1971

A Statistical Grouping of Corporations by their Financial Characteristics.

William (Bill) H. Williams
CUNY Hunter College

Michael L. Goodman

JOURNAL OF FINANCIAL AND QUANTITATIVE ANALYSIS
September 1971

A STATISTICAL GROUPING OF CORPORATIONS BY THEIR FINANCIAL CHARACTERISTICS

*W. H. Williams and M. L. Goodman**

I. Introduction

It appears to be a widely held view that corporations with similar operational characteristics ought to have similar financial characteristics. For example, one might expect that the financial characteristics of two drug companies would be similar. This seems entirely reasonable. Unfortunately, however, there does not appear to be any quantitative analysis of this point in the literature. Furthermore, discussions with our financial colleagues lead to the conclusion that, if such financial differentiation of corporations were possible, it is by no means obvious what the variables of differentiation would be. Consequently, such an analysis was undertaken and is described in this paper.

The basic question asked is whether the *statistical* grouping of corporations by their financial data characteristics is similar to their predetermined, *external*, industrial classification. If these groupings are similar, it would reveal that companies with similar operations also have similar financial characteristics.

II. Data

The data used in this analysis were taken from Standard and Poor's Compustat tapes which were modified and expanded somewhat by the American Telephone and Telegraph Company. One of the Standard and Poor's tapes contains information on approximately 900 industrial organizations and another contains information on 175 utilities. These two tapes were merged into one and the data were examined to ensure consistency over the time span for stock splits and other characteristics that would affect the comparability of the financial characteristics.

The data were available for a twenty-two year span from 1946 to 1967. Observations existed on fifty-seven different economic variables and several identification variables. Fourteen important variables were selected from the available fifty-seven. Some of these fourteen are functions of two or more of the original variables. The fourteen are listed in Table 1.

**The University of Michigan and Bell Telephone Laboratories, Incorporated, Murray Hill, New Jersey, respectively.*

TABLE 1

VARIABLES USED IN THE ANALYSIS

1. Price per share
2. Earnings per share
3. Dividends per share
4. Price-to-earnings ratio
5. Dividends-to-earnings ratio
6. Dividends-to-price ratio
7. Total market value of stock traded (= number of shares traded x price per share)
8. Total market value of stock (= number of shares outstanding x price per share)
9. Percent of stock traded (= number of shares traded ÷ number of shares outstanding)
10. Total market value of stock ÷ total assets
11. Total dividends ÷ total assets
12. Operating income ÷ total assets
13. Total assets ÷ number of shares outstanding
14. Debt equity ratio

III. Utility versus Industrial Discrimination

An examination of the grouping characteristics of industrials and utilities was undertaken first. Specifically, is it possible to distinguish between corporations classified as utilities and corporations classified as industrials on the basis of their financial characteristics *alone?* And if so, with what accuracy? In this part of the analysis, the seventh, ninth, and fourteenth variables were not used because information on the number of shares traded and the dept equity ratio were not available for utilities.

The first step was the creation of scatter plots with respect to all possible pairs of the eleven remaining variables. There were thus fifty-five such plots. Most of these plots showed much overlapping between the two different groups with no discrimination evident. But some of the plots (see Figures I, II, and III) showed a degree of discrimination in two dimensions. In these figures, utilities are identified by circles and industrials by pluses. From the form of the plots, it is quite clear that one is able to distinguish between the utilities as a group and the industrials as a group on the basis of certain financial characteristics alone.

To approach the question of accuracy, a two-group discriminant analysis was used (cf. Anderson [1]). This procedure looks at the data as points in eleven-dimensional space and assumes that the points belong to two groups in this eleven-dimensional space. A ten-dimensional hyperplane is then drawn through the space, and points (viz., the corporations) on one side of the plane are assigned to one group and points on the other side to the other group. One way of analyzing how well a technique like this works is to see how many observations are properly classified. When this technique was applied to the 1966 data with 875 industrials and 131 utilities, it was found that almost 98 percent of these organizations were properly identified. The summary of this analysis is given in the top half of Table 2. The F-statistic, which measures the significance of the difference between the centroids of the two groups, is overwhelmingly significant. Similarly, when the technique was applied to 1967 data, it was again found that 98 percent of the observations were correctly identified. On the basis of this analysis, it certainly seems clear that it is a relatively simple matter to classify corporations as industrials or utilities simply by being made aware of their financial characteristics.

IV. Discrimination within Industrial Classes

A discrimination between utilities and industrials was relatively easy to develop, but could the same techniques be extended to the more difficult case of separation among the industrial subclassifications?

The Standard and Poor's Corporation groups the industrials on the Compustat tape into 112 different categories. The basis for classification is the main product of the company. Five of the largest categories were arbitrarily selected: chemicals, drugs, domestic oils, steels, and electronics. The question now is whether these external classifications would in any way match the statistical groupings of the corporations obtained by using their financial characteristics. As in the earlier case, the first step was to construct scatter plots of the financial variables. In this case, the fourteen variables gave rise to ninety-one plots similar to those shown for the earlier discrimination. An examination of the plots suggested that industrial classes did seem to be grouped with respect to certain pairs of the variables. To test the importance of these groupings, a multigroup discriminant analysis technique was applied (cf. Anderson [1]).

This procedure divides the space into five regions and assigns each observation to the group in whose region it falls. The regions are determined on the basis of the data

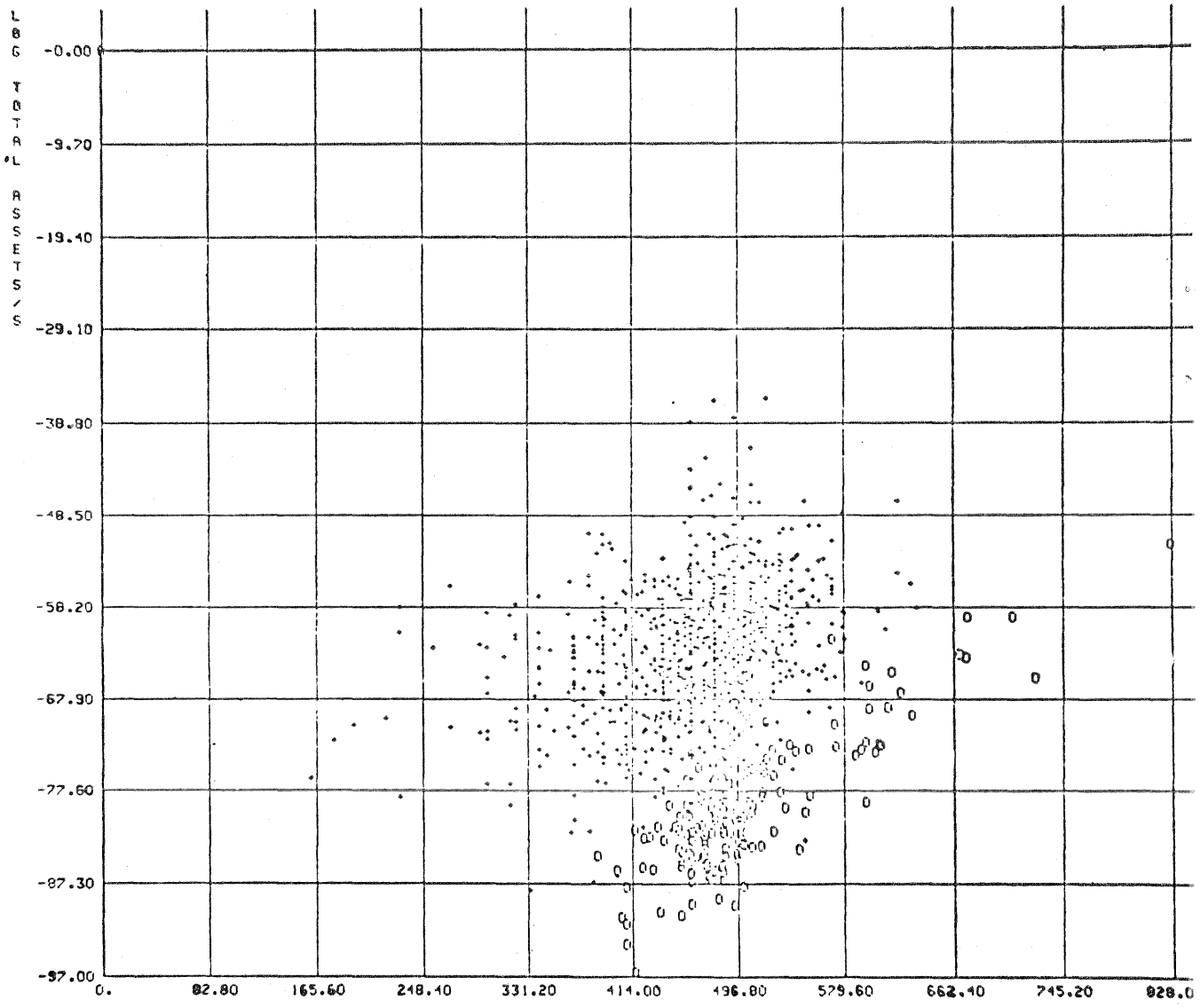
LOG TOTAL ASSETS/S
-0.00
-9.70
-19.40
-29.10
-38.80
-48.50
-58.20
-67.90
-77.60
-87.30
-97.00
0.
82.80
165.60
248.40
331.20
414.00
496.80
579.60
662.40
745.20
828.0
LOG DIVIDENDS/SHARE
STOCK ANALYSIS
INDUSTRIAL COUNT = 820
UTILITY COUNT = 173
XSCALE = -2

FIGURE I

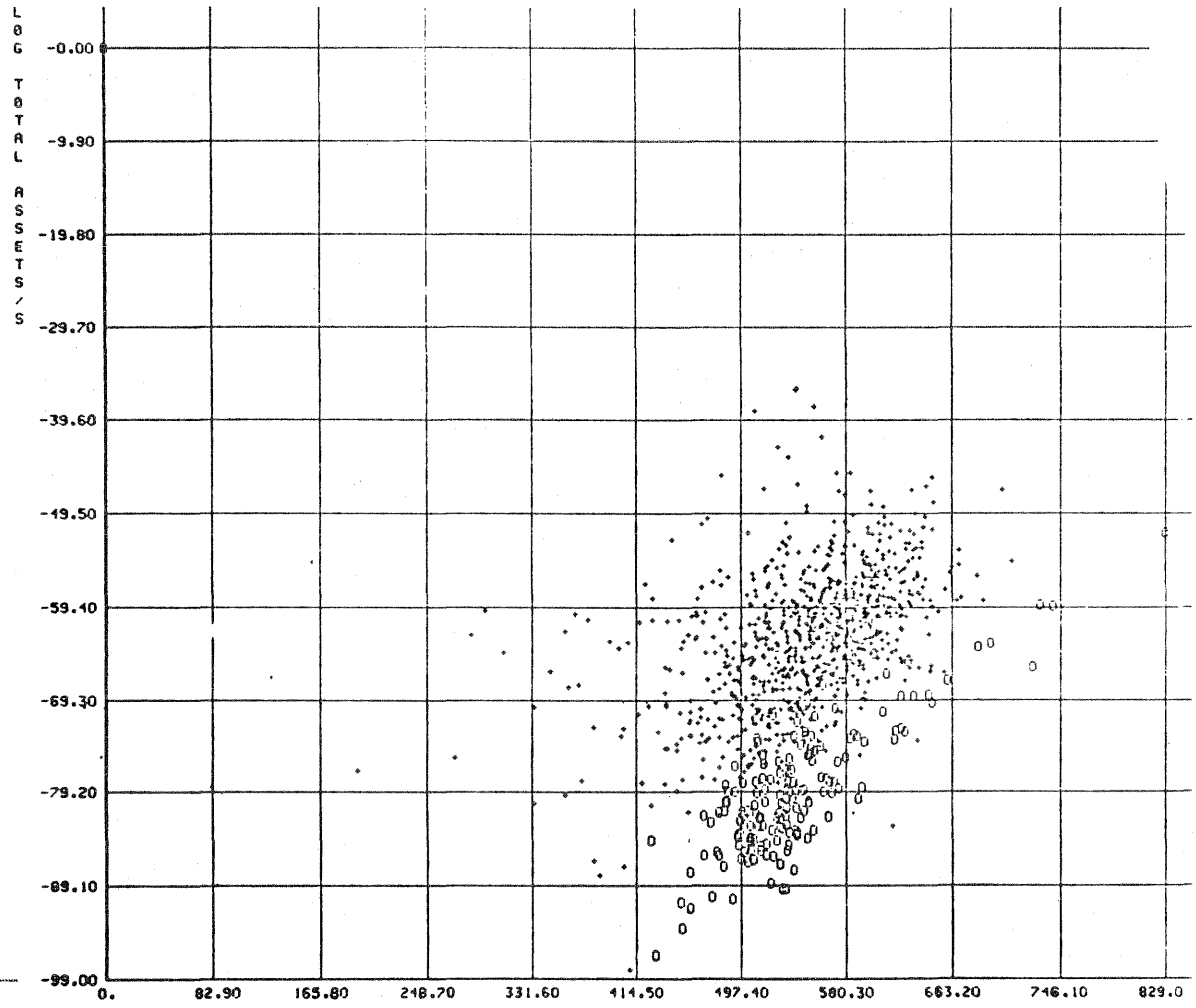
LOG TOTAL ASSETS/S
-0.00
-9.90
-19.80
-29.70
-39.60
-49.50
-59.40
-69.30
-79.20
-89.10
-99.00
0.
82.90
165.80
248.70
331.60
414.50
497.40
580.30
663.20
746.10
829.0
LOG EARNINGS/SHARE
STOCK ANALYSIS
INDUSTRIAL COUNT = 889
UTILITY COUNT = 173
XSCALE = -2
YSCALE = -1

FIGURE II

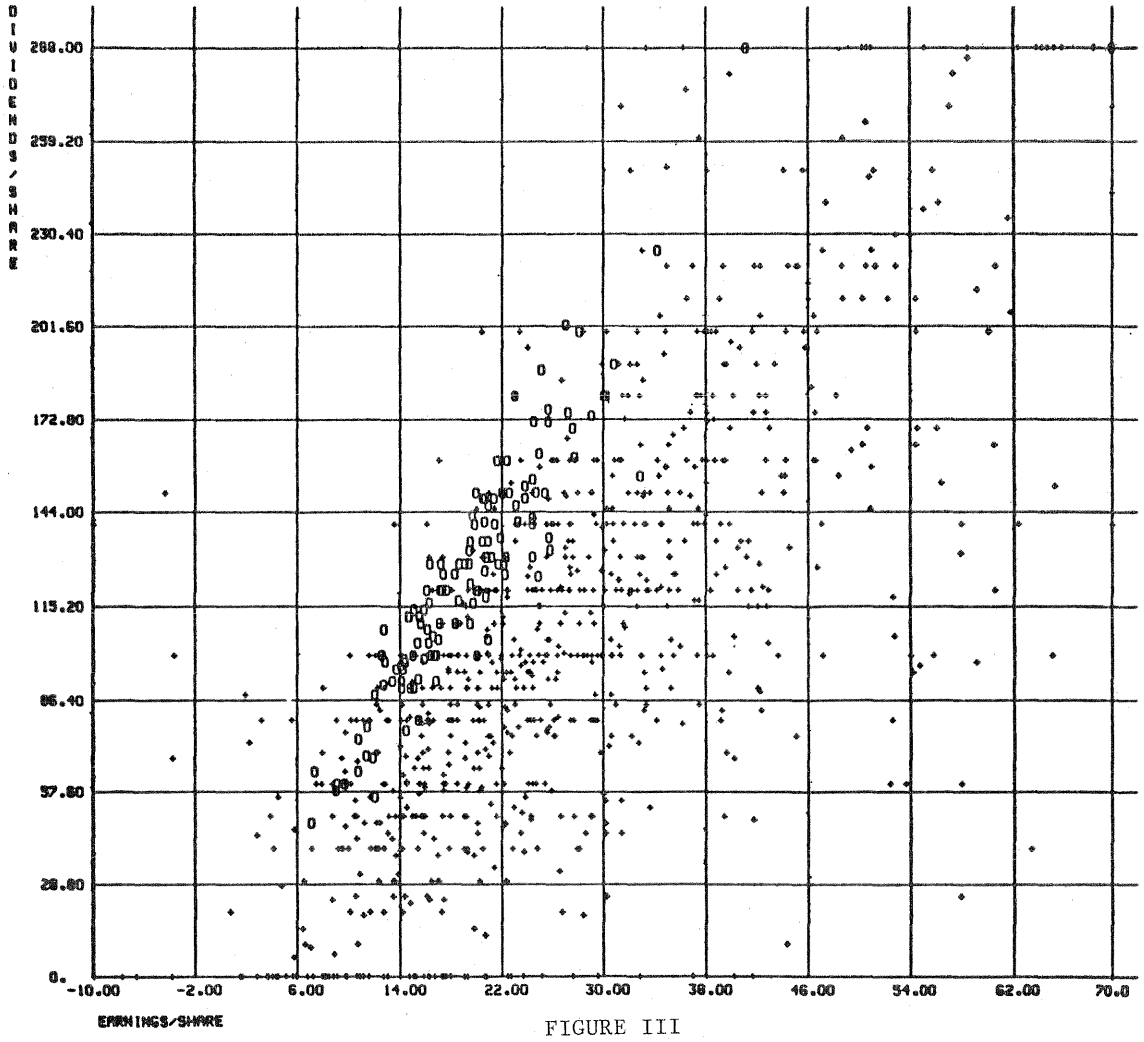
DIVIDENDS/SHARE
288.00
259.20
230.40
201.60
172.80
144.00
115.20
86.40
57.60
28.80
0.
-10.00
-2.00
6.00
14.00
22.00
30.00
38.00
46.00
54.00
62.00
70.0
EARNINGS/SHARE
STOCK ANALYSIS WITH 875 INDUSTRIALS AND 131 UTILITIES
XSCALE = -1
YSCALE = -2

FIGURE III

TABLE 2

DISCRIMINATING BETWEEN INDUSTRIALS AND UTILITIES

1966 Data

		Classed as 1.	2.	Total	Proportion Correct
1.	Industrials	866	9	875	.990
2.	Utilities	12	119	131	.908

Total: .979 correct identification

Mahalanobis D^2 = 23.84080

F(11,994) = 244.49065*

1967 Data

		Classed as 1.	2.	Total	Proportion Correct
1.	Industrials	829	9	838	.989
2.	Utilities	7	123	130	.946

Total: .983 correct identification

D^2 = 25.43787

F(11,956) = 257.56141*

*Significant at level .001

in the light of the external grouping of the observations, and one would hope that most of the points in a group would fall in the proper region of the space. In this analysis each of the twenty-two years was considered separately; that is, for each year a separate set of discriminant functions was computed. The number of companies misclassified over the twenty-two years was counted, and it was found that 72 percent of the companies were properly identified. This is not as good as the utility-industrial discrimination; however, one must remember that, if companies had been assigned to groups at random, one would expect to properly identify only 20 percent of them. In addition, the generalized Mahalanobis D^2-statistic, which is a measure of the distance between the centroids of the five groups, ranged from 201 to 604. This may be thought of as a chi-square statistic and it is overwhelmingly significant, so that the groups are very definitely separated. It is interesting, moreover, to note in Table 3 that some groups are confused more often than others. For example, chemicals and drugs are misclassified as one another quite often, although oils and electronics are seldom confused. The reason for this may well be that chemicals and drugs are producing somewhat similar products, and companies that produce similar products may indeed have similar financial characteristics in spite of their different basic classifications. It is also interesting to notice that certain companies are consistently misclassified. For example, among the chemicals, Chemtron, Interchemical and Koppers were properly identified only five times in the twenty-two discriminant analyses, and Kaweki Chemicals was never identified as a chemical in the nine years for which data were available. Among the drug companies, Kendall Company was never properly identified in eight years, and Kerr-McGee and Murphy Oil were seldom put into the oil classification. Among the steels, Dominion Foundries was classified as a steel company only once in seven years, while AMP Incorporated and General Signal were almost never classified as electronics companies.

In summary, the multivariate discriminant analysis found the existence of rather tight groups except for a few outlying companies. It is worth observing that eliminating these outlying companies from the analysis changed the percentage of correct identification only from 72 percent to 73 percent and produced a new set of misclassified companies. It would be more interesting to consider whether the originally misclassified companies are operating somewhat differently from their basic classification. For example, are some of the misclassified drug companies really operating more like chemical companies?

V. Use of Discriminant Functions for Measuring Movements of Groups

An additional use of the discriminant function was to measure year-to-year changes in the movement of groups. The procedure here is to compute the discriminant functions for a given year and then apply those functions to the data for the following year. This was done and it was found that more than 60 percent of the companies were still correctly identified, as against the 20 percent that would be expected if companies had been assigned at random. A plausible conclusion from this is that the centroids of the groups move very little from year to year in relation to the distance between them.

The technique can be used for forecasting which group a company will fall into. Unfortunately, it is not very accurate for periods in which the stock market shows a marked change from one year to the next. For example, in the years in which the stock market moves up markedly, many stocks are classified as drugs or chemicals, which are generally the two highest priced groups of stocks.

VI. Summary

This paper has reported on an investigation of the relationship between the external industrial classification of corporations and the statistical groupings that can be found

TABLE 3

DISCRIMINATION AMONG FIVE INDUSTRIAL CLASSES

Using 1946-1967 Data

		Classed as 1.	2.	3.	4.	5.	Total	Proportion Correct
1.	Chemicals	408	96	63	64	15	646	.632
2.	Drugs	47	261	6	4	12	330	.791
3.	Domestic Oils	36	22	298	37	1	394	.756
4.	Steel	40	21	59	398	14	532	.748
5.	Electronics	15	15	1	10	147	188	.782

Total: .723 correct identification

Using Last Year's Discriminant Functions on This Year's Data

		Classed as 1.	2.	3.	4.	5.	Total	Proportion Correct
1.	Chemicals	329	102	97	68	27	623	.528
2.	Drugs	57	232	9	6	15	319	.727
3.	Domestic Oil	50	24	271	33	2	381	.711
4.	Steel	79	33	100	268	35	515	.520
5.	Electronics	23	22	0	15	125	185	.676

Total: .606 correct identification

in their corporate financial measurements. It appears that utilities and industrials are quite different in their financial characteristics. Five industrial classes also appear to be quite distinct, although somewhat less distinct than the industrial-versus-utility comparisons. Consequently, it appears that financial variables do tend to distinguish the various industrial classifications and that, with only a corporation's financial characteristics known, its industrial classification may be reliably determined.

REFERENCE

[1] Anderson, T. W. *An Introduction to Multivariate Statistical Analysis.* New York, N. Y.: Wiley & Sons, 1958.