Performance evaluation and benchmark errors (II)*

How to correct performance figures when the benchmark portfolio is less than optimal.

Richard Roll

Performance evaluation with the capital asset pricing beta risk-adjustment is subject to an *ex ante* conceptual problem — the market index may be mean/variance inefficient. In the Summer 1980 issue of this Journal, I showed why the index's inefficiency would result in a different benchmark error for each individual stock and why this could cause incorrect assessments of portfolio managers' abilities. This sequel discusses the practical empirical problems of correcting those errors as well as the possibility for obtaining accurate judgments of ability.

In principle and under ideal conditions, the problem is quite simple. Every individual asset does possess a well-defined *ex ante* benchmark error: It is the asset's true expected excess return (above a risk-free rate) less the market index's true expected excess return multiplied by the asset's beta. Investors can easily make estimates of these items from historical data and can thus obtain historical estimates of the benchmark error for each asset. In principle, these can be employed directly to correct a given manager's observed risk-adjusted performance.

Unfortunately, there is a significant practical impediment to the use of historical estimates. Asset returns are exceedingly "noisy." This is especially true in the case of common stocks, where the intertemporal variability of returns is so large relative to the average return that very large samples are required in order to obtain reasonably accurate estimates. This has been noted many times before in the pages of this Journal and elsewhere.

Our analysis begins with a brief summary of the extent and nature of this problem. The paper then illustrates a method for using noisy historical data in measuring and assessing benchmark errors. With very long time periods, there *is* some hope of detecting and correcting benchmark errors, even in the presence of apparent non-stationarity.

EQUITY RETURNS ARE NOISY

This paper uses eight individual common stocks in all the empirical examples. I chose this number, because eight securities are enough to achieve much of the diversification possible with portfolios of individual stocks. It will be apparent that a larger number can only improve the results. I selected the group of eight firms (Table 1) to illustrate the noisiness problem because of their diversity of industry and their large individual size. The large size was desirable to mitigate the econometric difficulty of non-synchronous trading.[1] The author had no *a priori* knowledge about the performance of these firms.

The statistics in Table 1 are for annual results based on monthly data; that is, the mean and standard deviation of monthly returns were computed for each calendar year 1952 through 1978 inclusive. Then the means and standard deviations of these numbers were calculated across years.

In the case of Allied Chemical, for example, the standard deviation of the 27 mean annual returns was 25.3%. In each year, a standard error for the mean could have been computed from the standard deviation of monthly returns during that year, divided by

* This is a continuation of the discussion in the author's article on this subject in the Summer 1980 issue of this Journal.

1. Footnotes appear at the end of the article.

TABLE 1

Historical Returns Data for a Selected, Diverse Group of Large Individual Common Stocks, Monthly Data, for 27 Years, 1952-78, Inclusive

	Annual Mean Return (%/Annum)		Annual Standard Deviation of Returns (%/Annum)	
Company	Mean	Standard Deviation	Mean	Standard Deviation
Allied Chemical	6.33	25.3	78.9	30.8
Aluminum Company of America	10.8	28.8	90.9	22.1
Consolidated Edison	9.85	25.8	56.0	48.4
Eastman Kodak	13.8	27.5	65.5	22.8
General Electric	11.3	22.8	73.8	20.0
Greyhound	11.3	20.8	67.0	27.5
J.C. Penney	9.23	23.6	70.8	29.1
Texaco	11.6	18.1	66.6	17.0
One-month Treasury bills	3.90	1.90	.471	.294

TABLE 2

Approximate 95% Sampling Intervals for the Expected Annual Return and Standard Deviation of Monthly Returns for a Typical Individual Security (Monthly Data)

Calendar Period (Years)	Sample Mean Return (True Value: 10% Annum)	Standard Deviation of Return (True Value: 86.6%/Annum)
	95% Sampling Interval (%/Annum)	
1	−39.0 to +59.0	51.0 to 122.
2	−24.6 to +44.6	61.7 to 111.
3	−18.3 to +38.3	66.5 to 107.
5	−11.9 to +31.9	71.3 to 102.
10	−5.50 to +25.5	74.8 to 97.0
20	−.957 to +21.0	78.5 to 94.0
30	+1.05 to +18.9	80.0 to 92.7

Note: The table is based on an assumed standard error of the sample mean return of 25% per annum. This implies a true standard deviation of monthly returns of $\sqrt{12} \cdot 25 = 86.6\%$ per annum. For the confidence intervals of the standard deviation, I have assumed normally-distributed returns. This implies that s, the sample standard deviation, satisfies the condition

$$\text{Prob}\left[s < \sigma \sqrt{\frac{x^2}{T-1}} \right] = k$$

where x^2 is the k^{th} fractile of the x^2 distribution with $T-1$ degrees of freedom. (T is the number of months, and σ is the true standard deviation). The expected value of s is approximately

$$\sigma / \left[1 + \frac{3}{4} \frac{1}{T-1} \right]$$

See Johnson and Kotz [1970, pp. 62-65].

the square root of 12. On average, this standard error would have been $78.9/\sqrt{12} = 22.8\%$, which is close to but slightly smaller than the actual standard deviation of mean returns across years.

In order to accomplish accurate performance evaluation, we should estimate a benchmark error for each security in the manager's universe. This requires a reasonably accurate estimate of the security's position in the mean/standard deviation plane. The data in Table 1 suggest that even equity issues of actively-traded large firms require very large calendar periods before such accuracy is feasible. For instance, if a particular security displayed characteristics similar to Allied Chemical — say, a standard error of the annual mean return of 25% and a standard error of the annual standard deviation of returns of 30%, then 95% sampling intervals[2] for the true expected return and true standard deviation would be about those indicated in Table 2; even these values are probably understated, because they assume that returns are generated by stationary normal distributions.

Notice that we require almost thirty years of monthly data before the 95% sampling interval of an individual stock, whose true expected return is 10%, does not overlap zero. With sample sizes like those frequently employed in practice, say, three to five years, the possible range of variation is enormous relative to the expected return; 5% of all three-year samples display observed mean annual returns more than 28% from the true expected return. Only about one-half of all three-year periods would have sample mean returns within 10% of the true expected return.

The distribution of the sample standard deviation of returns is similarly disperse. Also, unlike the sample mean return, the sample standard deviation is biased (downward)[3] and follows an asymmetric distribution. Relative to its true value, however, the standard deviation seems to permit more accurate measurement.

For a given calendar period of available data, a finer gradation of returns, such as daily versus monthly, will not improve the expected return estimate. If the returns are independent over time, the standard error of the grand mean return over the calendar period has the same value regardless of the data gradation,[4] although a finer data gradation will result in a more precise estimate of the estimated standard deviation. The appendix proves these assertions for normally-distributed data.

An additional problem, imposed on top of the gross noisiness of common stock returns, is strong evidence of non-stationary distributions of returns. In other words, *true* means and standard deviations vary over time. There is no theoretical reason why this should not happen, and there are often good reasons why an individual company has had different riskiness and expected return in each historical period.

We can illustrate a simple way to detect non-stationarity with the data for eight large companies in Table 1. If returns were stationary, the standard error of the annual mean return should be about equal to the mean of annual standard deviations of returns divided by $\sqrt{12}$. For example, these two numbers for Allied Chemical are 25.3 and $78.9/\sqrt{12} = 22.8$, respectively. To test the significance of the difference, we can obtain an approximate t-statistic by dividing the difference by the standard error of the mean standard deviation, i.e., by $30.8/[\sqrt{12} \cdot \sqrt{27}] = 1.71$. (There were 27 annual observations.)[5] For Allied Chemical, the t-statistic is $(25.3 - 22.8)/1.71 \cong 1.5$, which indicates little non-stationarity, if any. As Table 3 shows, however, strong non-stationarity is indicated in some of the other stocks and in the Treasury bill rate.

For Consolidated Edison and Eastman Kodak, there is strong evidence of a non-stationary mean return over the 1952-1978 period. The non-stationarity in expected inflation has a very strong effect on Treasury bill returns.

Non-stationarity makes identification of benchmark errors much more difficult and reduces the predictive power of *estimated* benchmark errors from historical data. Thus, for the simple procedure outlined in the next section, I will try to finesse the problem by using only stocks that display little non-stationarity in the historical estimation period.

The noisiness problem is more difficult to solve and large samples seem to be the only solution. By excluding non-stationary stocks, long historical periods can yield reliable estimates of expected returns and standard deviations.

A PROCEDURE FOR MEASURING BENCHMARK ERRORS

Given the noisy, non-stationary nature of common stock returns, measurement and correction of benchmark errors may be difficult. I therefore offer the following methodology, not with any claim of optimality, but merely as a quick and dirty procedure that seems likely to work with monthly common stock data:

1. For a long historical period, 15 years or 180 months at least, calculate individual sample mean annual returns and standard deviations.
2. For each individual stock, test for non-stationarity by comparing the standard error of the mean annual return to the average annual standard deviation of returns (times $\sqrt{12}$). Discard stocks whose t-statistics of the difference are greater than some pre-specified value.
3. When a pre-specified number of stationary stocks is obtained, compute their covariance matrix over the 15 year control period.
4. Then use the sample covariance matrix to obtain a weight γ for the James/Stein estimator of each individual expected return (see Lavely et al. [1980]),

$$R_j^* = R_j + \gamma(\bar{R} - R_j) \qquad (1)$$

where the notation indicates

R_j^*: The James/Stein estimator of the expected return for stock j.

R_j: The sample mean return for stock j over the 15 years.

$\bar{R}$: The mean over all stocks of their sample mean returns.

γ: The James-Stein Weight.[6]

5. Using the estimates from (1), compute the sample α_j for each stock from the 15-year period,

$$\alpha_j = R_j^* - R_F - \beta_j (R_m - R_F) \qquad (2)$$

where R_F and R_m are the mean sample returns of the T-bill rate and the market index, respectively, and β_j is the sample "beta" from the 15 years.[7]

The α_j computed from (2) is the benchmark error for stock j. In order to correct a particular manager's performance measurement, this number should be subtracted from the actual return on stock j during a later period when the manager is being evaluated. Then, the corrected performance is simply the manager's portfolio "alpha" computed from the corrected individual alphas. Note that α_j can be either positive or negative and that the correction can reduce or increase the manager's assessed ability.

Because stocks that are too non-stationary are excluded from this procedure's set of measured benchmark errors, we may be unable to correct every stock in a manager's portfolio. Furthermore, there are obvious statistical estimation errors in the corrections themselves. Finally, if the number of stocks to be corrected is large, we need an even longer time period,

TABLE 3

Indications for Non-Stationarity for Eight Individual Common Stocks, Monthly Data for 27 Years, 1952-78 Inclusive

Company	From Annual Mean Returns	From Mean of Annual Standard Deviation of Returns	
	Estimated Standard Errors of Mean Return		t-statistic of Difference[a]
Allied Chemical	25.3	22.8	1.49
Aluminum Company of America	28.8	26.2	2.13
Consolidated Edison	25.8	16.2	3.56
Eastman Kodak	27.5	18.9	6.74
General Electric	22.8	21.3	1.36
Greyhound	20.8	19.3	.921
J.C. Penney	23.6	20.4	1.95
Texaco	18.1	19.2	−1.14
One-month Treasury bills	1.90	.136	107.4

[a] See text.

because the degrees of freedom in estimating the mean vector indicate the approximate difference between the number of time series observations and the number of stocks. Because stock returns are so noisy, this difference must be at least 180 months. Hence, if 500 stocks are to be corrected, we require more than 55 years of data! This latter problem can be brought under control by using more frequent observations, such as weekly or daily data, but then the non-synchronous trading problem (see note 1) becomes more acute.

For all these reasons, the procedure suggested above is only a beginning and probably can be greatly improved. We must improve it, because current methods lead to unfair assessments of portfolio managers.

THE PROCEDURE WORKS WITH THE S&P 500 INDEX

Benchmark errors are a function of the market index used. If the index is truly mean/variance efficient — if it is an "optimized" portfolio — there are no benchmark errors. The S&P 500, however, is almost assuredly *not* an optimized portfolio, for one very simple reason: It does not include dividends. To the extent that dividends display little variability over time and much less variability than price changes, the S&P 500 should be dominated by another index of the same stocks with dividends included. Thus, if it is employed in performance evaluation, the S&P 500 is very likely to produce benchmark errors. We shall see below that it actually does.

Using monthly returns data for New York Stock Exchange listed common stocks from 1952-1978, the procedure described in section III was carried out for successive 15-year periods. In each 15-year period, I used the first eight alphabetically-ordered stocks whose returns met the stationarity test with a t-statistic less than .5 to form a portfolio whose returns were examined in the 16th year. For example, eight stocks were selected from 1952-1966 and a portfolio of these stocks was tracked in 1967. Then I formed a new group of eight from data over 1953-1967 and tracked a new portfolio in 1968, and so on. In total there were 12 test years, 1967-1978, each preceded by a 15-year estimation period.

The portfolio whose returns were followed in the test year was weighted proportionately to the α_j's computed, during the 15-year estimation period, according to equation (2). The weight of stock j was proportional to,

$$\alpha_j - \min_j (\alpha_j).$$

This weighting scheme eliminated short positions. Better results might be obtained with more complex weighting schemes that allow short positions, but even this very simple scheme produces significant results. The absence of short positions also has the virtue of greater similarity to many managed portfolios.

The performance of these portfolios appears in Table 4.[8] We see two methods for computing α during the test year. The first uses the simple beta computed *during* the test year. The second is a refinement that attempts to correct two problems of the simple beta: (a) there are only 12 monthly observations available during the test year and the beta estimate has a large standard error; (b) the portfolio's beta may be biased because of non-synchronous trading. A Dimson [1979] beta using three monthly lags and no leads and using data for all 16 years of the estimation and control period was computed in an attempt to alleviate both problems. The α, of course, was computed with portfolio and market index excess returns from the test (16th) year only.

The results are highly variable year-by-year, which is to be anticipated in view of the noisy nature of returns. In each test year, the t-statistic of α is quite small. Yet, on average over the 12 test years, the t-statistic of α is significant; its own t-statistic computed from its variance over the 12 years is 2.10. The average α is 5.21 with a t-statistic of 1.82 using the simple beta and 6.14 with a t-statistic of 2.07 using the Dimson beta. Both are significant at the 5% level.

On average, then, even with an eight-stock portfolio, a manager could *appear* to outperform the S&P 500 by over 5% per annum simply because of benchmark errors. Undoubtedly, this apparent per-

TABLE 4

Portfolio Performance Due Solely to Benchmark Error, Eight Stock Portfolios, 1967-78

Year	Simple Beta During Test Year	Dimson Beta[a] From Estimation and Test Years
	Performance (%/Annum)	
1967	12.8	11.4
1968	9.76	9.39
1969	−1.82	−1.38
1970	−11.8	−11.4
1971	8.82	7.74
1972	−13.3	−6.57
1973	17.0	10.1
1974	1.13	−1.97
1975	14.4	23.0
1976	7.54	5.94
1977	12.7	21.2
1978	5.31	6.17
Average all years	5.21	6.14
t-statistic	1.82	2.07

[a] The Dimson [1979] beta with three (monthly) lags and no leads was used.

formance could be and probably has been improved upon by managers using more refined methods (such as choosing small firms and not correcting for non-synchronous trading).

TABLE 5

Individual Stocks in Test Portfolios by Year, 1967-1978

1967	1968
American Bank Note Co.	Amerada Petroleum Corp.
Calumet & Hecla, Inc.	American Can Co.
Champlain National Corp.	American Home Products Corp.
Cudahy Co.	Amstar Corp.
Eastman Kodak Co.	Atlas Chemical Inds. Inc.
Freeport Minerals Co.	Champlain National Corp.
G A T X Corp.	Detroit Edison Co.
General Electric Co.	Eastman Kodak Co.
1969	**1970**
Airco, Inc.	American Home Products Corp.
American Can Co.	Amstar Corp.
American Home Products Corp.	Bush Unvl. Inc.
Atlas Chem. Inds. Inc.	C I T Financial Corp.
Borden, Inc.	Callahan Mining Corp.
Callahan Mining Corp.	City Investing Co.
J. I. Case, Co.	City Stores, Inc.
Conoco, Inc.	Diamond International Corp.
1971	**1972**
Allied Chemical Corp.	Allied Chemical Corp.
AMAX, Inc.	Amax, Inc.
Archer Daniels Midland Co.	Archer Daniels Midland Co.
Asarco, Inc.	Associated Dry Goods Corp.
Associated Dry Goods Corp.	Borden, Inc.
Belding Heminway, Inc.	Bush Unvl. Inc.
C I T Financial Corp.	C I T Financial Corp.
Callahan Mining Corp.	Callahan Mining Corp.
1973	**1974**
A C F Inds., Inc.	Adams Express Co.
Allegheny Corp.	Airco, Inc.
Allied Chemical Corp.	Allis Chalmers Corp.
Allis Chalmers Corp.	American Std., Inc.
American Broadcasting Co., Inc.	American Tel & Telegraph Co.
American Home Products Corp.	Archer Daniels Midland Co.
Associated Dry Goods Corp.	Briggs & Stratton Corp.
Borden, Inc.	Bush Unvl. Inc.
1975	**1976**
Airco, Inc.	A C F Inds., Inc.
Allis Chalmers Corp.	Acme Cleveland Corp.
Amax, Inc.	American Can Co.
American Bakeries Co.	American Distilling Co.
American Can Co.	Archer Daniels Midland Co.
American Std., Inc.	Asarco, Inc.
Asarco, Inc.	Beatrice Foods Co.
Borden, Inc.	Borden, Inc.
1977	**1978**
Acme Cleveland Corp.	A C F Inds. Inc.
Adams Express Co.	Adams Express Co.
Aguirre Co.	Aguirre Co.
Allegheny Corp.	Allegheny Corp.
Allegheny Power Sys., Inc.	Allegheny Power Sys. Inc.
Amax, Inc.	American Can Co.
Asarco, Inc.	American Stores Co.
Beatrice Foods Co.	Armstrong Cork Co.

APPENDIX

STANDARD ERRORS OF SAMPLE MEAN AND SAMPLE STANDARD DEVIATION ACCORDING TO FREQUENCY OF OBSERVATION, NORMALLY-DISTRIBUTED RETURNS

For a given calendar period, the frequency of observation has no impact on the standard error of the mean return, but it does affect the standard error of the standard deviation of returns. Assume that:

1. Returns are continuously-compounded (and thus additive) and are normally distributed.
2. Using the most frequent observation interval, there are T available observations during a fixed calendar period.
3. For some less frequent observation interval, there are $K = T/N$ available observations, where N is the number of most frequent observations per longer interval.

For instance, daily observations could be the most frequent available, but we might desire to study results on a monthly basis. For $N = 21$ trading days per month and five years of data, we would have $K = 60$, $T = 1260$.

Using daily and monthly as an example, with T daily returns, we can express the sample grand mean return as a return per *month*:

$$\overline{R} = (N/T)\textstyle\sum_{t=1}^{T} R_t \qquad (A.1)$$

where R_t is the daily return on day t.

Now suppose that we had calculated monthly returns. For the τth month, the return is

$${}_mR_k = R_{N(\tau-1)-1} + \ldots + R_{N\tau} \qquad (A.2)$$

where the prescript indicates monthly returns. Computed from these monthly returns, the grand mean is simply:

$$\overline{R} = (1/K)\textstyle\sum_{\tau=1}^{K} {}_mR_\tau \qquad (A.3)$$

From (A.1), the standard error of the grand mean is $\sigma(\overline{R}) = N\sigma/\sqrt{T}$ where σ^2 is the daily return variance. From (A.3), the standard error is $\sigma(\overline{R}) = \sigma({}_mR)/\sqrt{K}$, where $\sigma({}_mR)$ is the standard deviation of monthly returns. With temporal independence, from (A.2) we have $\sigma({}_mR) = \sqrt{N}\,\sigma$ and thus $\sigma(\overline{R}) = \sqrt{N} \cdot \sigma/\sqrt{K} = N\sigma/\sqrt{T}$ as before, regardless of the interval frequency employed. Indeed, (A.1) and (A.3) are identical after a simple substitution, so they must have the same standard error. This is true for all distributions with finite means, not just the normal.

For the sample variance, however, the story is quite different. From daily observations, the estimate of the *monthly* variance is:

$$s_d^2 = \frac{N}{T-1}\sum_{t=1}^{T} (R_t - \overline{R})^2 \qquad (A.4)$$

while from monthly observations, the sample variance is:

$$s_m^2 = \frac{1}{K-1}\sum_{\tau=1}^{K} ({}_mR_\tau - \overline{R})^2 \qquad (A.5)$$

Given normally and independently distributed returns,

$$(T-1)\, s_d^2/N\sigma^2$$

is distributed as x^2 with $T-1$ degrees of freedom while

$$(K-1)\, s_m^2/\sigma({}_mR)$$

is distributed as x^2 with $K-1$ degrees of freedom. Using the identity $N\sigma^2 \equiv \sigma({}_mR)$, it can thus be shown that:

$$\text{Var}(s_d^2) = 2\sigma^4N^2/(T-1)$$

and

$$\text{Var}(s_m^2) = 2\sigma^4 N^2/(K-1),^9$$

while $E(s_d^2) = E(s_m^2) = N\sigma^2$. Although the expected values are identical, daily returns always produce a less variable estimate of the standard deviation. For 21 trading days per month and, say, five years of data, the standard error of the estimated standard deviation of monthly returns using monthly data is $\sqrt{1259/59} = 4.6$ times larger than when daily data are used.

It is interesting to note that the superior accuracy of the sample standard deviation computed from more frequent data does not help much in testing the sample mean against some hypothesized population value. For example, to test whether the mean return is significantly different from zero, two t-statistics are possible: for daily data, the mean return per day is R/N and the standard deviation of daily returns is $s_d/\sqrt{N}$. Thus, the t-statistic for a test of positive expected return is:

$$t_d = \sqrt{NT} \cdot (\bar{R}/N)/s_d = \sqrt{K}\,\bar{R}/s_d$$

For monthly data, the t-statistic is:

$$t_m = \sqrt{K}\,\bar{R}/s_m$$

and, $E(t_d) = E(t_m)$. Under the null hypothesis, both t-statistics are asymptotically normal with the same variance (unity). Thus, in large samples, there is little to be gained from using more frequent observations in testing a hypothesis about expected return. In very small samples, of course, there is some advantage, since t_d follows a t-distribution with more degrees of freedom $(T - 1)$ than $t_m(K - 1)$. This also implies that observation frequency should be larger than about T/30, since the t-distribution for less frequent data would be quite non-normal.

[1] The effect of non-synchronous trading on risk assessment has been investigated by Scholes and Williams [1977] and Dimson [1979]. Roll [1980] argues that the apparent return premia of small firms after risk adjustment can be attributed to improper risk measurement caused by non-synchronous trading.

[2] The interval in which 95% of all sample values will fall.

[3] The sample variance,

$$s^2 = \frac{1}{T-1}\sum(x - \bar{x})^2,$$

is unbiased, but $E(s) < \sqrt{E(s^2)}$.

[4] This is exactly true for continuously-compounded returns and approximately correct for discretely-compounded returns.

[5] This t-statistic should not be regarded as rigorous, but only as an indication. It is asymptotically distributed as Student's t, but in small samples the two components of the numerator may be correlated and the mean of standard deviations may be only approximately normal.

[6] There are several different weighting schemes available but since stocks are cross-sectionally correlated, a weight similar to that recommended by Efron and Morris [1976] is probably indicated. Their weight, used here, is:

$$\gamma = \text{Min}\,\{1, (K - 2)/[(T - K - 3)\,R'\hat{V}^{-1}R]\}$$

where R is the sample mean return vector, $\hat{V}$ is the sample covariance matrix for K stocks and T periods. In the results below, K = 8, T = 180 months. Notice that the James/Stein method has the effect of shrinking each individual stock's mean return estimate toward the grand mean of all stocks, and the degree of shrinkage is inversely proportional to return variability relative to mean return.

[7] A refinement of this procedure would entail a James/Stein estimator of β_j. However, simulation by Jobson and Korkie [undated] indicate that such a refinement is not worth very much. The simple product moment covariance matrix is almost as good as a James/Stein covariance matrix estimate.

[8] The individual securities in each portfolio are reported by year in Table 5. The index was the S&P Composite 500 after June 1962 (when it became available on the CRSP data base) and a value-weighted index of NYSE stocks without dividends from 1952-June 1962. Using the value-weighted index throughout makes only a third significant digit alteration in the results.

[9] See any good statistics textbook, e.g., Neter, Wasserman, and Whitmore [1978, pp. 164, 305].

REFERENCES

Dimson, Elroy, "Risk Measurement When Shares are Subject to Infrequent Trading," *Journal of Financial Economics*, 7 (June 1979), 197-226.

Efron, Bradley and Carl Morris, "Families of Minimax Estimators of the Mean of a Multivariate Normal Distribution," *The Annals of Statistics* 4 (1976) 11-21.

Jobson, J. D. and Bob Korkie, "Improved Estimation and Selection Rules for Moskowitz Portfolios," Dept. of Finance and Management, University of Alberta (undated).

Johnson, Norman L. and Samuel Kotz, *Continuous Univariate Distributions, I.* (New York: Houghton-Mifflin, 1970).

Lavely, Joe, Gordon Wakefield, and Bob Barrett, "Toward enhancing beta estimates," *Journal of Portfolio Management* (Summer 1980), 43-46.

Neter, John, William Wasserman, and G. A. Whitmore, *Applied Statistics* (Boston: Allyn and Bacon, 1978).

Roll, Richard, "A Possible Explanation of the Small Firm Effect," (working paper 11-80, Graduate School of Management, UCLA, October, 1980).

Roll, Richard, "Performance evaluation and benchmark errors," *Journal of Portfolio Management* (Summer 1980), 5-12.

Scholes, Myron and Joseph Williams, "Estimating Betas from Nonsynchronous Data," *Journal of Financial Economics*, 5 (December 1977), 309-327.