Measuring the Size Factor In Equity Returns

Does company size effect investment returns? Most people would probably agree that it does, but how do you measure its effect? In this paper, the author proposes a new method for the direct calculation of the effect of size on equity returns.

Robert Fernholz, Ph.D.
is Chief Investment Officer of Princeton-based Intech, a subsidiary of The Prudential Insurance Company of America, that uses mathematical strategies to manage institutional equity portfolios. Earlier in his career, he served on the faculties of Princeton University and the City University of New York. He has published articles on mathematics, statistics, and finance, in academic and professional journals. He received his A.B. in Mathematics from Princeton University and his Ph.D. in Mathematics from Columbia University.

OBSTACLES TO MEASURING THE SIZE EFFECT

Company size is known to be an important factor affecting equity returns, but statistical estimation of the impact that this factor has on returns is complicated by its unstable and nonlinear nature. In this paper, a new method is proposed for the direct calculation of the effect of size on equity returns. The method is based on the analysis of changes in the distribution of capital in the market. And, since the method does not involve statistical estimation, instability and nonlinearity do not interfere with it. Application of the proposed method indicates that the exposure of some managers' returns to the size factor may be significantly greater than is implied by traditional estimates.

It has been understood for some time that the size of companies is an important factor that affects stock returns, and hence portfolio performance (see Ross (1976), Fama and French (1993,1995,1996)). In fact, Fernholz and Garvy (1999) presented evidence that the size factor explained over half of the annual variation in median manager relative returns versus the S&P 500 over the period from 1971 to 1998.

Approaches

Traditionally the effects of structural factors on equity returns have been estimated by statistical methods related to least-squares regression (see Sharpe (1988)), but there are at least two significant problems with that approach. First, there seldom exists a natural, well-defined variable to represent a factor, and, second, regression analysis is frequently complicated by instability, nonlinearity of response, correlation among multiple factors, and sensitivity to outliers in the data. Although regression methods have been used traditionally, there is no *a priori* reason why these techniques must be used. Regression minimizes the mean squared residual; and there is no reason to believe that this is always an appropriate objective when measuring the effect that a factor has on portfolio returns. For example, regression analysis would probably not be used to measure the effect of a particular stock on the return of a portfolio, since in this case direct measurement is simpler and more accurate, and the objective of minimizing the effect of all the other stocks in the portfolio makes no sense.

In order to use regression to measure the effect of size on portfolio returns, an appropriate explanatory variable must be found. However, there is no natural variable to represent the size factor in regression: it could be represented, for example, by the relative return of the largest 100 stocks in the S&P 500 Index versus the remaining 400, or by the return on the Russell 1000 Index versus the Russell 2000 Index, or by the change in market diversity (as in Fernholz and Garvy (1999)) and so forth. Moreover, no single variable can accurately represent a nonlinear, multidimensional relationship. In fact, BARRA (1997) announced that they would use two variables to represent the size factor in order to attempt to capture the nonlinearity of its effect. The arbitrary nature of the choice of the variable used to represent the size factor in the regression-based methods casts further doubt on the efficacy of these traditional techniques.

PROPOSED METHOD

Here we propose a direct and natural method to calculate the effect of size based on the analysis of changes in the distribution of capital in the stock market. Changes in the distribution of capital are caused by the ebb and flow of capital between the larger and smaller stocks. If capital flows into the larger stocks, then the capital distribution becomes more concentrated; and if capital ebbs back into the smaller stocks, the distribution becomes more diverse. We show that the effect on portfolio returns due to changes in the capital distribution can be measured directly and accurately. This provides a precise method of calculating the effect of size that is free from the multifarious difficulties of the traditional regression-based techniques.

As an application of our method, we analyze the performance of a simulated "active core" manager relative to the benchmark S&P 500 Index over the ten-year period from 1989 to 1998. We show that, compared to our method, traditional regression techniques significantly underestimate the effect of size on the returns of this manager.

The Distributional Component of the Relative Return

Suppose that an equity market contains *n* stocks. Suppose that the capitalization weights of the stocks are $w_1 \geq w_2 \geq \ldots \geq w_n$, arranged in descending order. The ordered set of these weights, $w_1, w_2, \ldots, w_n$, is called the *capital distribution* of the market, and if we plot the weights on a chart, we generate a decreasing curve called the *capital distribution curve*. For each point on the capital distribution curve, the vertical coordinate represents the capitalization weight of a particular stock and the horizontal coordinate represents the rank of that stock. The area under the curve will be equal to 1, the sum of all the weights. The capital distribution curve is steeper when capital is more concentrated into the larger stocks, and is flatter when capital is more evenly distributed over the market.

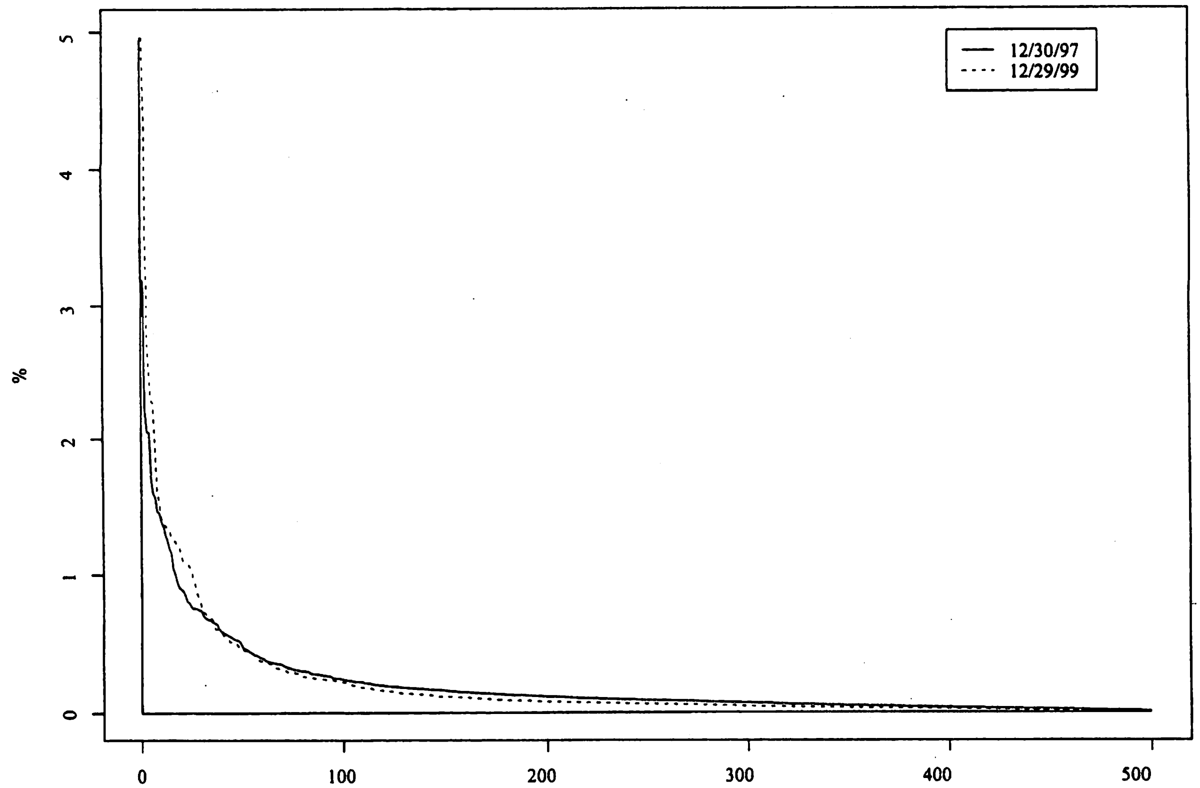
A pair of capital distribution curves for the S&P 500 Index is shown in Figure 1 (*see page* 13). The solid line is the curve for December 30, 1997; the broken line is the curve for December 29, 1999. The points on the curves in Figure 1 represent the weights of the stocks in the S&P 500 Index expressed in percent. From the two curves we can see that there was more concentration of capital in the larger stocks at the end of 1999 than at the end of 1997. This means that there was a flow of capital from the smaller stocks to the larger stocks during 1998 and 1999. For a given portfolio, the return relative to the S&P 500 Index would have been affected by this shift in capital. If the portfolio held a higher proportion of the larger stocks than the S&P 500, its relative return probably would have benefited from the shift, and if it held a lower proportion, its relative return probably would have suffered.

When large stocks outperform small stocks, the capital distribution curve steepens; when small stocks outperform large stocks, the capital distribution curve flattens out. Differences in the relative returns of large and small stocks occur if and only if there are changes in the capital distribution. Consider the following example.

Suppose we have a two-stock market in which Stock *X* has twice the capitalization of Stock *Y*, so the capital distribution of the market is (2/3, 1/3). Suppose now that over a given period *X* drops by 50% and *Y* rises by 100%, so *Y* ends up with twice the capitalization of *X*, but the capital distribution of the market, which is always in descending order, remains unchanged at (2/3, 1/3). It is tempting to conclude that over this period smaller stocks have outperformed larger stocks since *X*, which started the period with twice the capitalization of *Y*, went down, and *Y* went up. But this facile analysis is flawed. Indeed, the initially smaller stock *Y* outperformed the initially larger stock *X* until they both were the same size, but after that the now larger stock *Y* outperformed the now smaller stock *X*. Hence, on average over the period, the performance of small stocks and large stocks in this market was exactly the same. If small stocks had outperformed large stocks over the period, the capital distribution at the end of the period would have to be closer to equal weights than it was at the beginning of the period.

The changes in the capital distribution of the market over the period being considered contain exactly the information needed in order to calculate the effect of size

Figure 1
Capital Distribution Curves for the S&P 500 Index.
December 30, 1997 (solid line) and December 29, 1999 (broken line).

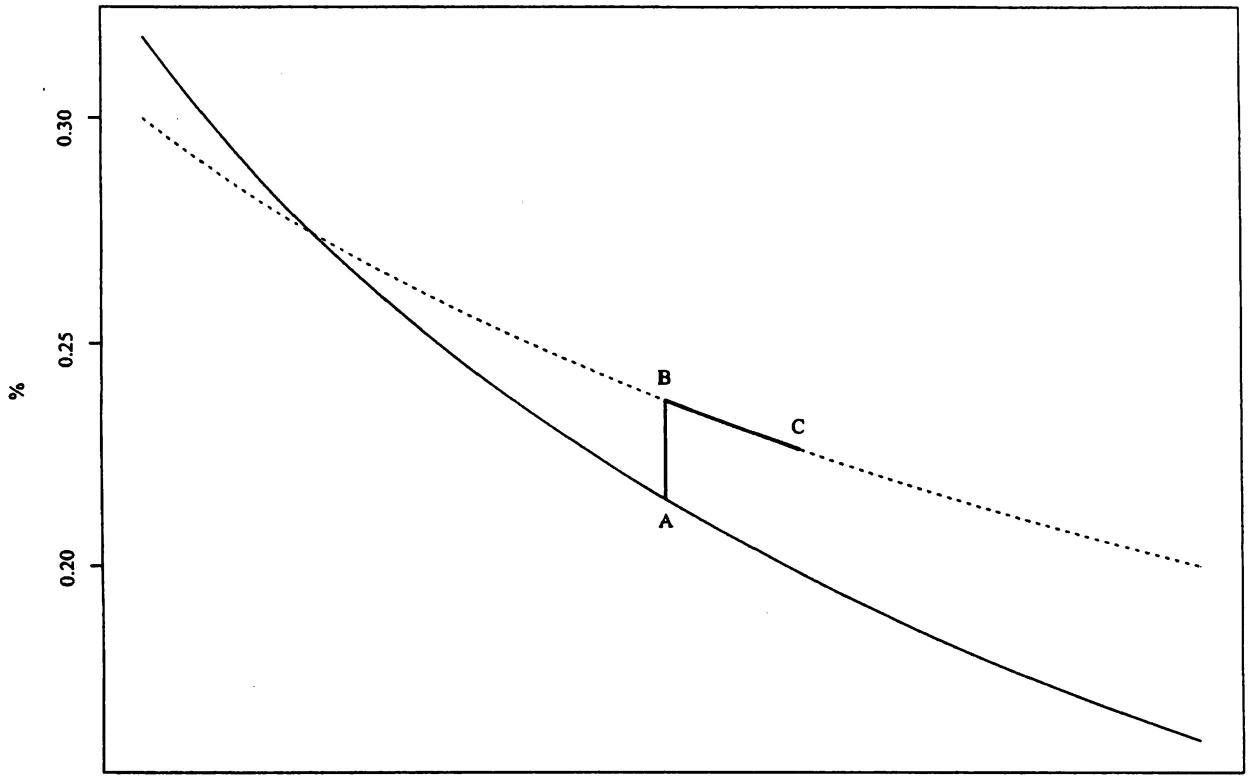


on the relative return of stocks and portfolios. If we were to use a regression model with the return of the larger half of the stocks relative to the smaller half as the explanatory variable, this method could provide an estimate of the effect of size over the period. But this estimate will not be sensitive to differences among the returns of the stocks within either one of the two halves. Perhaps if we used the relative returns of the five quintiles in a multivariate model, our estimate might improve. Or better yet, we could use deciles. And so forth. However, in order to capture the *entire* effect of size on returns, the information provided by the changes in the capital distribution is both necessary and sufficient. For this reason, the effect of size on portfolio returns can be calculated directly from the changes in the capital distribution without the need for statistical estimation or approximation.

In the Appendix (*see page* 17), we shall present a detailed mathematical description of the method we propose; here we introduce the basic ideas. Figure 2 (*see page* 14) represents a portion of two capital distribution curves, the solid line for the curve at the beginning of a given period of time, and the broken line for the curve at the end of the period. As usual, the vertical axis in Figure 2 represents the capitalization weights of the stocks; and the horizontal axis represents their ranks. If the capitalization weight of a stock increases over the period, then its capitalization has increased relative to the market's capitalization. Therefore its return has been greater than the market's return.

In Figure 2, suppose that a particular stock starts the period at Point A, and ends the period at Point C. The vertical distance from A to C measures the change in

Figure 2
The Distributional Component of the Relative Return Is that Component Caused by a Shift in the Capital Distribution Curve.



the capitalization weight of the stock, and this change represents the relative return of the stock versus the market. Since *C* is higher on the vertical axis than *A*, the weight of the stock increases over the period. Thus, the stock outperforms the market. However, *C* is to the right of *A*, so the stock fell to a lower rank over the period.

<u>The Distributional Component</u>

If the rank of the stock had not changed over the period, then the stock would have ended up at Point *B*. The implied relative return corresponding to the move from *A* to *B* is defined to be the distributional component of the relative return of the stock over the period. Hence, the distributional component is precisely that part of the relative return of the stock that is due to the shift in the capital distribution curve. The distributional component is a natural representation of the effect of size on stock returns, and can be calculated directly without the use of statistical estimation.

After the distributional component has been removed from the relative return of the stock, a residual component remains. In Figure 2, the residual component is the implied relative return corresponding to the move from *B* to *C*, so the residual component is negative in this case. Clearly, the residual component is positive or negative according to whether the stock rises or falls in rank in the market.

For a portfolio of stocks, it is natural to define the distributional component to be the implied relative portfolio return generated by the distributional component of the relative return of each of the individual stocks. This is precisely the part of the relative portfolio return that is caused by the change in the capital distribution. As with individual stocks, a residual component remains after the distributional component has been subtracted from the relative return of the portfolio. This residual component will be positive if, on average, the stocks in the portfolio rise in rank, and will be negative if they fall.

The method we have outlined here for direct calculation of the distributional component will sometimes produce results that are significantly different from those of traditional methods. In the next section we shall apply our method to a particular portfolio, and compare the results to those generated by the traditional regression techniques.

ANALYSIS OF A SIMULATED MANAGER

In this section we shall calculate the distributional component of the relative return of a simulated "active core" equity manager. We shall compare the distributional component calculated by our method with estimates of the effect of size produced by traditional regression techniques. Our results indicate that for this manager the distributional component is much greater than implied by the traditional techniques.

The goal of the active core management style is to generate annual return about one or two percent higher than a benchmark large-stock index such as the S&P 500, while at the same time maintaining control over the standard deviation of the return relative to the benchmark. Active core portfolios can be quite large, sometimes holding several hundred stocks selected from the benchmark index, and this is the type of portfolio our simulated manager holds. The relative return of active managers frequently has a significant exposure to the size factor (see Fernholz and Garvy (1999)), and our simulated manager shares this characteristic.

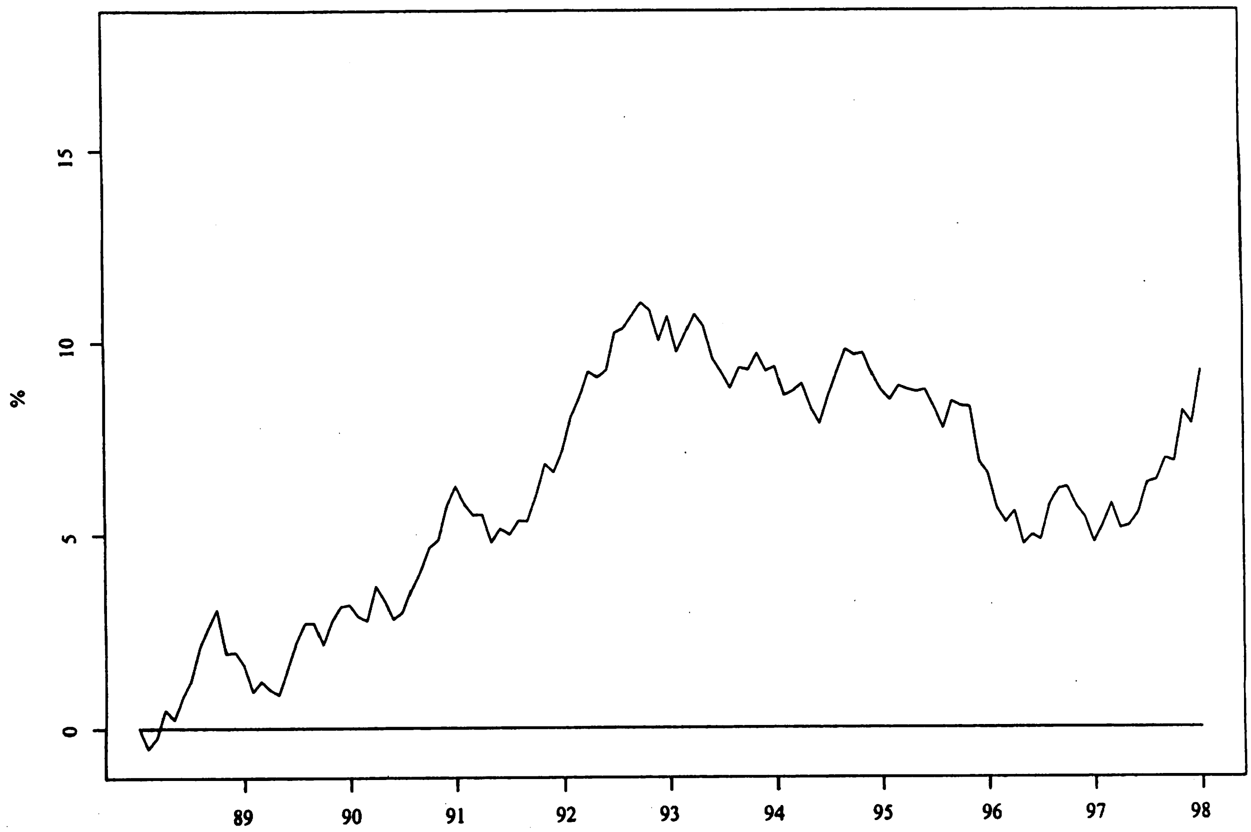
The cumulative monthly relative logarithmic return (log-return) of the simulated manager versus the S&P 500 Index over the period from January 1, 1989 to December 31, 1998 is presented in Figure 3 (*see page 16*). As we can see, the manager outperformed the benchmark by about 2% a year for the five years from 1989 to 1993, went into a slump for four years, and then came back in 1998. Let us see how size affected this performance over the period.

The distributional component of the manager's return is presented in Figure 4 (*see page* 17). Here we see that from 1989 to 1993 there was very little cumulative effect of the distributional component, however from 1994 on it declined at about 2% a year. In Figure 5 (*see page* 18) we see that the residual component of the manager's return was about 2% a year for the first five years, and then flattened out from 1994 to 1997 for some reason. However, in 1998 the residual component came back strongly. In any case, from 1994 to 1997 the flat residual component combined with a significantly negative distributional component gave the manager four years of poor performance relative to the benchmark.

Now let us see how the traditional regression analysis estimates the effect of size over the period. Figure 6 (*see page* 19) presents two estimates of the cumulative effect of size in the manager's return. The solid line is the estimate when the explanatory variable in the regression is the relative log-return of the largest 25 stocks in the S&P 500 Index versus the Index itself. The broken line is the corresponding estimate using the largest 100 stocks versus the S&P 500. Both the curves in Figure 6 are of about the same magnitude. From the look of these charts, the estimates in Figure 6 have roughly the same shape as that in Figure 4, but the magnitude is only about one-fourth as great (the scale of Figure 6 is the same as that of Figures 3, 4, and 5). Hence we see that regression provides a much smaller estimate of the effect of size on the relative return of this manager than was indicated by our direct calculation.

Let us consider one more estimate of the effect of size using regression analysis: let us use our calculated values of the distributional component as the explanatory variable in the regression. In this case we find that the regression coefficient is approximately 0.51 and that this explains about 16% of the monthly variation of the relative log-return. Hence, even in this case where the explanatory variable is our calculated distributional component, least-squares regression estimates the effect of size to be about half the value we calculated directly.

Figure 3
10-Year Cumulative Relative Log-Return of a Simulated Core Manager.

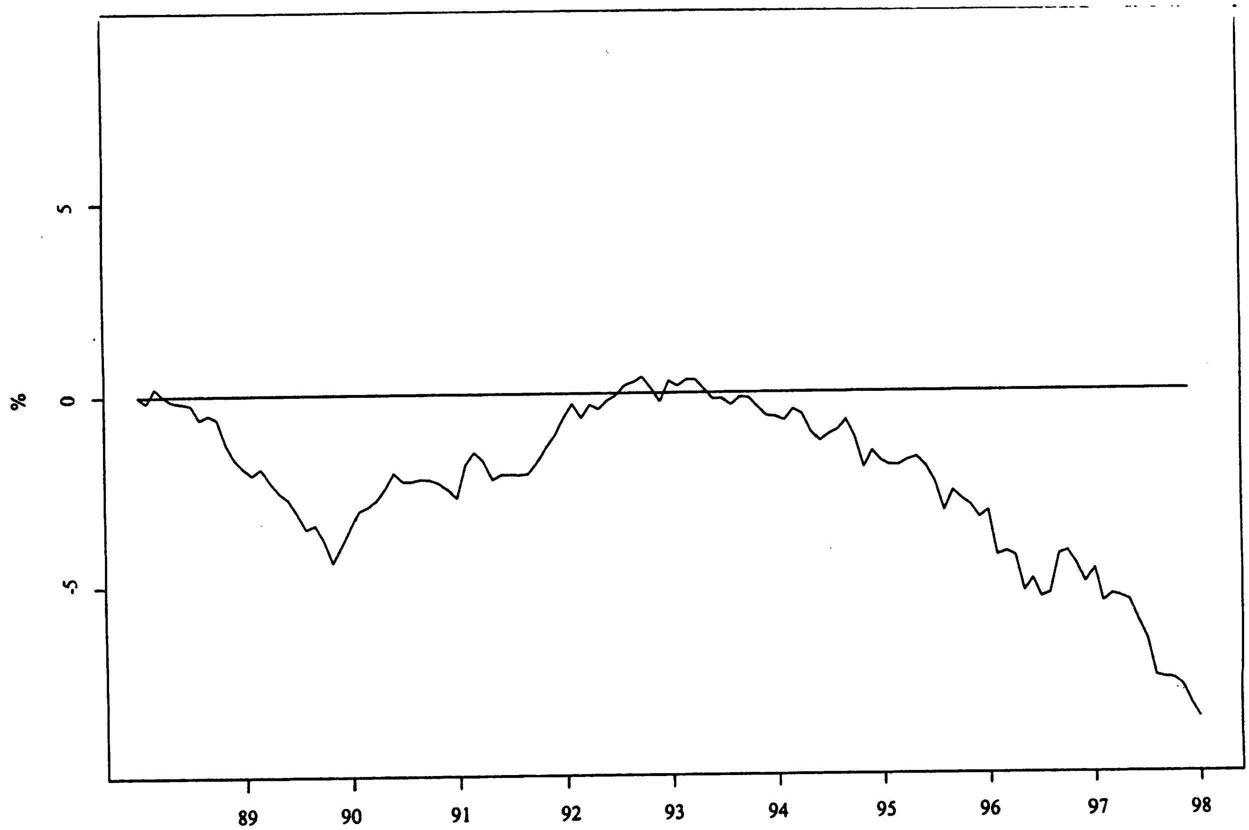


Our calculation differs so much from the regression estimates because the two techniques are designed to achieve different results. Our calculation directly measures the effect of shifts in the distribution of capital in the market, whereas regression minimizes the sum of the squares of the residuals. The residuals contain the combined effects of all other factors that may impact portfolio performance. It is difficult to see any rationale for an estimate that minimizes the combined effects of all the other factors.

CONCLUSION

We have presented a direct method for measuring the effect of company size on the relative return of an equity portfolio. Our measurement depends on calculating the effect on the relative portfolio return caused by changes in the capital distribution of the market. Our direct method of measurement is free of the problems associated with the traditional regression-based estimates of the effect of size. We have shown that our direct method sometimes can produce results that differ significantly from the traditional estimates.

Figure 4
Distributional Component of the Log-Return in Figure 3.



APPENDIX

Calculation of the Distributional Component

In this appendix we present a mathematical description of the calculation of the distributional component of the relative return of a stock or portfolio. We shall consider a period from time T_0 to time $T_1 > T_0$, and suppose that during this period the number of stocks in the market is fixed: there are no additions or deletions, and the companies neither merge nor split up. Note that since any time period can be broken up into shorter periods that satisfy this assumption, our results can be used for longer periods as well.

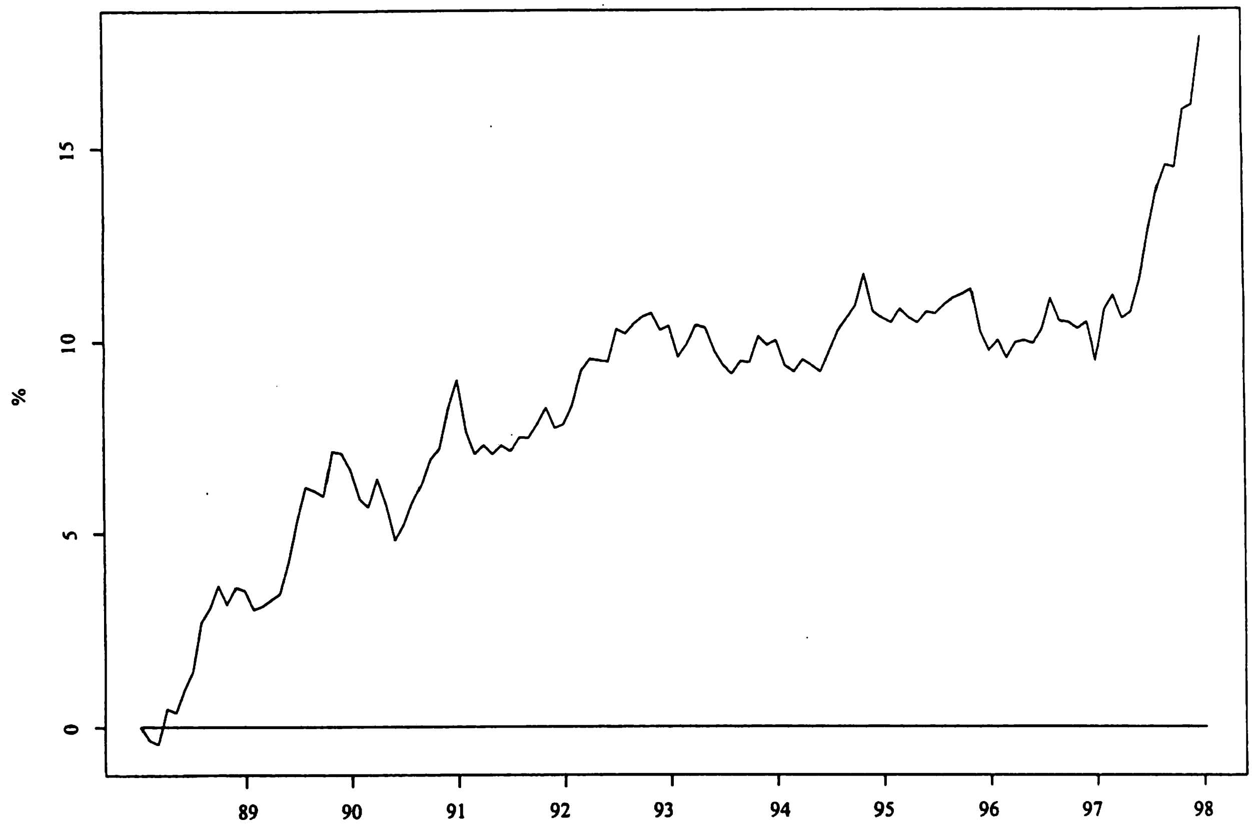
Suppose that the market contains n stocks and let

$$X_1 \geq X_2 \geq \ldots \geq X_n \tag{1}$$

represent the capitalization's of the stocks at time T_0 in descending order. In this case the total capitalization of the market at time T_0 is

$$M = X_1 + X_2 + \ldots + X_n.$$

Figure 5
Residual Component of the Log-Return in Figure 3.



Let X'_i represent the capitalization of the i-th stock at time T_1, so the total capitalization of the market at time T_1 will be

$$M' = X'_1 + X'_2 + \ldots + X'_n.$$

Let us assume for the moment that no dividends or other distributions are paid over the period, in which case the log-return of the i-th stock will be

$$\log(X'_i / X_i),$$

and the log-return of the market will be

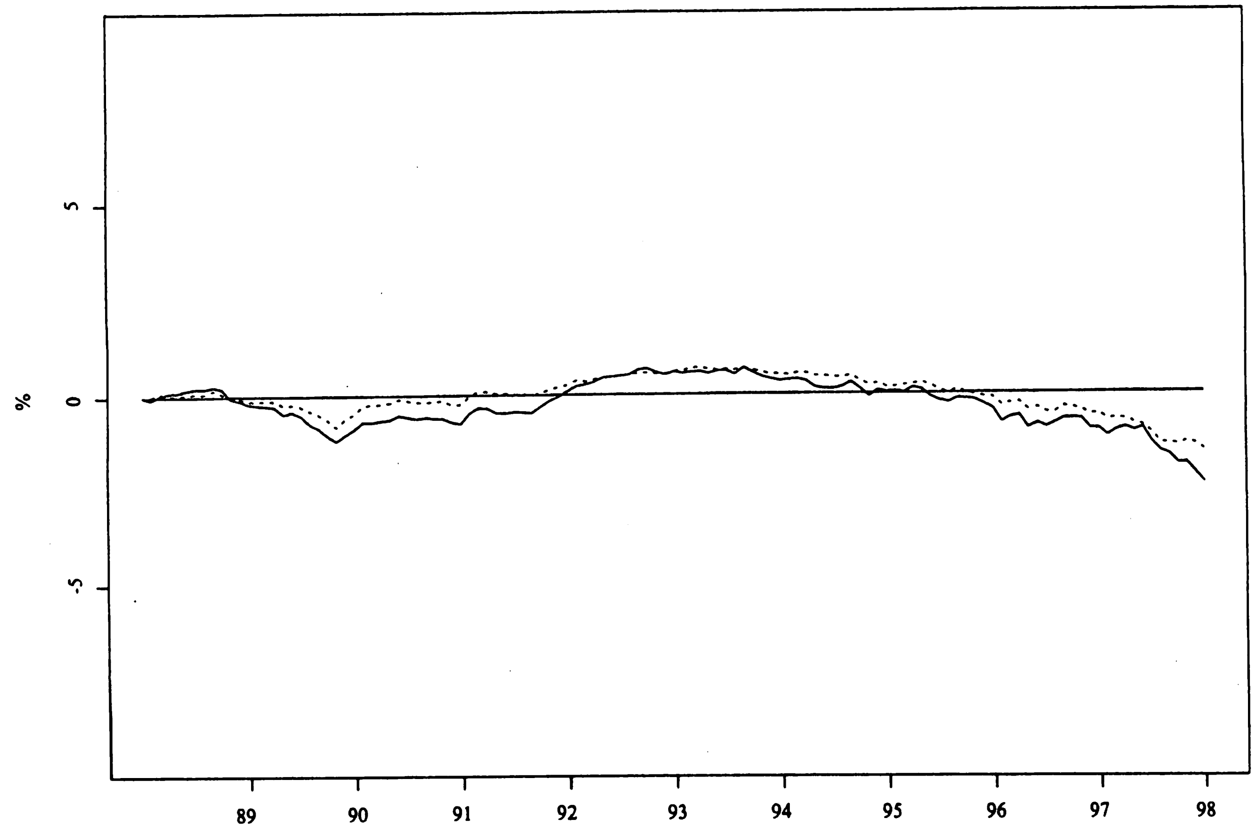
$$\log(M'/M).$$

Let us consider now the *capitalization weights*

$$w_i = X_i/M,$$

for $i=1, \ldots, n$, at time T_0, and

Figure 6
Size Component Estimated by Regression of Log-Return in Figure 3. Solid line: vs. Relative Log-Return of Largest 25 S&P 500 Stocks. Broken line: vs. Relative Log-Return of Largest 100 S&P 500 Stocks.



$$w'_i = X'_i / M',$$

for $i=1, \ldots, n$, at time T_1. The log-return of the i-th stock relative to the market is

$$\begin{aligned}\log(X'_i/X_i) - \log(M'/M) &= \log(X'_i/M') - \log(X_i/M) \\ &= \log(w'_i) - \log(w_i),\end{aligned} \qquad (2)$$

so the relative log-return of the stocks can be represented in terms of the change in their capitalization weights.

It follows from Equation (1) (*see page* 17) that

$$w_1 \geq w_2 \geq \ldots \geq w_n.$$

Hence, as in Section 2, the ordered set of weights $(w_1, w_2, \ldots, w_n)$ represents the *capital distribution* at time T_0. It is unlikely that the weights w'_i at time T_1 are in descending order, but we can rearrange them with a permutation p such that,

$$w'_{p(1)} \geq w'_{p(2)} \geq \ldots \geq w'_{p(n)}.$$

In this case, the capital distribution at time T_1 will be $w'_{p(1)}, w'_{p(2)}, \ldots, w'_{p(n)}$.

The *distributional component* of the log-return of the i-th stock is defined to be

$$\log(w'_{p(i)}) - \log(w_i). \tag{3}$$

By this definition, the distributional component corresponds to the transition from Point A to Point B in Figure 2 (*see page* 14). The distributional component is precisely the contribution to the relative log-return of the stock due to the shift in the capital distribution. If there is no shift in the capital distribution, then there is no distributional component in any of the stocks' log-returns. If a stock maintains its rank in the capital distribution, then the distributional component will be equal to the relative log-return of the stock.

Now suppose we have a portfolio of stocks; we can assume without loss of generality that each stock has a single share outstanding, and that the portfolio holds fractional shares. In this case, if the portfolio holds s_i (fractional) shares of the i-th stock, then the total value of the portfolio P at time T_0 is

$$P = s_1 X_1 + \ldots + s_n X_n.$$

The number of shares of individual stocks in the portfolio can be negative, representing short sales, but the portfolio value must always be positive.

Let us assume that the portfolio makes no trades over the period. (This causes no loss of generality since any time period can be partitioned into shorter periods in which this assumption is valid.) In this case, the portfolio value at time T_1 will be

$$P' = s_1 X'_1 + \ldots + s_n X'_n.$$

Hence, the log-return of the portfolio will be

$$\log(P'/P),$$

and the relative log-return will be

$$\begin{aligned}\log(P'/P) - \log(M'/M) &= \log(P'/M') - \log(P/M)\\ &= \log(s_1 w'_1 + \ldots + s_n w'_n) - \log(s_1 w_1 + \ldots + s_n w_n).\end{aligned}$$

As in Equation (2) (*see page* 19), the relative log-return of the portfolio can be represented in terms of the change in the capitalization weights of the stocks.

As in Equation (3) (*above*) we define the distributional component of the relative log-return of the portfolio to be

$$\log(s_1 w'_{p(1)} + \ldots + s_n w'_{p(n)}) - \log(s_1 w_1 + \ldots + s_n w_n). \tag{4}$$

This distributional component measures the contribution to the relative log-return of the portfolio due to the shift in the capital distribution.

This methodology is valid for a broad market such as the market of all exchange-traded U.S. stocks. However, our model is not appropriate if the market is replaced by an index such as the S&P 500 Index, in which smaller stocks are systematically dropped and replaced by larger ones. In this case we must consider the S&P 500 itself to be a portfolio of stocks within the broad market. To calculate the contribution of the distributional component for a portfolio relative to the S&P 500, we first calculate the distributional component of the portfolio relative to the market and then subtract from this the distributional component of the S&P 500 relative to the market. This procedure is related to the correction for "leakage" discussed in Fernholz, Garvy, and Hannon (1998) and Fernholz (1999).

In the event that dividends are paid over the period, the log-returns of the stocks will have to be modified to include the dividends. Since any shift in the capital distribution depends only on capital gains and losses, the calculation of the distributional component in Equation (3) (*see page* 20) and Equation (4) (*see page* 20) will remain unchanged.

REFERENCES

BARRA, "BARRA Releases New Version of U.S. Equity Model," *Press Release*, May 1997, Available at http://www.barra.com/PressReleases/5597-pr.asp.

Fama, E. and K. French, "Common Risk Factors in the Returns on Stocks and Bonds," *Journal of Financial Economics*, 1993, Vol. 33, pp. 3-56.

Fama, E. and K. French, "Size and Book-To-Market Factors in Earnings and Return," *Journal of Finance*, 1995, Vol. 50, pp. 131-155.

Fama, E. and K. French, "Multifactor Explanations of Asset Pricing Anomalies," *Journal of Finance*, 1996, Vol. 51, pp. 55-87.

Fernholz, R. "Equity Portfolios Generated by Functions of Ranked Market Weights," To appear in *Finance and Stochastics*, 2001.

Fernholz, R. and R. Garvy, "Diversity Changes Affect Relative Performance," *Pensions & Investments*, May 1999, Vol. 112.

Fernholz, R., R. Garvy, and J. Hannon, "Diversity-Weighted Indexing," *Journal of Portfolio Management*, Winter 1998, Vol. 24, No. 2, pp. 74-82.

Ross, S. "The Arbitrage Theory of Capital Asset Pricing," *Journal of Economic Theory*, 1976, Vol. 13, pp. 341-360.

Sharpe, W. "Determining a Fund's Effective Asset Mix," *Investment Management Review*, November/December 1988, pp. 59-69.