.379	.568	.577	.228	.442	.446
	3rd	.618	.727	.766	.447	.617	.629
	4th	.943	.971	.942	.862	.954	.939
Ri	1st	.084	.253	.232	.059	.136	.149
	2nd	.322	.490	.515	.206	.372	.351
	3rd	.604	.713	.717	.428	.555	.572
	4th	.834	.898	.969	.847	.899	.888

TABLE XV

QUARTILE R^2 VALUES IN SECOND NINE-YEAR PERIOD

Data Set	Equation Form: / Quartile	Levels: Market Index Only	Levels: Market & Simple Group Indices	Levels: Market & Weighted Group Indices	Changes: Market Index Only	Changes: Market & Simple Group Indices	Changes: Market & Weighted Group Indices
C1.i	1st	.045	.188	.173	.067	.220	.199
	2nd	.155	.398	.392	.206	.402	.366
	3rd	.367	.649	.635	.451	.592	.578
	4th	.776	.915	.887	.834	.884	.883
Ii	1st	.049	.361	.361	.075	.270	.242
	2nd	.229	.551	.588	.221	.490	.474
	3rd	.477	.744	.795	.529	.727	.774
	4th	.869	.980	.957	.834	.966	.987
Ri	1st	.062	.255	.244	.047	.195	.233
	2nd	.182	.406	.387	.212	.398	.392
	3rd	.361	.572	.572	.442	.542	.605
	4th	.899	.939	.920	.950	.867	.958
C2.i	1st	.041	.142	.129	.076	.233	.230
	2nd	.169	.366	.317	.253	.463	.444
	3rd	.450	.607	.578	.463	.683	.678
	4th	.753	.962	.958	.857	.885	.977

TABLE XVI

MEAN ABSOLUTE VALUES OF INTER-FIRM CORRELATION COEFFICIENTS

Coefficients Calculated from: \ Data Set:	First Nine Years			Second Nine Years			
	C1i	Ii	Ri	C1i	Ii	Ri	C2i
Original Variables:							
Levels:	.89	.86	.88	.83	.76	.83	.88
Changes:	.35	.32	.32	.34	.34	.34	.35
Residuals of:							
Levels:							
Using Market Index Only	.40	.39	.40	.44	.41	.42	.42
Using Market and Simple Group Indices	.36	.36	.36	.36	.36	.36	.39
Using Market and Weighted Group Indices	.36	.36	.35	.36	.35	.36	.40
Changes:							
Using Market Index Only	.33	.33	.34	.34	.34	.35	.36
Using Market and Simple Group Indices	.33	.33	.33	.33	.32	.34	.33
Using Market and Weighted Group Indices	.34	.33	.32	.33	.33	.32	.33

APPENDIX B

LIST OF FIRMS

SERIAL NUMBER	COMPUSTAT INDUSTRY NUMBER	INDUSTRY NAME	COMPUSTAT COMPANY NUMBER	COMPANY NAME
1	1000	METALS-MISC	43413	ASARCO INC
2	1000	METALS-MISC	117421	BRUSH WELLMAN INC
3	1000	METALS-MISC	453258	INCO LTD
4	1000	METALS-MISC	608744	MOLYCORP INC
5	1000	METALS-MISC	882887	TEXASGULF INC
6	1021	COPPER ORES	131069	CALLAHAN MINING CORP
7	1042	GOLD MINING	134411	CAMPBELL RED LAKE MINES
8	1042	GOLD MINING	257075	DOME MINES LTD
9	1042	GOLD MINING	374586	GIANT YELLOWKNIFE MINES
10	1042	GOLD MINING	437614	HOMESTAKE MINING
11	1042	GOLD MINING	581238	MC INTYRE MINES LTD
12	1211	COAL-BITUMINOUS	276461	EASTERN GAS & FUEL ASSOC
13	1211	COAL-BITUMINOUS	656780	NORTH AMERICAN COAL
14	1211	COAL-BITUMINOUS	725701	PITTSTON CO
15	1211	COAL-BITUMINOUS	790155	ST. JOE MINERALS CORP
16	1311	OIL-CRUDE PRODUCERS	257093	DOME PETROLEUM LTD
17	1499	MISC NONMETAL MINING	356715	FREEPORT MINERALS CO
18	1511	GENERAL BLDG CONTRACTORS	618448	MORRISON-KNUDSEN
19	1621	HEAVY CONSTRUCTION-EX HWY&ST	20771	ALPHA PORTLAND INDS
20	1621	HEAVY CONSTRUCTION-EX HWY&ST	754722	RAYMOND INTL INC
21	2000	FOOD-PACKAGED FOODS	487836	KELLOGG CO
22	2000	FOOD-PACKAGED FOODS	853139	STANDARD BRANDS INC
23	2020	FOOD-DAIRY PRODUCTS	99599	BORDEN INC
24	2020	FOOD-DAIRY PRODUCTS	143483	CARNATION CO
25	2020	FOOD-DAIRY PRODUCTS	500755	KRAFTCO CORP
26	2051	FOOD-BREAD & CAKE BAKERS	24069	AMERICAN BAKERIES CO
27	2051	FOOD-BREAD & CAKE BAKERS	460754	INTERSTATE BRANDS
28	2051	FOOD-BREAD & CAKE BAKERS	876553	TASTY BAKING CO
29	2051	FOOD-BREAD & CAKE BAKERS	934051	WARD FOODS INC
30	2052	FOOD-BISCUIT BAKER	629527	NABISCO INC
31	2065	CONFECTIONERY	427866	HERSHEY FOODS CORP
32	2065	CONFECTIONERY	890516	TOOTSIE ROLL INDS INC
33	2065	CONFECTIONERY	982526	WRIGLEY (WM.) JR. CO
34	2070	FATS & OILS	42083	ARMADA CORP
35	2082	BEVERAGES-BREWERS	35231	ANHEUSER-BUSCH INC
36	2082	BEVERAGES-BREWERS	422884	HEILEMAN (G.) BREWING INC
37	2082	BEVERAGES-BREWERS	681453	OLYMPIA BREWING
38	2082	BEVERAGES-BREWERS	693715	PABST BREWING CO
39	2082	BEVERAGES-BREWERS	806823	SCHLITZ (JOS.) BREWING
40	2085	BEVERAGES-DISTILLERS	635655	NATIONAL DISTILLERS &CHEMICL
41	2085	BEVERAGES-DISTILLERS	744635	PUBLICKER INDS INC
42	2086	BEVERAGES-SOFT DRINKS	191162	COCA-COLA BOTTLING CO OF NY
43	2086	BEVERAGES-SOFT DRINKS	256129	DR PEPPER CO
44	2086	BEVERAGES-SOFT DRINKS	713448	PEPSICO INC
45	2086	BEVERAGES-SOFT DRINKS	780240	ROYAL CROWN COLA CO
46	2111	TOBACCO-CIGARETTE MFG	24703	AMERICAN BRANDS INC
47	2111	TOBACCO-CIGARETTE MFG	532202	LIGGETT GROUP
48	2111	TOBACCO-CIGARETTE MFG	718167	PHILIP MORRIS INC
49	2111	TOBACCO-CIGARETTE MFG	761753	REYNOLDS (R.J.) INDS
50	2111	TOBACCO-CIGARETTE MFG	912775	U.S. TOBACCO CO
51	2121	TOBACCO-CIGAR MFG	73239	BAYUK CIGARS INC
52	2200	TEXTILE PRODUCTS	77491	BELDING HEMINWAY
53	2200	TEXTILE PRODUCTS	206813	CONE MILLS CORP
54	2200	TEXTILE PRODUCTS	235773	DAN RIVER INC
55	2200	TEXTILE PRODUCTS	387478	GRANITEVILLE CO
56	2200	TEXTILE PRODUCTS	547779	LOWENSTEIN (M.) & SONS INC
57	2270	CARPET & FLOOR COVERINGS	42321	ARMSTRONG CORK CO
58	2270	CARPET & FLOOR COVERINGS	207192	CONGOLEUM CORP
59	2300	TEXTILE APPAREL MFG	6284	ADAMS-MILLIS CORP
60	2300	TEXTILE APPAREL MFG	189486	CLUETT, PEABODY & CO
61	2300	TEXTILE APPAREL MFG	626320	MUNSINGWEAR INC
62	2400	FOREST PRODUCTS	299209	EVANS PRODUCTS CO
63	2510	HOME FURNISHING	501026	KROEHLER MFG CO
64	2510	HOME FURNISHING	608030	MOHASCO CORP
65	2510	HOME FURNISHING	828709	SIMMONS CO
66	2520	OFC & PUB BLDG-FURN & FIX	29465	AMERICAN SEATING CO
67	2600	PAPER	102187	BOWATER CORP LTD-ADR
68	2600	PAPER	228669	CROWN ZELLERBACH
69	2600	PAPER	257561	DOMTAR LTD
70	2600	PAPER	408306	HAMMERMILL PAPER CO
71	2600	PAPER	460146	INTL PAPER CO
72	2600	PAPER	737628	POTLATCH CORP

SERIAL NUMBER	COMPUSTAT INDUSTRY NUMBER	INDUSTRY NAME	COMPUSTAT COMPANY NUMBER	COMPANY NAME
73	2600	PAPER	793453	ST. REGIS PAPER CO
74	2600	PAPER	809877	SCOTT PAPER CO
75	2600	PAPER	905530	UNION CAMP CORP
76	2650	PAPER-CONTAINERS	252669	DIAMOND INTL CORP
77	2650	PAPER-CONTAINERS	315711	FIBREBOARD CORP
78	2650	PAPER-CONTAINERS	861589	STONE CONTAINER CORP
79	2700	PUBLISHING	959265	WESTERN PUBLISHING
80	2711	PUBLISHING-NEWSPAPERS	887360	TIMES MIRROR CO
81	2721	PUBLISHING-PERIODICALS	887224	TIME INC
82	2731	PUBLISHING-BOOKS	398784	GROLIER INC
83	2731	PUBLISHING-BOOKS	411631	HARCOURT BRACE JOVANOVICH
84	2731	PUBLISHING-BOOKS	554790	MACMILLAN INC
85	2731	PUBLISHING-BOOKS	580645	MCGRAW-HILL INC
86	2752	COMMERCIAL PRINTING & LITH	257867	DONNELLEY (R. R.) & SONS CO
87	2801	CHEMICALS-MAJOR	19087	ALLIED CHEMICAL CORP
88	2801	CHEMICALS-MAJOR	25321	AMERICAN CYANAMID CO
89	2801	CHEMICALS-MAJOR	150843	CELANESE CORP
90	2801	CHEMICALS-MAJOR	383883	GRACE (W.R.) & CO
91	2801	CHEMICALS-MAJOR	427056	HERCULES INC
92	2801	CHEMICALS-MAJOR	611662	MONSANTO CO
93	2801	CHEMICALS-MAJOR	775371	ROHM & HAAS CO
94	2801	CHEMICALS-MAJOR	857721	STAUFFER CHEMICAL CO
95	2801	CHEMICALS-MAJOR	905581	UNION CARBIDE CORP
96	2802	CHEMICALS-INTERMEDIATE	680665	OLIN CORP
97	2802	CHEMICALS-INTERMEDIATE	759200	REICHHOLD CHEMICALS INC
98	2803	CHEMICALS-SPECIALTY	9292	AIRCO INC
99	2803	CHEMICALS-SPECIALTY	227111	CROMPTON & KNOWLES CORP
100	2803	CHEMICALS-SPECIALTY	237424	DART INDS
101	2803	CHEMICALS-SPECIALTY	317315	FILTROL CORP
102	2803	CHEMICALS-SPECIALTY	457641	INMONT CORP
103	2803	CHEMICALS-SPECIALTY	500602	KOPPERS CO
104	2803	CHEMICALS-SPECIALTY	554205	MACANDREWS & FORBES
105	2835	DRUGS-ETHICAL	2824	ABBOTT LABORATORIES
106	2835	DRUGS-ETHICAL	26609	AMERICAN HOME PRODUCTS CORP
107	2835	DRUGS-ETHICAL	532457	LILLY (ELI)& CO
108	2835	DRUGS-ETHICAL	589331	MERCK & CO
109	2835	DRUGS-ETHICAL	717081	PFIZER INC
110	2835	DRUGS-ETHICAL	812302	SEARLE (G.D.) & CO
111	2835	DRUGS-ETHICAL	832377	SMITHKLINE CORP
112	2835	DRUGS-ETHICAL	934488	WARNER-LAMBERT CO
113	2836	DRUGS-PROPRIETARY	110097	BRISTOL-MYERS CO
114	2836	DRUGS-PROPRIETARY	463349	IROQUOIS BRANDS LTD
115	2836	DRUGS-PROPRIETARY	599292	MILES LABORATORIES INC
116	2836	DRUGS-PROPRIETARY	859264	STERLING DRUG INC
117	2836	DRUGS-PROPRIETARY	875314	TAMPAX INC
118	2837	DRUGS-MEDICAL & HOSP SUPPLY	26681	AMERICAN HOSPITAL SUPPLY
119	2837	DRUGS-MEDICAL & HOSP SUPPLY	71892	BAXTER TRAVENOL LABORATORIES
120	2841	SOAP	194162	COLGATE-PALMOLIVE CO
121	2841	SOAP	904784	UNILEVER N V
122	2844	COSMETICS	54303	AVON PRODUCTS
123	2844	COSMETICS	165339	CHESEBROUGH-POND'S INC
124	2844	COSMETICS	302808	FABERGE INC
125	2844	COSMETICS	375766	GILLETTE CO
126	2844	COSMETICS	761525	REVLON INC
127	2850	PAINT	739732	PRATT & LAMBERT INC
128	2899	CHEM & CHEM PREPARATIONS	212867	CONWOOD CORP
129	2899	CHEM & CHEM PREPARATIONS	255093	DIVERSEY CORP
130	2899	CHEM & CHEM PREPARATIONS	315405	FERRO CORP
131	2899	CHEM & CHEM PREPARATIONS	492746	KEWANEE INDS
132	2899	CHEM & CHEM PREPARATIONS	629853	NALCO CHEMICAL CO
133	2899	CHEM & CHEM PREPARATIONS	866645	SUN CHEMICAL CORP
134	2911	PETROLEUM REFINING	453038	IMPERIAL OIL LTD-CL A
135	2912	OIL-INTEGRATED DOMESTIC	211813	CONTINENTAL OIL CO
136	2912	OIL-INTEGRATED DOMESTIC	565845	MARATHON OIL CO
137	2912	OIL-INTEGRATED DOMESTIC	626717	MURPHY OIL CORP
138	2912	OIL-INTEGRATED DOMESTIC	718507	PHILLIPS PETROLEUM CO
139	2912	OIL-INTEGRATED DOMESTIC	747419	QUAKER STATE OIL REFINING
140	2912	OIL-INTEGRATED DOMESTIC	830575	SKELLY OIL CO
141	2912	OIL-INTEGRATED DOMESTIC	853734	STANDARD OIL CO (OHIO)
142	2912	OIL-INTEGRATED DOMESTIC	907770	UNION OIL CO OF CALIFORNIA
143	2913	OIL-INTEGRATED INTL	110889	BRITISH PETROLEUM CO LTD
144	2913	OIL-INTEGRATED INTL	302290	EXXON CORP

SERIAL NUMBER	COMPUSTAT INDUSTRY NUMBER	INDUSTRY NAME	COMPUSTAT COMPANY NUMBER	COMPANY NAME
145	2913	OIL-INTEGRATED INTL	402460	GULF OIL CORP
146	2913	OIL-INTEGRATED INTL	607080	MOBIL OIL CORP
147	2913	OIL-INTEGRATED INTL	881694	TEXACO INC
148	2950	BLDG MAT-ROOF&WALLBOARD	339711	FLINTKOTE CO
149	2950	BLDG MAT-ROOF&WALLBOARD	636316	NATIONAL GYPSUM CO
150	2950	BLDG MAT-ROOF&WALLBOARD	770553	ROBERTSON (H H) CO
151	3000	TIRE & RUBBER GOODS	216831	COOPER TIRE & RUBBER
152	3000	TIRE & RUBBER GOODS	382388	GOODRICH (B.F.) CO
153	3000	TIRE & RUBBER GOODS	382550	GOODYEAR TIRE & RUBBER CO
154	3000	TIRE & RUBBER GOODS	564402	MANSFIELD TIRE & RUBBER CO
155	3000	TIRE & RUBBER GOODS	608302	MOHAWK RUBBER CO
156	3000	TIRE & RUBBER GOODS	909160	UNIROYAL INC.
157	3210	GLASS PRODUCTS	219327	CORNING GLASS WORKS
158	3210	GLASS PRODUCTS	693506	PPG INDS
159	3221	CONTAINERS-METAL&GLASS	24843	AMERICAN CAN CO
160	3221	CONTAINERS-METAL&GLASS	33047	ANCHOR HOCKING CORP
161	3221	CONTAINERS-METAL&GLASS	211452	CONTINENTAL GROUP
162	3221	CONTAINERS-METAL&GLASS	228255	CROWN CORK & SEAL CO INC
163	3221	CONTAINERS-METAL&GLASS	635128	NATIONAL CAN CORP
164	3221	CONTAINERS-METAL&GLASS	690768	OWENS-ILLINOIS INC
165	3241	BUILDING MATERIALS-CEMENT	23904	AMCORD INC
166	3241	BUILDING MATERIALS-CEMENT	374532	GIANT PORTLAND CEMENT CO
167	3241	BUILDING MATERIALS-CEMENT	451542	IDEAL BASIC INDS INC
168	3241	BUILDING MATERIALS-CEMENT	524858	LEHIGH PORTLAND CEMENT CO
169	3241	BUILDING MATERIALS-CEMENT	542290	LONE STAR INDS
170	3241	BUILDING MATERIALS-CEMENT	571443	MARQUETTE CO
171	3241	BUILDING MATERIALS-CEMENT	585072	MEDUSA CORP
172	3241	BUILDING MATERIALS-CEMENT	606215	MISSOURI PORTLAND CEMENT CO
173	3270	CONCRETE GYPSUM & PLASTER	912027	U.S. GYPSUM CO
174	3291	ABRASIVE PRODUCTS	141375	CARBORUNDUM CO
175	3295	MINERAL-EARTH GRD OR OTHRWSE	69869	BASIC INC
176	3297	REFRACTORIES	370622	GENERAL REFRACTORIES CO
177	3310	STEEL-MAJOR	42195	ARMCO STEEL CORP
178	3310	STEEL-MAJOR	87509	BETHLEHEM STEEL CORP
179	3310	STEEL-MAJOR	457470	INLAND STEEL CO
180	3310	STEEL-MAJOR	637844	NATIONAL STEEL CORP
181	3310	STEEL-MAJOR	760779	REPUBLIC STEEL CORP
182	3310	STEEL-MAJOR	912656	U.S. STEEL CORP
183	3311	STEEL-MINOR	93545	BLISS & LAUGHLIN INDS
184	3311	STEEL-MINOR	217687	COPPERWELD CORP
185	3311	STEEL-MINOR	458702	INTERLAKE INC
186	3311	STEEL-MINOR	549866	LUKENS STEEL CO
187	3311	STEEL-MINOR	582273	MCLOUTH STEEL CORP
188	3311	STEEL-MINOR	707355	PENN-DIXIE INDS
189	3331	PRIMARY SMELTING & REFINING	32393	ANACONDA CO
190	3331	PRIMARY SMELTING & REFINING	217525	COPPER RANGE CO
191	3331	PRIMARY SMELTING & REFINING	457686	INSPIRATION CONS COPPER CO
192	3331	PRIMARY SMELTING & REFINING	489314	KENNECOTT COPPER CORP
193	3331	PRIMARY SMELTING & REFINING	717265	PHELPS DODGE CORP
194	3334	ALUMINUM	13716	ALCAN ALUMINIUM LTD
195	3334	ALUMINUM	22249	ALUMINUM CO OF AMERICA
196	3334	ALUMINUM	483008	KAISER ALUMINUM & CHEM CORP
197	3334	ALUMINUM	761763	REYNOLDS METALS CO
198	3350	ROLLING & DRAW NON-FER METL	77455	BELDEN CORP
199	3350	ROLLING & DRAW NON-FER METL	369298	GENERAL CABLE CORP
200	3350	ROLLING & DRAW NON-FER METL	629156	N L INDS
201	3350	ROLLING & DRAW NON-FER METL	761406	REVERE COPPER & BRASS INC
202	3350	ROLLING & DRAW NON-FER METL	895861	TRIANGLE INDS
203	3429	HARDWARE	854616	STANLEY WORKS
204	3430	BLDG MAT-HEAT-AIR COND-PLUMB	29717	AMERICAN STANDARD INC
205	3430	BLDG MAT-HEAT-AIR COND-PLUMB	574599	MASCO CORP
206	3430	BLDG MAT-HEAT-AIR COND-PLUMB	690734	OWENS-CORNING FIBERGLAS CORP
207	3430	BLDG MAT-HEAT-AIR COND-PLUMB	878895	TECUMSEH PRODUCTS CO
208	3430	BLDG MAT-HEAT-AIR COND-PLUMB	892892	TRANE CO
209	3449	METAL WORK-MISC	19411	ALLIED PRODUCTS
210	3449	METAL WORK-MISC	150033	CECO CORP
211	3449	METAL WORK-MISC	853819	STANDARD PRESSED STEEL CO
212	3499	FABRICATED METAL PRODUCTS	253651	DIEBOLD INC
213	3499	FABRICATED METAL PRODUCTS	826690	SIGNODE CORP
214	3511	MACHINERY STEAM GENERATING	200273	COMBUSTION ENGINEERING INC
215	3511	MACHINERY STEAM GENERATING	350244	FOSTER WHEELER CORP
216	3522	MACHINERY-AGRICULTURAL	19645	ALLIS-CHALMERS CORP

SERIAL NUMBER	COMPUSTAT INDUSTRY NUMBER	INDUSTRY NAME	COMPUSTAT COMPANY NUMBER	COMPANY NAME
217	3531	MACHINERY-CONST & MAT HANDLG	118745	BUCYRUS-ERIE CO
218	3531	MACHINERY-CONST & MAT HANDLG	149123	CATERPILLAR TRACTOR CO
219	3531	MACHINERY-CONST & MAT HANDLG	181396	CLARK EQUIPMENT CO
220	3531	MACHINERY-CONST & MAT HANDLG	302491	FMC CORP
221	3533	MACHINERY-OIL WELL	406216	HALLIBURTON CO
222	3540	MACHINE TOOLS	115223	BROWN & SHARPE MFG CO
223	3540	MACHINE TOOLS	172172	CINCINNATI MILACRON INC
224	3540	MACHINE TOOLS	375046	GIDDINGS & LEWIS INC
225	3540	MACHINE TOOLS	609150	MONARCH MACHINE TOOL CO
226	3540	MACHINE TOOLS	830643	SKIL CORP
227	3540	MACHINE TOOLS	867323	SUNDSTRAND CORP
228	3540	MACHINE TOOLS	934408	WARNER & SWASEY
229	3550	MACHINERY-SPECIALTY	291210	EMHART CORP
230	3550	MACHINERY-SPECIALTY	524462	LEESONA CORP
231	3550	MACHINERY-SPECIALTY	597715	MIDLAND-ROSS CORP
232	3550	MACHINERY-SPECIALTY	689002	OTIS ELEVATOR CO
233	3560	MACHINERY-INDUSTRIAL	167898	CHICAGO PNEUMATIC TOOL CO
234	3560	MACHINERY-INDUSTRIAL	365550	GARDNER-DENVER CO
235	3560	MACHINERY-INDUSTRIAL	590825	MESTA MACHINE CO
236	3570	OFFICE & BUSINESS EQUIPMENT	122781	BURROUGHS CORP
237	3570	OFFICE & BUSINESS EQUIPMENT	253579	DICTAPHONE CORP
238	3570	OFFICE & BUSINESS EQUIPMENT	285551	ELECTRONIC ASSOC
239	3570	OFFICE & BUSINESS EQUIPMENT	628862	NCR CORP
240	3570	OFFICE & BUSINESS EQUIPMENT	724479	PITNEY-BOWES INC
241	3570	OFFICE & BUSINESS EQUIPMENT	925853	VICTOR COMPTOMETER CORP
242	3573	ELECTRONIC COMPUTER EQUIP	438506	HONEYWELL INC
243	3579	OFFICE MACHINES N/ELSEWHERE	631226	NASHUA CORP
244	3580	MACHINES-SERVICE INDUSTRY	902878	UMC INDS
245	3580	MACHINES-SERVICE INDUSTRY	922612	VENDO CO
246	3600	ELEC & ELEC LEADERS	369604	GENERAL ELECTRIC CO
247	3600	ELEC & ELEC LEADERS	960402	WESTINGHOUSE ELECTRIC CORP
248	3610	ELEC EQUIP	580628	MCGRAW-EDISON CO
249	3610	ELEC EQUIP	677194	OHIO BRASS CO
250	3610	ELEC EQUIP	884315	THOMAS & BETTS CORP
251	3622	ELEC INDUSTRIAL CONTROLS	232165	CUTLER-HAMMER INC
252	3622	ELEC INDUSTRIAL CONTROLS	852206	SQUARE D CO
253	3630	ELEC HOUSEHOLD APPLIANCE	439272	HOOVER CO
254	3630	ELEC HOUSEHOLD APPLIANCE	578592	MAYTAG CO
255	3630	ELEC HOUSEHOLD APPLIANCE	810640	SCOVILL MFG CO
256	3630	ELEC HOUSEHOLD APPLIANCE	829302	SINGER CO
257	3642	LIGHTING FIXTURES	884425	THOMAS INDS INC
258	3651	RADIO-TV MANUFACTURERS	989399	ZENITH RADIO CORP
259	3662	RADIO-TV TRANSMTTNG EQUIP-AP	620076	MOTOROLA INC
260	3670	ELECTRONICS	23141	AMBAC INDS INC
261	3670	ELECTRONICS	31897	AMP INC
262	3670	ELECTRONICS	208291	CONRAC CORP
263	3670	ELECTRONICS	303693	FAIRCHILD CAMERA&INSTRUMENT
264	3670	ELECTRONICS	421596	HAZELTINE CORP
265	3670	ELECTRONICS	429812	HIGH VOLTAGE ENGINEERING
266	3670	ELECTRONICS	755111	RAYTHEON CO
267	3679	ELECTRONIC COMPONENTS	561246	MALLORY (P.R.) & CO
268	3679	ELECTRONIC COMPONENTS	849339	SPRAGUE ELECTRIC CO
269	3679	ELECTRONIC COMPONENTS	882508	TEXAS INSTRUMENTS INC
270	3711	MOTOR VEHICLES	171196	CHRYSLER CORP
271	3711	MOTOR VEHICLES	345370	FORD MOTOR CO
272	3711	MOTOR VEHICLES	370442	GENERAL MOTORS CORP
273	3713	AUTO TRUCKS	231021	CUMMINS ENGINE
274	3713	AUTO TRUCKS	359370	FRUEHAUF CORP
275	3713	AUTO TRUCKS	964066	WHITE MOTOR CORP
276	3714	AUTO PARTS & ACCESSORIES	17634	ALLEN GROUP
277	3714	AUTO PARTS & ACCESSORIES	43339	ARVIN INDS INC
278	3714	AUTO PARTS & ACCESSORIES	99725	BORG-WARNER CORP
279	3714	AUTO PARTS & ACCESSORIES	118835	BUDD CO
280	3714	AUTO PARTS & ACCESSORIES	278058	EATON CORP
281	3714	AUTO PARTS & ACCESSORIES	313549	FEDERAL-MOGUL CORP
282	3714	AUTO PARTS & ACCESSORIES	530000	LIBBEY-OWENS-FORD CO
283	3714	AUTO PARTS & ACCESSORIES	566472	MAREMONT CORP
284	3714	AUTO PARTS & ACCESSORIES	630395	NAPCO INDS INC
285	3714	AUTO PARTS & ACCESSORIES	872649	TRW INC
286	3714	AUTO PARTS & ACCESSORIES	887389	TIMKEN CO
287	3721	AEROSPACE	97023	BOEING CO
288	3721	AEROSPACE	231561	CURTISS-WRIGHT CORP

SERIAL NUMBER	COMPUSTAT INDUSTRY NUMBER	INDUSTRY NAME	COMPUSTAT COMPANY NUMBER	COMPANY NAME
289	3721	AEROSPACE	369550	GENERAL DYNAMICS CORP
290	3721	AEROSPACE	400181	GRUMMAN CORP
291	3721	AERUSPACE	539821	LOCKHEED AIRCRAFT CORP
292	3721	AEROSPACE	884102	THIOKOL CORP
293	3721	AEROSPACE	913017	UNITED TECHNOLOGIES CORP
294	3740	RAILROAD EQUIPMENT	361448	GATX CORP
295	3740	RAILROAD EQUIPMENT	370856	GENERAL STEEL INDS
296	3740	RAILROAD EQUIPMENT	745791	PULLMAN INC
297	3740	RAILROAD EQUIPMENT	854701	STANRAY CORP
298	3811	ENGR LAB & RESEARCH EQUIP	31105	AMETEK INC
299	3811	ENGR LAB & RESEARCH EQUIP	640745	NEPTUNE INTL CORP
300	3821	MECH-MEAS-CONT INSTR	370838	GENERAL SIGNAL CORP
301	3822	AUTOMATIC TEMPERATURE CONT	478366	JOHNSON CONTROLS INC
302	3822	AUTOMATIC TEMPERATURE CONT	770519	ROBERTSHAW CONTROLS
303	3831	OPTICAL	71707	BAUSCH & LOMB INC
304	3861	PHOTOGRAPHIC	77851	BELL & HOWELL CO
305	3861	PHOTOGRAPHIC	277461	EASTMAN KODAK CO
306	3861	PHOTOGRAPHIC	604059	MINNESOTA MINING & MFG CO
307	3861	PHOTOGRAPHIC	731095	POLAROID CORP
308	3948	LEISURE TIME PRODUCTS	1688	AMF INC
309	3999	MFG INDS	410306	HANDY & HARMAN
310	3999	MFG INDS	457659	INSILCO CORP
311	3999	MFG INDS	776338	RONSON CORP
312	3999	MFG INDS	912078	U.S. INDS
313	4210	TRUCKING	209237	CONSOLIDATED FREIGHTWAYS INC
314	4210	TRUCKING	588602	MERCHANTS INC
315	4210	TRUCKING	690326	OVERNIGHT TRANSPORTATION
316	4210	TRUCKING	769739	ROADWAY EXPRESS INC
317	4400	SHIPPING	615798	MOORE MCCORMACK RESOURCES
318	4511	AIR TRANSPORT	23771	AMERICAN AIRLINES INC
319	4511	AIR TRANSPORT	105425	BRANIFF INTL CORP
320	4511	AIR TRANSPORT	210795	CONTINENTAL AIR LINES INC
321	4511	AIR TRANSPORT	276191	EASTERN AIR LINES
322	4511	AIR TRANSPORT	667281	NORTHWEST AIRLINES INC
323	4511	AIR TRANSPORT	698057	PAN AMERICAN WORLD AIRWAYS
324	4511	AIR TRANSPORT	811641	SEABOARD WORLD AIRLINES
325	4511	AIR TRANSPORT	893349	TRANS WORLD AIRLINES
326	4511	AIR TRANSPORT	902550	UAL INC
327	4511	AIR TRANSPORT	957586	WESTERN AIR LINES INC
328	4712	FREIGHT FORWARDING	291101	EMERY AIR FREIGHT CORP
329	4712	FREIGHT FORWARDING	894015	TRANSWAY INTERNATIONAL CORP
330	4811	TELEPHONE COMPANIES	30177	AMERICAN TELE & TELEGRAPH
331	4811	TELEPHONE COMPANIES	371028	GENERAL TELEPHONE&ELECTRONCS
332	4811	TELEPHONE COMPANIES	959805	WESTERN UNION CORP
333	4830	RADIO-TV BROADCASTERS	24735	AMERICAN BROADCASTING
334	4830	RADIO-TV BROADCASTERS	124845	CBS INC
335	4830	RADIO-TV BROADCASTERS	862131	STORER BROADCASTING CO
336	5050	WHOLESALE-METALS & MINERALS	148411	CASTLE (A.M.) & CO
337	5093	WHOLESALE-SCRAP&WASTE MTRLS	676346	OGDEN CORP
338	5140	WHOLESALE-FOODS	252435	DI GIORGIO CORP
339	5140	WHOLESALE-FOODS	339130	FLEMING CO'S, INC
340	5311	RETAIL-DEPARTMENT STORES	31141	AMFAC INC
341	5331	RETAIL-VARIETY STORES	626643	MURPHY (G.C.) CO
342	5411	RETAIL-FOOD CHAINS	72797	BAYLESS (A. J.) MARKETS INC
343	5411	RETAIL-FOOD CHAINS	501044	KROGER CO
344	5411	RETAIL-FOOD CHAINS	786514	SAFEWAY STORES INC
345	5661	RETAIL-SHOE STORES	280875	EDISON BROTHERS STORES
346	5661	RETAIL-SHOE STORES	585745	MELVILLE CORP
347	5812	EATING PLACES	441074	HOST INTL INC
348	6140	FINANCE	125569	C.I.T. FINANCIAL CORP
349	6145	FINANCE-SMALL LOAN	26879	AMERICAN INVESTMENT CO
350	6145	FINANCE-SMALL LOAN	81721	BENEFICIAL CORP
351	6145	FINANCE-SMALL LOAN	119007	BUDGET INDS
352	6145	FINANCE-SMALL LOAN	441812	HOUSEHOLD FINANCE CORP
353	6145	FINANCE-SMALL LOAN	530710	LIBERTY LOAN CORP
354	7011	HOTELS-TOURIST COURTS-MOTELS	432848	HILTON HOTELS CORP
355	7213	LINEN SUPPLY-INDL LAUNDERS	812370	SEARS INDS INC
356	7399	SERVICES-GENERAL BUSINESS	25231	AMERICAN CONSUMER INDS
357	7810	MOTION PICTURES	901221	TWENTIETH CENTURY-FOX FILM
358	9997	CONGLOMERATES	826622	SIGNAL COS
359	9997	CONGLOMERATES	883203	TEXTRON INC

VITA

William Kemp Carter

Candidate for the Degree of

Doctor of Philosophy

Thesis: AN ALTERNATIVE TO CONVENTIONAL INDUSTRY DEFINITIONS IN MODELING EARNINGS GENERATION PROCESSES

Major Field: Business Administration

Biographical:

Personal Data: Born in Los Angeles, California, October 31, 1951, the son of Mr. and Mrs. E. Aubert Carter.

Education: Graduated from South Pike High School, Magnolia, Mississippi, in May, 1969; attended William Carey College, Hattiesburg, Mississippi, 1969-1970; received Bachelor of Science in Business Administration from the University of Southern Mississippi in 1973, received Master of Science degree in Accounting from the University of Mississippi in 1974; enrolled in doctoral program at Oklahoma State University in 1974; completed requirements for the Doctor of Philosophy degree in December, 1977.

Professional Experience: Staff Accountant with the firm of Burke, Burke, and MacArthur, CPA, Hattiesburg, Mississippi, in 1973 and 1974; graduate teaching assistant at the University of Southern Mississippi, 1973-1974; teaching associate at Oklahoma State University, 1974-1977.

Professional Organizations: Member of the Institute of Internal Auditors and the American Accounting Association.