Volume Title: Bank Stock Prices and the Bank Capital Problem

Volume Author/Editor: David Durand

Volume Publisher: NBER

Volume ISBN: 0-87014-368-9

Volume URL: http://www.nber.org/books/dura57-1

Publication Date: 1957

Chapter Title: Appendix A: Technical Notes on the Regression Analysis of Bank Stock Prices

Chapter Author: David Durand

Chapter URL: http://www.nber.org/chapters/c4629

Chapter pages in book: (p. 52 - 59)

# APPENDIX A

## *Technical Notes on the Regression Analysis of Bank Stock Prices*

### *The Regression Equation*

The regression function used in this study was the exponential

$$(1) \qquad P' = k\,B^b D^d E^e,$$

where $B$ is book value per share, $D$ is dividends per share, and $E$ is earnings per share. For the operation of fitting this equation, the logarithms of the variables have to be extracted, as is well known, and the equation that is actually fitted has the form

$$(2) \qquad \log P' = \log k + b \log B + d \log D + e \log E = \log P - \epsilon,$$

where $P$ represents observed price, and $\epsilon$ represents the deviations (assumed random) of log $P$ about the regression function.

Several modifications of (2) were employed in a vain search for additional variables that might affect bank stock prices. The most extensively used modification was

$$(3) \qquad \log P' = \log k + b \log B + d \log D + e \log E + c_1 \log C + c_2 \log A/C + c_3 (\log A/C)^2,$$

where $C$ is total capital and $A$ is total assets. This modification was designed to test the possible effects of two variables, size of bank and the ratio of assets to capital. The squared term in (3) was necessary to test the expected nonlinearity, even in the logarithms, of the relation between the $A/C$ ratio and price. It seemed altogether likely that the market might prefer, other things being equal, stocks of banks with moderate capitalization ratios – that is, neither too conservative nor too far extended. And should this interesting relation exist, clearly it could not be detected by means of a single linear term in the regression equation.

*Statistical Reliability and Confidence Limits*

Considerable statistical testing was conducted in order to appraise the reliability of the regression coefficients, or weights. As an example, three sets of coefficients are compared with their conventional 95 per cent confidence limits, as follows:

| | Book Value (*b*) | Dividends (*d*) | Earnings (*e*) |
|---|---|---|---|
| Group I, 1952 | | | |
| Regression coefficients | 0.40 | —0.02 | 0.64 |
| Conventional 95% limits | ±0.37 | ±0.24 | ±0.28 |
| Group II, 1952 | | | |
| Regression coefficients | 0.29 | 0.74 | 0.09 |
| Conventional 95% limits | ±0.21 | ±0.20 | ±0.19 |
| Group II, average | | | |
| Regression coefficients | 0.27 | 0.66 | 0.16 |
| Conventional 95% limits | ±0.07 | ±0.06 | ±0.06 |

But conventional 95 per cent limits do not entirely meet the requirements of this problem because the confidence level is actually less than 95 per cent – possibly very much less – whenever joint statements are made. And joint statements are unavoidable in stock price analysis. Not only is it desirable to set limits to individual coefficients, but, since we wish to compare the importance of factors, we should seek also to set limits to the differences between factors. Moreover, in view of the subsequent discussion of stock splits, it is desirable to set limits to the sum $b + d + e$.

To meet the need for joint statements, the tabulation below presents joint 95 per cent limits[1] wherein the confidence level is at least 95 per cent, regardless of the number of joint statements made.

| | Group I, 1952 | | Group II, 1952 | | Group II, Average | |
|---|---|---|---|---|---|---|
| | Coef. | Limit | Coef. | Limit | Coef. | Limit |
| $b$ | 0.40 | ±0.55 | 0.29 | ±0.31 | 0.27 | ±0.09 |
| $d$ | —0.02 | ±0.35 | 0.74 | ±0.29 | 0.66 | ±0.08 |
| $e$ | 0.64 | ±0.42 | 0.09 | ±0.27 | 0.16 | ±0.09 |
| $b - d$ | 0.42 | ±0.84 | —0.45 | ±0.53 | —0.39 | ±0.16 |
| $e - b$ | 0.24 | ±0.92 | —0.20 | ±0.53 | —0.11 | ±0.17 |
| $e - d$ | 0.66 | ±0.50 | —0.65 | ±0.43 | —0.50 | ±0.13 |
| $b + d + e$ | 1.02 | ±0.08 | 1.12 | ±0.10 | 1.09 | ±0.03 |

[1]For a discussion of joint limits with applications to the bank stock study, see the author's "Joint Confidence Regions for Multiple Regression Coefficients," *Journal of the American Statistical Association,* Vol. 49 (1954), pp. 130-46.

These confidence limits, whether conventional or joint, are too large for comfort. This is particularly true of limits for coefficients in individual years, but even here the limits are not so large as to destroy all value of the weights. For example, the limits for the difference $d - e$ are small enough to support the statement that earnings outweigh dividends for group I banks in 1952, while the reverse holds for group II banks in the same year. Difficulties will arise, however, when one wishes to establish the slopes of indifference lines, as in Chart 11, or to formulate some other fairly precise estimate of the degree to which one factor outweighs another in a specified group and year. The size of the confidence limits indicates that precise estimates of this sort are not possible.

The limits for group averages are smaller than those for individual years – owing to the larger number of degrees of freedom – and much more encouraging. But the possibility of heterogeneity from year to year casts some doubt on the reliability of these limits. In fact, the designer of statistical procedures in this field faces a serious dilemma: he can easily increase the size of his samples and the number of degrees of freedom, but in doing so he will run the risk of introducing heterogeneity within samples. At some point, the disadvantages of heterogeneity will undoubtedly outweigh any possible advantages from increased sample size.

The statistical problem of testing the apparent variations in the regression coefficients from group to group or from year to year has been discussed elsewhere.[2] Very briefly, the evidence of heterogeneity in the regression from group to group is fairly strong, and there is also evidence of heterogeneity in the residual variances. The evidence concerning year-to-year variation is less convincing, but hardly lacking; in addition there is a suggestion of serial correlation among the residuals, which is a deterrent to increasing experience by the expedient of pooling data for several years.

### *Computations*

Virtually all of the computations required for the regression analysis were performed on IBM punched-card equipment – including the conversion of raw data into logarithms, the summation of squares and products, and the solution of the simultaneous equations.

For the conversion of raw data into logarithms, it was desired to restrict the converted variables to three digits and at the same time

[2]David Durand, "Bank Stocks and the Analysis of Covariance," *Econometrica,* Vol. 23 (1955), pp. 30-45.

to reduce the rounding error as much as possible. Since preliminary investigation indicated that the range of variation after conversion would be approximately from –0.20 to 4.00, it was apparent that doubled three digit logarithms could be used effectively, for the doubling would not increase the number of digits, and it would reduce the rounding error by one-half. The conversion to doubled logarithms was performed with the aid of a specially prepared punched-card table of antilogarithms, containing such entries as

| Log | Antilog (unrounded) | Doubled-log conversion value |
|---|---|---|
| 0.9975 | 9.94260 | 2.00 |
| 1.0025 | 10.0577 | 2.01 |
| 1.0075 | 10.1741 | 2.02 |
| ... | ... | ... |

In the operation, the antilog cards would be sorted ahead of the detail cards, and the doubled-log values would then be gang punched from the antilog cards onto the subsequent detail cards. For example, all detail cards bearing values obeying the inequality

$$10.0577 \leqq X < 10.1741$$

would receive the rounded doubled-log value 2.01.

The summation of squares and products was performed on a 602 calculating punch by means of routine techniques. The solution of the simultaneous equations was performed on the 602 by a method similar to that described by Verzuh.[3] This method has the interesting feature that the forward solution and the back solution are performed simultaneously. One can solve a set of, say, six equations in six unknowns, while simultaneously obtaining solutions for five equations in the first five unknowns, four equations in the first four unknowns, and so on. Thus, the process of solving for the regression coefficients in (3) yielded also the regression coefficients of (2).

*Ratios in Regression Analysis*

The exponential type of regression equation, (1) or (2), has advantages in the study of stock prices because of its flexibility in handling ratios. Simple algebraic manipulations will transform (1) into such alternative forms as

[3]Frank M. Verzuh, "The Solution of Simultaneous Linear Equations with the Aid of the 602 Calculating Punch," *Mathematical Tables and Other Aids to Computation* (MTAC), Vol. 3, July 1949, pp. 453-62.

(4) $$P' = k\,B^{b+d+e}\,(D/B)^d\,(E/B)^e$$

(5) $$P'/B = k\,B^{b+d+e-1}\,(D/B)^d\,(E/B)^e.$$

To justify these transformations, it suffices to examine the sum of squares that is being minimized in the solution for the regression coefficients. For (5), for example, the sum of squares is

$$\Sigma\,[\log P - \log B - \log k - (b+d+e-1)\log B - d\,(\log D - \log B) - e\,(\log E - \log B)]^2$$

and this is identically equal to the sum

$$\Sigma\,(\log P - \log k - b\log B - d\log D - e\log E)^2,$$

which is maximized to obtain the coefficients of (1).

In equation (5), all but one of the variables are ratios. Moreover, the odd variable, $B$, drops out under the special condition $b + d + e = 1$, in which case (5) becomes

(6) $$P'/B = k\,(D/B)^d\,(E/B)^e.$$

The use of (6), which is simpler than (5), as the basic regression function would be equivalent, of course, to fitting (1) or (5) subject to the linear restriction $b + d + e = 1$. Since the sum of exponents may be expected to approximate unity in this particular application (see especially pages 58 and 59), the simpler equation (6) may provide a fair approximation. But in applications where the sum departs considerably from unity, (6) would not be suitable.

In correlation studies (6) may be unsuitable, even when $b + d + e = 1$, because the act of dividing each side of (1) by $B$ is apt to bias the correlation coefficients. This subject has been discussed in the literature for over fifty years, but mistakes are still made.[4] As shown above, the squared deviations of log $P$ about log $P'$ are identical with the squared deviations of log $P/B$ about log $P'/B$. But the deviations of log $P$ about its mean are ordinarily quite different from the deviations of log $P/B$ about its mean; therefore, the two correlation coefficients

[4]See, for example: Karl Pearson, "On a Form of Spurious Correlation Which May Arise When Indices Are Used in the Measurement of Organs," *Proceedings of the Royal Society,* Vol. 60 (1897), sections 8 and 9, p. 489; G. Udney Yule and M. G. Kendall, *An Introduction to the Theory of Statistics,* 13th edition revised, London, 1949, section 16.8; and Jerzy Neyman, *Lectures and Conferences on Mathematical Statistics and Probability,* 2nd edition, Graduate School, Department of Agriculture, 1952, pp. 143-54.

$$\sqrt{1-\frac{\Sigma(\log P-\log P')^2}{\Sigma(\log P-\overline{\log P})^2}} \text{ and } \sqrt{1-\frac{\Sigma(\log P/B-\log P'/B)^2}{\Sigma(\log P/B-\overline{\log P/B})^2}}$$

are not equal in general.

*Required Rates of Return*

It is a relatively easy matter to solve equation (5) for $E/B$ as follows:

$$(7) \qquad E/B=\left[\frac{B^{1-b-d-e}P'/B}{k\,(D/B)^d}\right]^{\frac{1}{e}}$$

In particular, when $P'/B = 1$, equation (7) estimates the return on capital necessary to support bank stocks at 100 per cent of book value when book value and the ratio of $D/E$ are given. Similar manipulations can be used to obtain solutions for $E/B$ when $D/E$ is given, for $D/B$ when $E/B$ is given, or for other relations that may be desired.

Some justification is required for the operations leading to (6) and similar equations. Under the classical assumptions of linear regression, equation (5) defines the expected value of log $P/B$, given specified values of $B$, $D/E$, and $E/B$. When this equation is solved for $E/B$ (or for another independent variable) the result does not define the expected value of log $E/B$; instead it defines the value of $E/B$ required to yield a specified expected value for log $P/B$, given $B$ and $D/E$. Although the expected value for log $E/B$ could be obtained — under some assumptions[5] — by regressing this variable on the others, that value would not meet the requirements of the bank stock problem. We do not wish to estimate the rate of return when bank stocks sell at 100 per cent of book value, but rather to ascertain the rate of return that seems most likely to cause bank stocks to sell at 100 per cent of book value; and this is obtained from the solution of (7) under the classical assumptions of linear regression.[6]

[5]In many modern discussions of regression, it is assumed that the dependent variable has a probability distribution, but that the independent variables are fixed known parameters — possibly subject to variation in a controlled experiment. Under these assumptions, it is of course nonsense to speak of a most probable value for one of the independent variables or to regress one of these variables on any others.

[6]For further discussion see "The Interpretation of Certain Regression Methods and Their Use in Biological and Industrial Research," by Churchill Eisenhart, *Annals of Mathematical Statistics,* Vol. X (1939), pp. 162-86.

*High Priced versus Low Priced Stocks: Split-ups*

As is well known, the exponential type of regression function (1) is homogeneous and the sum of its exponents has implications for evaluating high priced versus low priced stocks, and hence for estimating the probable effects of stock splits. The same type of function is, of course, often used as a production function. In that context the sum of exponents indicates the degree of returns to scale – a sum equal to one implying constant returns, a sum greater than one implying increasing returns, and a sum less than one implying decreasing returns. In the context of stock price analysis, the sum of exponents indicates the degree to which the ratios $P/B$, $P/D$, and $P/E$ are affected by the actual level of $B$, $D$, and $E$. Consider two stocks for which the amounts of $B$, $D$, and $E$ are strictly proportional, for example:

| | | |
|---|---|---|
| *B* | $100.00 | $25.00 |
| *D* | 6.00 | 1.50 |
| *E* | 10.00 | 2.50 |

On the basis of these figures alone, one share of the first stock represents the same investment value in terms of $B$, $D$, and $E$ as four shares of the second, but this does not guarantee that the market price of the first will be four times the market price of the second. In fact, a popular market belief holds that high priced stocks are unsuited for small portfolios and therefore restricted in market appeal; hence the first stock above might not be able to command fully four times the market price of the second.

The condition of proportionality for stock prices is analogous to that of constant returns to scale and is identified by the unit sum,

$$b + d + e = 1,$$

for the exponents of (1). Under this condition, high priced stocks should sell for as much in relation to book value, dividends, and earnings as low priced stocks, and stock splits should exert no upward pressure through broadening the market. On the other hand, the condition for high priced stocks to sell proportionally below low priced stocks, which is analogous to decreasing returns to scale, is identified by a sum of exponents less than one. Finally, the condition for high priced stocks to sell proportionally higher is identified by a sum greater than one.

A glance at Table 2 indicates that the sum of exponents (weights) varies all the way from 0.90 to 1.14, but on the whole the sum tends to exceed 1.00 rather than to be less. In fact, 38 of the 48 unaveraged sums exceed 1.00, two equal it, and eight are less. This evidence clearly fails to support the hypothesis that high priced stocks sell for less in proportion to book value, dividends, and earnings than do low priced stocks. On the other hand, it does not necessarily support the contrary view, for other explanations are easily found. For example, the market may possibly regard the higher priced stocks in some of the six groups as possessing better quality than the lower priced stocks; and in this event, the sum of the weights could exceed 1.00 without any implications concerning proportionality or the effects of stock splits.