The size factor in equity returns

Robert Fernholz

INTECH
One Palmer Square
Princeton, NJ 08542

February 11, 2000

Abstract

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 size component of 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 conventional estimates.

1 Introduction

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 size explained over half of the annual variation in median manager relative returns versus the S&P 500 over the period from 1971 to 1998.

Traditionally the structural components of equity returns have been estimated by statistical methods related to regression analysis (see Sharpe (1988)). However, it is well known that this type of analysis can be complicated by instability, nonlinearity of response, and correlation among multiple factors. Here we propose a direct method to calculate the size component in equity returns. Unlike regression analysis, our calculation is direct and is not sensitive to instability or nonlinearity.

Although the size component of portfolio return is traditionally measured by regression techniques, there is no *a priori* reason why these techniques must be used. Regression minimizes the mean squared residual, and this may not always be appropriate when measuring return components. For example, since stock returns can be measured directly, regression analysis is probably inappropriate for measuring the component of portfolio return due to a particular stock. To use regression, appropriate explanatory variables 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 of the Russell 1000 Index versus the Russell 2000 Index, or by the change in market diversity, as in Fernholz and Garvy (1999). Moreover, no single variable can accurately represent a nonlinear, multidimensional relationship. In fact, BARRA (1997) announced that they would use two variables to represent size in order to attempt to capture the nonlinearity of its effect. The arbitrary nature of the size variable used in the regression-based methods casts doubt on the efficacy of the traditional techniques.

We propose a direct method to calculate the size component 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. The size component measures the effect on return caused by this ebb and flow of capital between the larger and smaller stocks.

As an application of our method, we analyze the relative performance versus the S&P 500 Index of a simulated "active core" manager over the ten year period from 1989 to 1998. We show that, relative to our method, conventional regression techniques significantly underestimate the size component for this manager.

2 The size component of the relative return

Suppose that an equity market or benchmark index contains n stocks. Suppose that the capitalization weights of the stocks are $w_1 \geq w_2 \geq \cdots \geq w_n$, arranged in descending order. If we plot these weights on a chart, we generate a decreasing curve called the *capital distribution curve* of the market or index. 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.

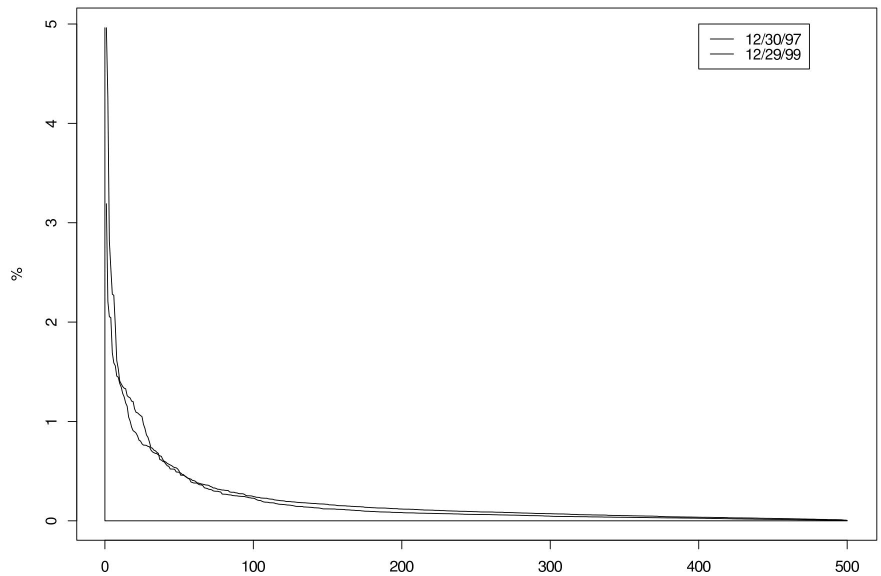

Figure 1: Capital distribution curves for the S&P 500 Index.
December 30, 1997 (solid line) and December 29, 1999 (broken line).

A pair of capital distribution curves for the S&P 500 Index are shown in Figure 1. The solid line is the curve for December 30, 1997 and 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 benefitted from the shift, and if it held a lower proportion, its relative return probably would have suffered. The size component of the relative return is that component caused by shifts in the capital distribution curve.

In the Appendix we present a detailed description of our method for calculating the size component for a portfolio. Here we shall introduce the basic ideas behind the method. Figure 2 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 proportionally more than the market's capitalization, and therefore it has outperformed the market.

Suppose 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 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, and 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.

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 *size component* of the relative return of the stock over the period. Hence, the size component is

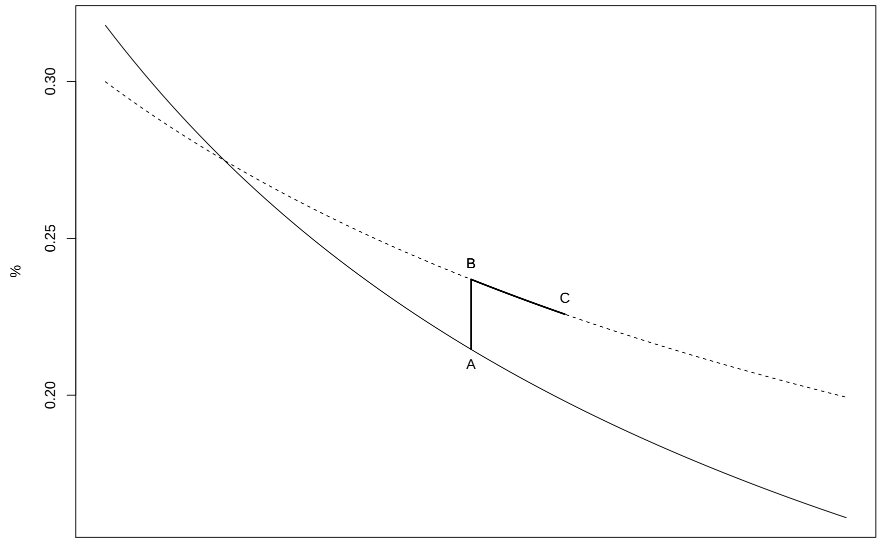

Figure 2: The size component of the relative return is that component caused by a shift in the capital distribution curve.

precisely that part of the relative return of the stock that is due to the shift in the capital distribution curve. Since the vertical distance from A to B is greater than the vertical distance from A to C, the size component is greater than the relative return in this case. After the size component has been removed from the relative return of the stock, a residual component remains. The residual component is the implied relative return corresponding to the move from B to C, and since C is lower on the vertical axis than B, this residual component is negative.

For a portfolio of stocks, the size component is equal to the implied relative portfolio return generated by the size 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 shifts in the capital distribution curve. After the size component has been removed from the relative return of the portfolio, a residual component remains. This residual component will be positive if, on average, the stocks in the portfolio rise in rank, in which case the portfolio is likely to outperform the market. Selecting stocks that rise in rank can be considered evidence of "alpha generation" by the portfolio manager, and the efficacy of this stock selection is measured by the residual component of the relative return.

The method we have outlined here for calculating the size component will sometimes produce results that are significantly different from those of conventional methods, which usually depend on some form of regression. In the next section we shall apply our method to a particular portfolio, and compare the results to those generated by conventional regression techniques.

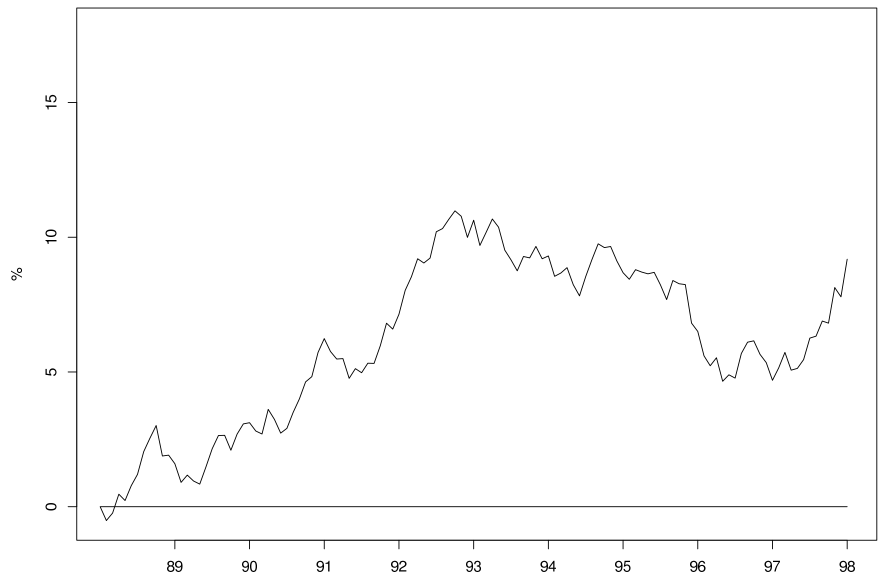

Figure 3: 10 year cumulative relative log-return of a simulated active core manager.

3 Analysis of a simulated manager

In this section we shall calculate the size component of the relative return of a simulated "active core" equity manager. We shall compare the size component calculated by our method with the size component as estimated by conventional regression techniques. Our results indicate that for this manager the size component is much greater than implied by the conventional 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 size component (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. 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 size component of the manager's return is presented in Figure 4. Here we see that from 1989 to 1993 there was very little cumulative effect of the size component, however from 1994 on it declined at about 2% a year. Hence, if the manager's alpha generation was about 2% a year, it would have been neutralized by the size component over the last five years of the simulation.

As we see from Figure 5 the residual alpha-generation 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 alpha-generation came back strongly. In any case, from 1994 to 1997 the flat alpha generation combined with a significantly negative size component gave the manager four years

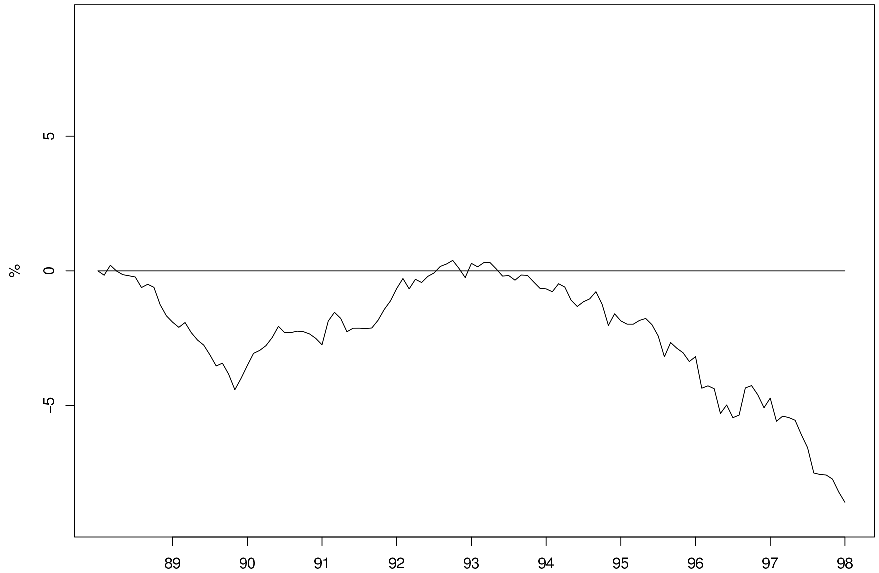

Figure 4: Size component of log-return in Figure 3.

of poor performance relative to the benchmark.

Now let us see how the size component looked using conventional regression analysis. Figure 6 presents two estimates of the cumulative size component 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 size component 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 size effect for this manager than does our method of direct calculation.

Let us consider one more estimate of the size component using regression analysis: let us use our calculated values of the size component as the explanatory variable in the regression. In this case we find that the regression coefficient is approximately .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 size component, least-squares regression estimates the size component at about half of its calculated value.

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. As we have seen, the residuals measure the effect of the changes in rank of the stocks in the portfolio, and hence relate to the efficacy of the manager's stock selection. It is difficult to see any rationale for an estimate that minimizes the manager's stock selection ability. Minimization of the residuals may be appropriate for style analysis as proposed by Sharpe (1988), but it appears to be inappropriate for performance attribution.

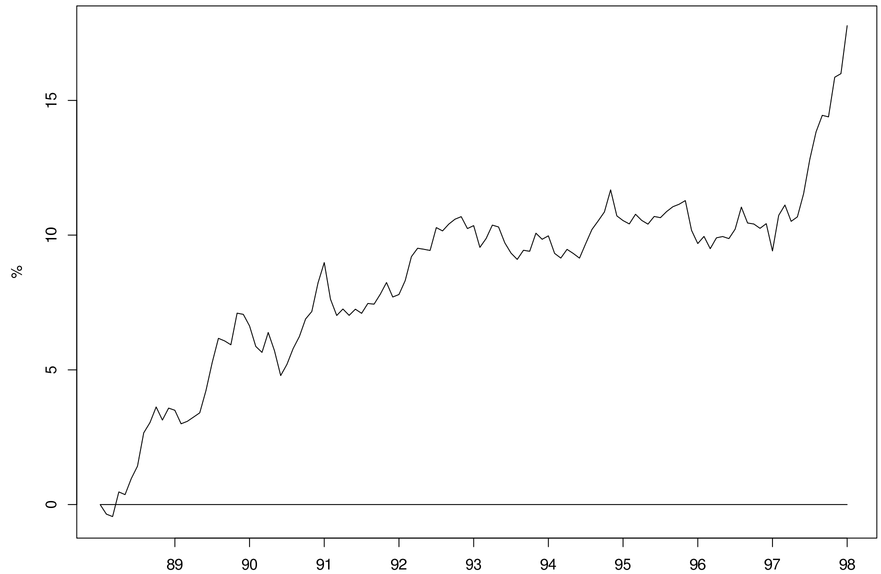

Figure 5: Residual component of the log-return in Figure 3.
This represents the manager's "alpha generation."

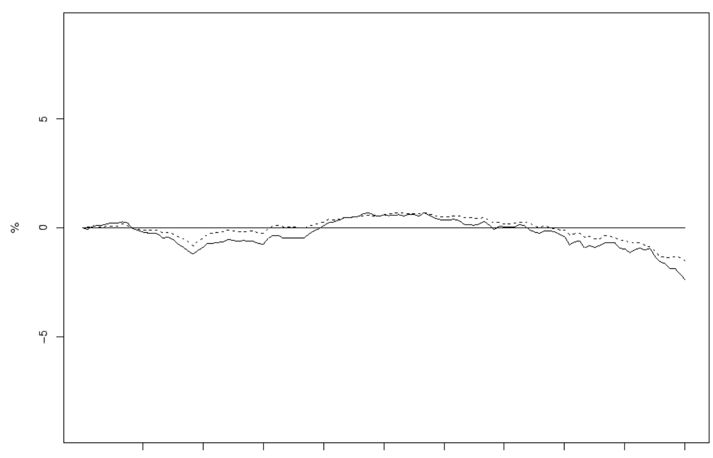

Figure 6: Size component estimated by regression of log-return in Figure 3.
Solid line: vs. relative log-return of largest 25 S&P 500 stocks.
Broken line: vs. relative log-return of largest 100 S&P 500 stocks.

4 Conclusion

We have presented a direct method for calculating the size component of the relative return of an equity portfolio. The calculation depends on analyzing the effect on the return caused by shifts in the distribution of capital in the market. We have shown that the size component calculated in this manner can differ markedly from the conventional estimates of this component using regression analysis. For the purpose of estimating the size component, the rationale behind our methodology appears to be more appropriate than the rationale behind regression analysis. Hence, the conventional estimates for the size component may be misleading, and consequently may lead to inaccurate assessment of managers' stock selection ability.

A Appendix: Calculation of the size component

Let us 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

$$X_1 \geq X_2 \geq \cdots \geq X_n \tag{A.1}$$

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

$$M = X_1 + \cdots X_n.$$

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

$$M' = X_1' + \cdots + X_n'.$$

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

$$\log\bigl(X_i'/X_i\bigr).$$

and the log-return of the market will be

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

Let us consider now the *capitalization weights*

$$w_i = X_i/M,$$

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

$$w_i' = X_i'/M'$$

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

$$\begin{aligned} \log\bigl(X_i'/X_i\bigr) - \log\bigl(M'/M\bigr) &= \log\bigl(X_i'/M'\bigr) - \log\bigl(X_i/M\bigr) \\ &= \log w_i' - \log w_i, \end{aligned} \tag{A.2}$$

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

It follows from (A.1) that

$$w_1 \geq w_2 \geq \cdots \geq w_n.$$

Hence, the *capital distribution curve* at time T_0 is generated by the weights $\{w_1, \ldots, w_n\}$. 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

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

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

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

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

By this definition, the size component corresponds to the transition from point A to point B in Figure 2. The size component is precisely the contribution to the relative return of the stock due to the shift in the capital distribution curve. If there is no shift in the capital distribution curve, then there is no size component in any of the stocks' returns. If a stock maintains its rank in the capital distribution, then the size 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 at time T_0 is

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

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

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

Hence, the log-return of the portfolio will be

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

and the relative log-return will be

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

As in (A.2), the relative log-return of the portfolio can be represented in terms of the change in the capitalization weights of the stocks.

As in (A.3) we define the size component of the relative return of the portfolio to be

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

This size component measures the contribution to the relative return of the portfolio due to the shift in the capital distribution curve.

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 size component for a portfolio relative to the S&P 500, we first calculate the size component of the portfolio relative to the market and then subtract from this the size 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 curve depends only on capital gains and losses, the calculation of the size component in (A.3) and (A.4) will remain unchanged.

References

BARRA (1997, May 5). BARRA Releases New Version of U.S. Equity Model. Press release. Available at `http://www.barra.com/PressReleases/5597-pr.asp`.

Fama, E. and K. French (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics 33*, 3–56.

Fama, E. and K. French (1995). Size and book-to-market factors in earnings and returns. *Journal of Finance 50*, 131–155.

Fama, E. and K. French (1996). Multifactor explanations of asset pricing anomalies. *Journal of Finance 51*, 55–87.

Fernholz, R. (1999). Equity portfolios generated by functions of ranked market weights. Technical report, Social Science Research Network, `http://papers.ssrn.com/`.

Fernholz, R. and R. Garvy (1999, May 17). Diversity changes affect relative performance. *Pensions & Investments*, 112.

Fernholz, R., R. Garvy, and J. Hannon (1998, Winter). Diversity-weighted indexing. *Journal of Portfolio Management 24*(2), 74–82.

Ross, S. (1976). The arbitrage theory of capital asset pricing. *Journal of Economic Theory 13*, 341–360.

Sharpe, W. (1988, November/December). Determining a fund's effective asset mix. *Investment Management Review*, 59–69.