Internal Estimation of Leakage

Leakage in a large-cap equity index occurs when some of the stocks in the index decline in capitalization and are subsequently replaced by larger capitalization stocks from the ambient stock universe (see Fernholz (1996)). Correction for leakage is necessary in order to accurately determine the change in diversity of the index. Measurement of leakage can be difficult in portfolios where the boundary of the portfolio with the ambient universe is not sharply defined. (The boundary of a portfolio is the subset composed of those stocks that lie adjacent to stocks outside the portfolio when the universe is sorted by capitalization.) In this report we develop an estimate of leakage that depends only on the stocks within the portfolio, and not on those of the ambient universe.

Some portfolios have sharp boundaries, some do not. An example of a portfolio with a sharp boundary is the Russell 1000 Index, which holds the 1000 largest capitalization companies in the U.S., so its boundary consists of the smallest stock in the portfolio. Examples of portfolios without sharp boundaries include the market portfolio and the S&P 500 Index. The market portfolio holds all observable publicly-traded stocks, so the ambient universe consists of stocks for which pricing data are not available, e.g., privately held companies. Without pricing data, the behavior of stocks near the boundary cannot be observed and hence cannot be used for leakage measurement. The S&P 500 is composed of 500 high-capitalization companies that Standard and Poor's believes to be representative of the U.S. economy, but they are not the 500 largest companies. Hence, the boundary between the S&P 500 and the rest of the market is fuzzy and not sharply defined as in the case of the Russell 1000.

Suppose we have a large-cap index of n stocks and we wish to estimate the leakage that occurs with respect to a measure of diversity $\mathbf{S}$. Let $X_1 \geq X_2 \geq \cdots \geq X_n$ represent the initial capitalizations of the stocks in the index arranged in descending order. Suppose a unit of time passes and we have returns $r_1, \ldots, r_n$ on the stocks. Suppose that these returns are net of dividends, so the stocks' capitalizations are now

$$X'_i = X_i(1 + r_i),$$

for $i = 1, \ldots, n$. The new capitalizations X'_i of the stocks are no longer in descending order, so let us rearrange them with a permutation p such that

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

To estimate leakage, let us choose m such that $1 < m < n$, and define weights

$$\pi_i = X_i / \big(X_1 + \cdots + X_m\big),$$

for $i = 1, \ldots, m$, and

$$\pi'_{p(i)} = X'_{p(i)} / \big(X'_{p(1)} + \cdots + X'_{p(m)}\big),$$

for $i = 1, \ldots, m$. Then the leakage with respect to $\mathbf{S}$ is estimated by

$$\lambda_m = \log \mathbf{S}(\pi'_{p(1)}, \ldots, \pi'_{p(m)}) - \log \mathbf{S}(\pi_1, \ldots, \pi_m).$$

To correct for leakage, λ_m is added to the change in diversity over the period. In practice, we have found that often the best results occur for $m \approx .8n$.

Let us now observe some examples of the effect of leakage correction using the estimate λ_m. The market we consider is composed of the stocks in the CRSP data base of NYSE, AMEX, and NASDAQ stocks, after the removal of ADRs, REITs, closed-end funds, and stocks for which the market weight at no time reached .00005. Figure 1 shows the cumulative variation in diversity of the market portfolio measured by $\mathbf{D}_p$ (see Fernholz (1999)) with $p = .5$ over the period from 1927 to 1997. (In each of the charts, the straight line is at the average value of the time series.) Figure 2 shows the same variation of diversity corrected for leakage by λ_m with $m = .85n$. Figures 3 and 4 refer to the relative capitalization of a small-stock index as a fraction of the total market capitalization. Here the small-stock index holds all but the 100 largest capitalization stocks in the market. The relative capitalization behaves much as a measure of diversity, and hence leakage correction is appropriate (see Fernholz (1999) and Fernholz (1998)). Figure 3 shows the cumulative change in relative capitalization of the small-stock index without leakage correction, and Figure 4 shows the same time series with leakage corrected by λ_m with $m = .8n$. Figures 5 and 6 show, respectively, the uncorrected and corrected changes in diversity for an index composed of the 500 largest stocks in the market from 1927 to 1997. For first decade or so there are fewer than 500 stocks in the data base, so this index comprises the whole universe over that period. In this case, diversity is measured by $\mathbf{D}_p$ with $p = .5$, and λ_m is calculated with $m = .8n$.

Robert Fernholz
10/12/98

References

Fernholz, R. (1996). Leakage in diversity weighted index portfolios. Technical report, INTECH, Princeton, NJ.

Fernholz, R. (1998, May/June). Crossovers, dividends, and the size effect. *Financial Analysts Journal 54*(3), 73–78.

Fernholz, R. (1999). Portfolio generating functions. In M. Avellaneda (Ed.), *Quantitative Analysis in Financial Markets*, River Edge, NJ. World Scientific.

Market Portfolio

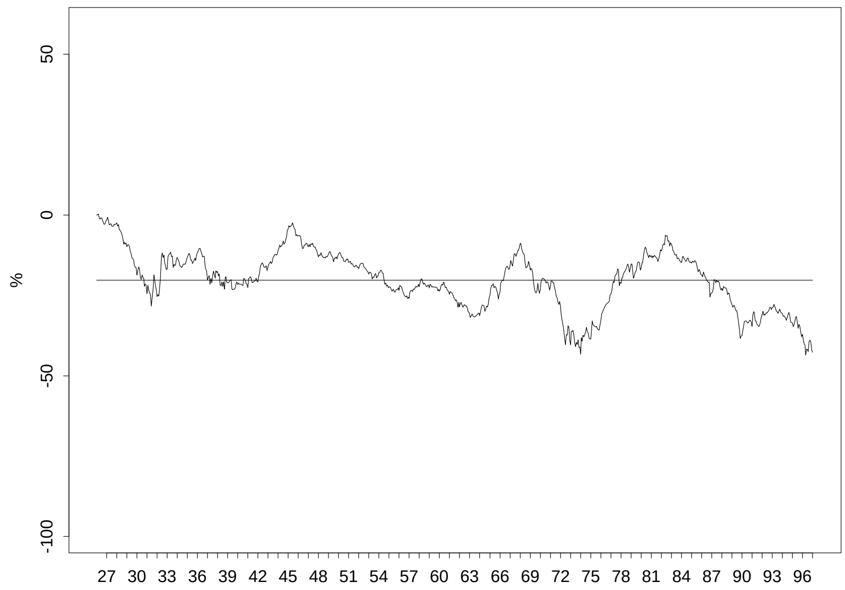

Figure 1: Change in $\mathbf{D}_p$ without leakage correction.

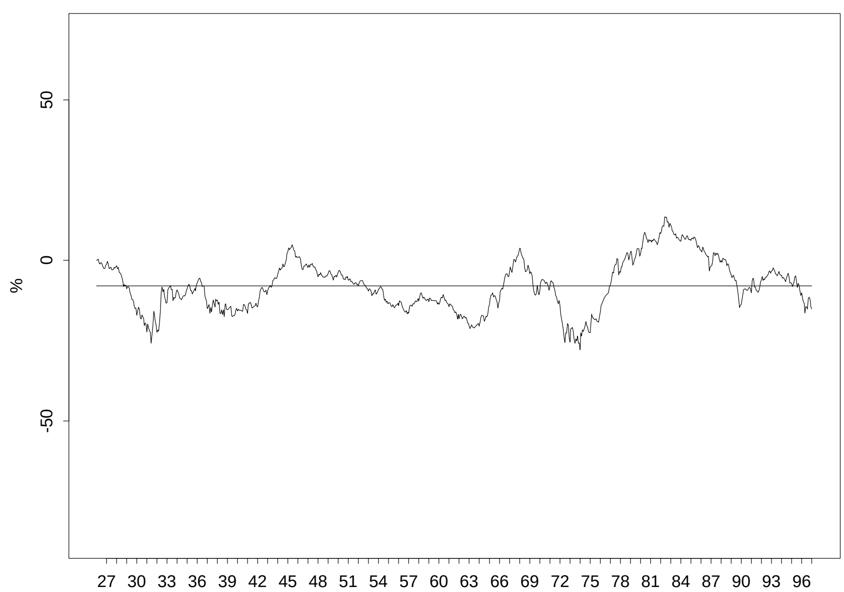

Figure 2: Change in $\mathbf{D}_p$ with leakage correction.

Small-Cap Index

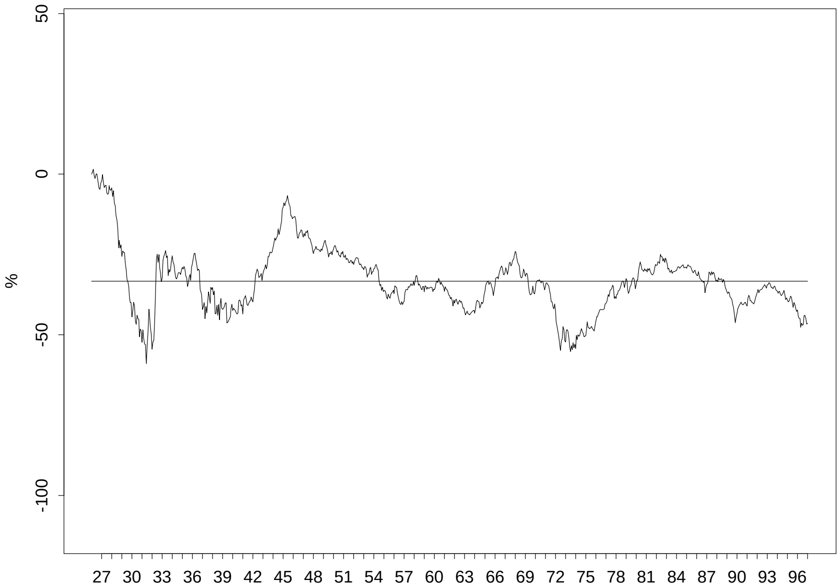

Figure 3: Change in relative capitalization without correction.

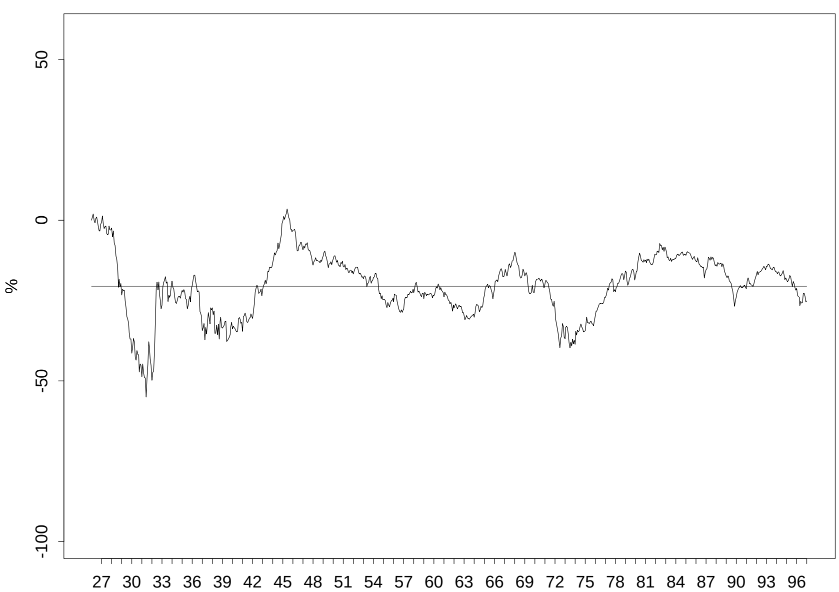

Figure 4: Change in relative capitalization with correction.

Largest 500 Stocks

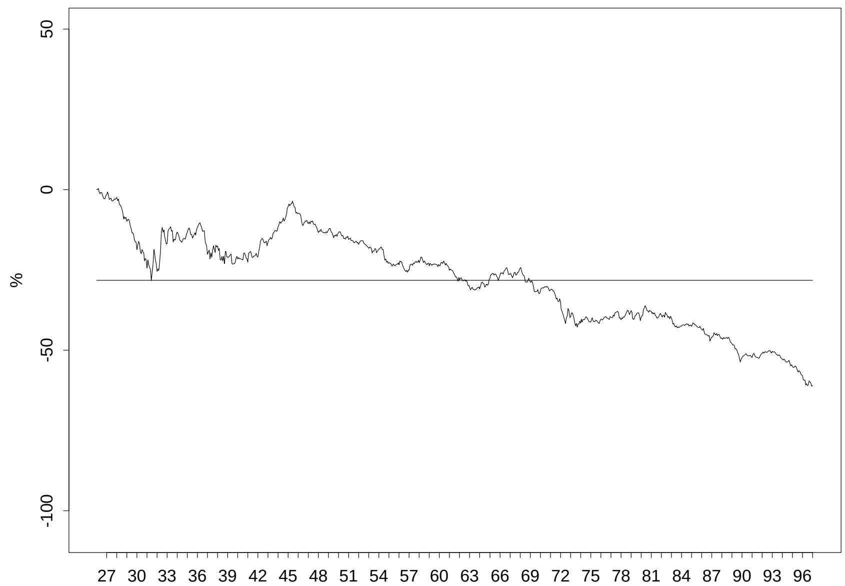

Figure 5: Change in $\mathbf{D}_p$ without leakage correction.

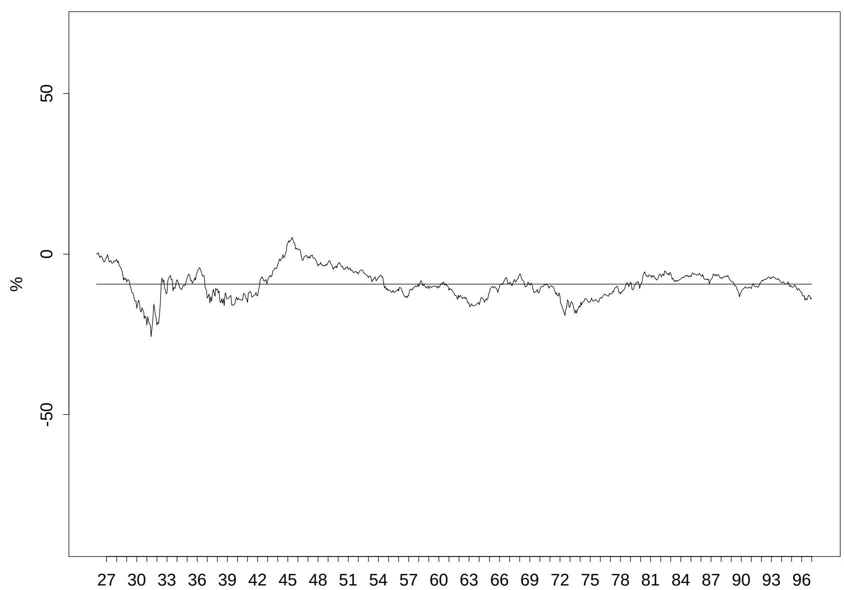

Figure 6: Change in $\mathbf{D}_p$ with leakage correction.