1973

Ambiguities in the Cross-Section Analysis of Per Share Financial Data

E G. Davis
Carleton University

D M. Dunn

William (Bill) H. Williams
CUNY Hunter College

AMBIGUITIES IN THE CROSS-SECTION ANALYSIS OF PER SHARE FINANCIAL DATA

E. G. DAVIS, D. M. DUNN AND W. H. WILLIAMS*

I. INTRODUCTION

IN ANALYZING the financial characteristics of a particular corporation, it is standard procedure to adjust the data to reflect the current number of shares outstanding. This is necessary to put the various financial measurements on a basis in which the historical time series observations are comparable. If such corrections were not made, many of the data series (e.g., stock prices) would contain disconcerting jumps, proportional in height to the magnitude of any stock splits. Such jumps would cause obvious technical difficulties; for example, they can easily distort autocorrelations by either masking real correlations or signalling false ones.

To illustrate explicitly, if the corporation under study underwent a two for one (say) stock split, then to make the historical financial data comparable through time, the customary procedure is to adjust the previously reported per share measurements (for example earnings and price per share) to the *current* basis. When these corporate series are made internally consistent, standard statistical techniques can be used in the financial analysis. This includes multivariate analysis of the various time series available on a single corporation. For statistical purposes, a single corporation can be analyzed on an *any* year basis, provided the observations are comparable. Arithmetically, it is simply a matter of a common divisor. Most often, analysts work on a current basis, however, because that is the basis which is most meaningful at that time. This is the procedure followed by the popular Standard and Poor's Compustat tapes.

However, it is shown in this paper that per share observations are *not* appropriate for *inter*company analysis. Thus, while price per share and earnings per share may be studied in the analysis of a single corporation, they cannot be legitimately used in cross-sectional studies. The results can be highly misleading, for the arbitrary number of shares outstanding can have a dramatic impact. For example, one could study the data available for the oil industry in 1967 on the basis of the number of shares outstanding in either 1967 or 1972. Since both are studies of 1967, one might intuitively expect the same results in either case. Unfortunately this is not so. Not only do the obvious statistical characteristics such as univariate means and variances of the per share variables change, but so also do regression and correlation coefficients between these variables. The mean and variance changes may not

* E. G. Davis is presently Assistant Professor of Economics at Carleton University, Ottawa, Canada. D. M. Dunn and W. H. Williams are Members of Technical Staff in the Mathematics and Statistics Research Center, Bell Laboratories, Murray Hill, New Jersey, 07974.

always be critical, but changes in the relationships among variables are certainly very important. Furthermore there is no simple way to compare correlations computed on the basis of two different years. Such a correlation is influenced by the number of shares which each company has outstanding at the time of the analysis. In Section II we shall illustrate this by use of an artificial example and by some real examples taken from the domestic oil industry.

It would not be necessary to discuss this bias if cross-sectional studies using per share financial measurements did not appear in the literature. But they do; the literature contains numerous examples of cross-section studies which involve stock price and earnings per share. A possible explanation is that much of the financial modeling that appears in the literature legitimately attempts to develop relationships among the various financial flows of a single corporation. However, when the empirical analysis is undertaken, cross-sectional data are used without making appropriate modifications.

Clear examples of this are contained in Diamond [2], Friend and Puckett [4] and Gordon [5, 7]. Fisher [3] develops corporate models based on price per share but upon reaching the empirical stage he correctly switches to a dividend scaled version of stock price. Fisher, however, does not seem to recognize the problem completely because he makes the statement that estimates of the regression parameters in the unscaled models yield unbiased estimates of the parameters. This is misleading because the parameter estimates depend upon the number of shares outstanding. Gordon [5] seems to recognize these difficulties with cross-sectional studies but presents his results anyway, apparently without fully recognizing the impossibility of interpretation. In the same paper, Gordon also exhibits results appropriately scaled by book value, but in his later book [7] he again presents empirical results which contains these cross-section bias difficulties. Included are the results associated with his Nolev [p. 149], Simlev [p. 155] and Adlev [p. 167] models. As we shall show, the interpretation of any per share cross-sectional model is anything but clear.

## II. Magnitude of the Cross Section Distortion

A. *A Numerical Example*

To give some idea of the potential magnitude of the distortion, we shall first discuss a constructed example. A population of twenty-four corporations was generated such that the market price per share of each company was normally distributed with a mean of 30 and a standard deviation of 5, $N(30, 5)$. Earnings per share were generated independently from a normal distribution, $N(2, 0.35)$. Then, the prices and earnings were randomly paired, so that the simulated industry of twenty-four corporations had an expected price-earnings correlation of zero. The sample correlation between price/share and earnings/share is -0.10. Next we supposed that one company underwent a three for one stock split. This company was randomly selected from the half of the corporations with the largest prices since companies with higher stock prices are presumably those most likely to undergo a stock split. The regression and correlation parameters were then recalculated and compared. This simulates

TABLE 1
ARTIFICIAL STOCK SPLITS

Stock Split	Correlation	Estimated Intercept	Estimated Slope
Original 24 companies	−0.10	32.27	−1.17
One split, 3/1	0.37	20.51	4.74
Two splits, 2/1, 4/1	0.56	15.65	7.08
Three splits, 2/1, 3/1, 4/1	0.52	13.88	7.98
Four splits, 2/1, 3/1, 4/1, 5/1	0.71	9.74	9.89

an analysis of one year done with two different bases, specifically the number of shares before and after the split. The estimated correlation between price per share and earnings per share changes abruptly to 0.37. Similarly, other companies were randomly selected for stock splits and the regression parameters computed. The stock splits and estimated parameters are shown in Table 1. Figure 1 shows a plot of the data before any of the simulated splits
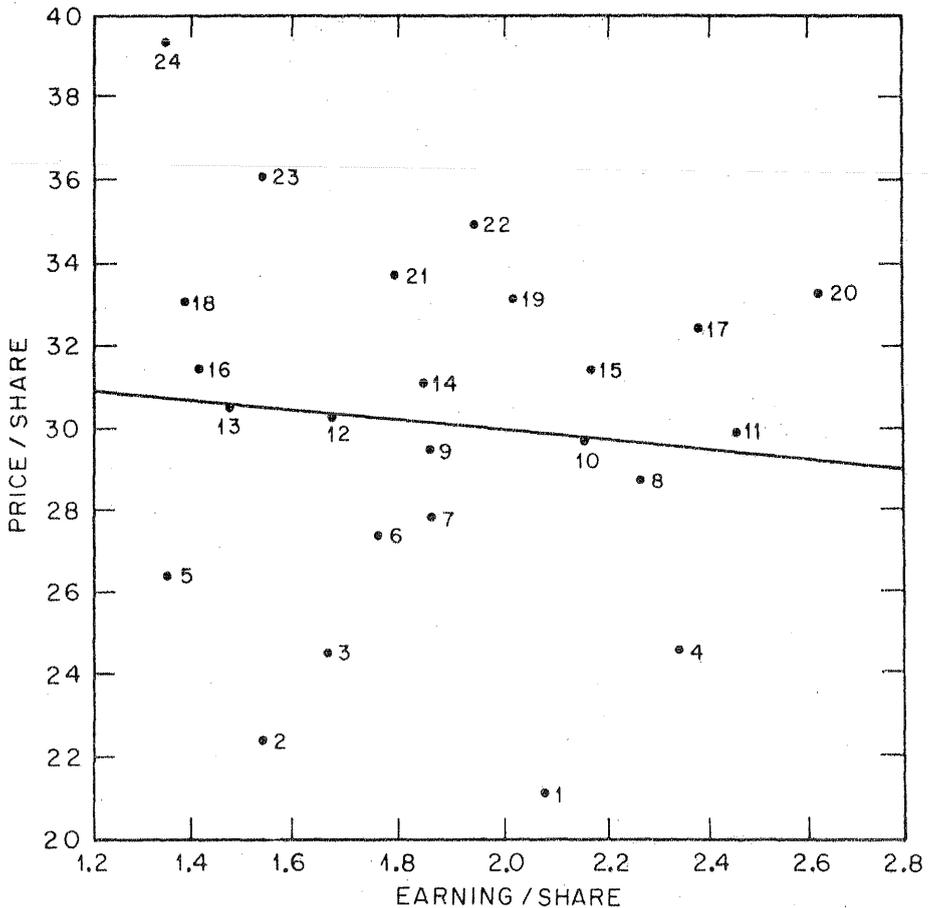
EXAMPLE DERIVED FROM NORMAL DISTRIBUTION

FIGURE 1

and Figure 2 shows the data after four splits had occurred. The change in the configuration of the points (and hence the change in regression parameters) is clearly visible.

As is evident from Table 1, the estimated regression parameters undergo equally dramatic changes. Notice that while in this case we started with a small correlation and wound up with a large one, we could have also started with a large correlation and distorted it downward. The key point, however, is that the parameter estimates are impossible to interpret whatever their estimated value.

B. *The Domestic Oils*

To get some idea of the variability in estimated parameters that one might encounter in practice, we picked 16 companies from the domestic oil industry. This industry was selected because it had been found (see Chen et al. [1])
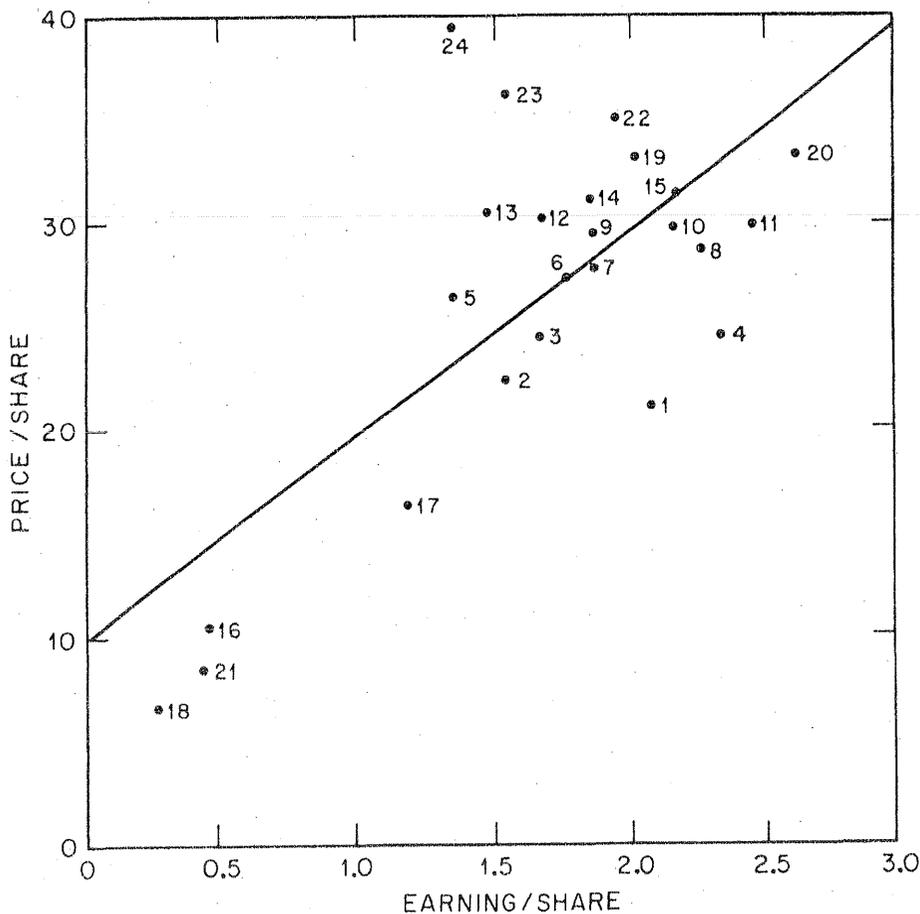
NORMAL DATA WITH 4 STOCK SPLITS, 2/1, 3/1, 4/1 AND 5/1

FIGURE 2

to be one of the more homogeneous groups appearing on the Standard and Poor's Compustat Tapes. However, the data used in this study (gathered by the authors from standard data sources) were more frequent in time than those found on the Compustat Tapes and consisted of monthly observations on the closing per share stock price, quarterly observations on the earnings and dividends per share, and yearly observations on numerous other financial variables (for example: total sales, income tax rate, and net income). In each case observations were available from January 1960 through December 1969 (ten years of data).

For each year of available data we estimated the least squares regression slopes and correlation coefficients for the regression of earnings/share on price/share using that year as a basis as well as all subsequent years. So, for example, we have parameters estimated for 1965 on the basis of 1965, 1966, 1967, 1968 and 1969. Thus we can see how the results would differ if we had studied the year 1965 in 1969 on a 1969 basis (say) as opposed to a 1965 basis. Table 2 shows the estimated correlation coefficients for
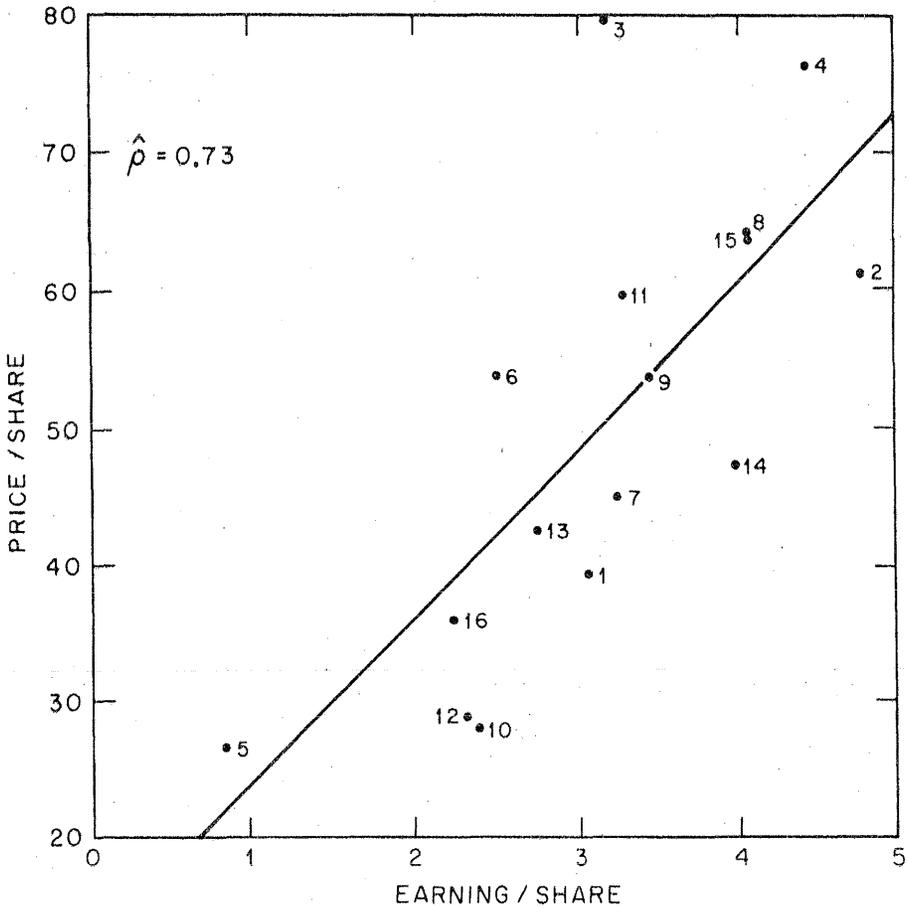
TABLE 2

CORRELATION BETWEEN PRICE PER SHARE AND EARNINGS PER SHARE FOR DOMESTIC OILS

		Basis			
		1965	1966	1967	1968	1969
Year	1965	.84	.86	.88	.86	.90
	1966		.80	.85	.82	.90
	1967			.73	.70	.81
	1968				.57	.79
	1969					.77

1965-1969. There is some change in the correlation coefficients and a tendency for slopes to increase as the basis year changes. This increase is similar to (if somewhat less dramatic than) the situation modeled in the artificial example. The effect of a single stock split on the regression slope is a function of the relative positions of the estimated regression and the line joining the points which represent the company's position before and after the split.

Plots showing 1964 on a 1964 and 1969 basis are shown in Figures 3 and 4 for illustration. The observations enclosed in boxes are those companies which have undergone stock splits. The number adjacent to the box indicates the cumulative stock splits which that company underwent in the period from 1964 through 1969. It is obvious, but worthy of comment, that if more firms should decide on stock splits (in particular, numbers 7, 11, 13, and 14) then the correlation could drop to a level more like that using 1964 as a basis. Hence, not only can correlations be induced in actual data but they can also be removed.

It should be noted that there is no natural variation in the rows of Table 2. An acceptable method of data analysis, given a particular year (row) and a fixed set of companies would dictate that the estimated parameters be *identical* for all of the different basis years. They clearly are not.
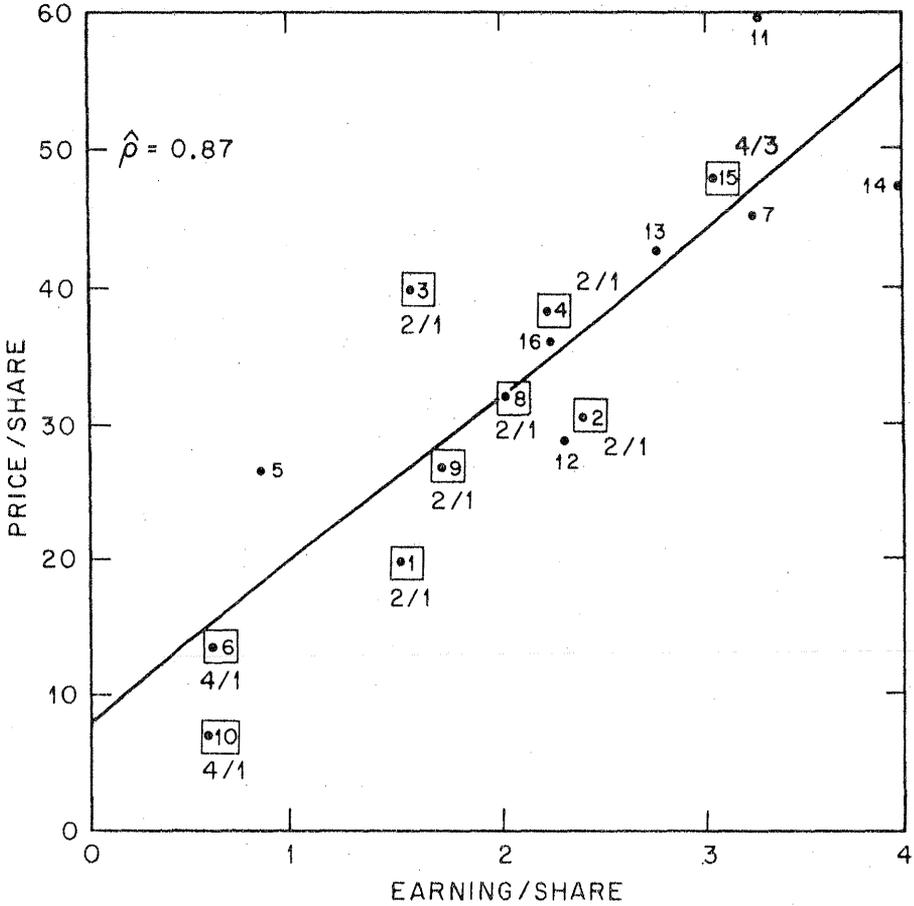
DOMESTIC OILS – 1964 ON 1964 BASIS

FIGURE 3

Analysis of 1964 on 1964 Basis

The other half of Table 2 was in fact calculated but it is not presented because the analysis of a given year on the basis of an earlier year will not normally happen in practice.

III. SUMMARY AND DISCUSSION

In this paper we have shown that there are serious difficulties present in cross-sectional analyses which involve per share data. Such data are a function of an essentially arbitrary divisor (the number of shares outstanding) which makes any results completely meaningless for cross-sectional analysis. This unfortunate characteristic holds for any measurement which is on a per share basis even though in this paper our examples have involved only price per share and earnings per share.

FIGURE 4
Analysis of 1964 on 1969 Basis

The use of per share data is an attempt to appropriately scale a total, such as total earnings. The reason for this is that the use of a "total" in cross-section analysis also distorts the results due to the differences in size among the members of the population studied. Consequently, variables such as total earnings cannot be compared in a cross-sectional study because they simply reflect corporate size. One company's earnings may be substantially larger than another but when compared with a size measure, the second company may be clearly seen to be relatively more efficient and profitable than the first. Explicit recognition of the role of size needs to be included in cross-sectional analyses by the use of objective rather than arbitrary size measures. For example, one might consider earnings per gross sales dollar.

Finally, it should be realized that a complete analysis of these financial

variables must include explicit recognition of the multivariate nature of the problem. Once a measure of size has been chosen and included in the analysis, one could then study other variables which are orthogonal to it. Such a study, including an investigation of the relations which exist between approximately scaled market price and earnings, will be the subject of a future work. In this paper we have aimed our discussion at the illustration of the difficulties brought about by stock split distortions.

REFERENCES

1. H. Chen, et al. "A Statistical Study of the Grouping of Corporations," *Proceedings of the Business and Economics Statistics Section—American Statistical Association*, (Detroit, 1970), pp. 447-451.
2. James J. Diamond. "Earnings Distribution and the Valuation of Shares: Some Recent Evidence," *Journal of Financial and Quantitative Analysis*, Vol. 2, (March 1967), pp. 14-30.
3. G. R. Fisher. "Some Factors Influencing Share Prices," *The Economic Journal*, Vol. 71, (March 1961), pp. 121-41.
4. Irwin Friend and Marshall Puckett. "Dividends and Stock Prices," *American Economic Review*, Vol. 54, No. 5, (September 1964), pp. 656-682.
5. M. J. Gordon. "Dividends, Earnings, and Stock Prices," *Review of Economics and Statistics*, Vol. 41, (May 1959), pp. 99-105.
6. M. J. Gordon. "The Savings Investment and Valuation of a Corporation," *Review of Economics and Statistics*, Vol. 44, (February 1962), pp. 37-51.
7. Myron J. Gordon. *The Investment, Financing and Valuation of the Corporation*, (Homewood, Illinois: Richard D. Irwin, Inc.), 1962.
8. Myron J. Gordon. Testimony F.C.C., Docket No. 16258, F.C.C. Staff Exhibit No. 17, (1966).