ON COMPUTING MEAN RETURNS AND THE SMALL FIRM PREMIUM

Richard ROLL*

University of California, Los Angeles, CA 90024, USA

The mean return computational method has a substantial effect on the estimated small firm premium. The buy-and-hold method, which best mimics actual investment experience, produces an estimated small-firm premium only one-half as large as the arithmetic and re-balanced methods which are often used in empirical studies. Similar biases can be expected in mean returns when securities are classified by any variable related to trading volume.

1. Introduction

There is a potentially serious problem in estimating expected return differences between small and large firms. Even with exactly the same sample observations, the method used to compute sample mean returns can have a substantial effect on the estimates.

With an arithmetic computational method, daily returns on individual stocks are averaged across both firms and days to obtain the mean daily return on an equally-weighted portfolio; then the portfolio's mean daily return is compounded to obtain an estimate of the expected return over a longer interval. With a buy-and-hold method, individual stock returns are first obtained for the longer interval by linking together the daily individual returns; then an equally-weighted portfolio's mean return is computed by averaging the longer-term (individual) returns.

Defining a 'longer interval' as one year, the arithmetic method produces an average annual return difference of 14.9 percent between AMEX and NYSE stocks[1] over the 19 complete calendar years, 1963–1981 inclusive. The buy-and-hold method gives an annual return difference of only 7.45 percent. Assuming that annual returns are statistically independent, the arithmetic

---

[1]The effect of smallness can be measured by the difference in returns of stock listed on the American Exchange (AMEX) and the New York Exchange (NYSE) because AMEX issues are, on average, much smaller than NYSE issues. Most of the results presented here are based on the AMEX–NYSE differential because it is convenient and easy to use. Some confirmatory results based directly on measured size will also be presented.

method's return differential had an associated t-statistic of 3.07 while the buy-and-hold method yielded a t-statistic of 1.53.

Speculation on possible causes of the small firm premium has occupied the attention of many finance theorists over the past few years; but perhaps this attention has been premature. If the estimated small firm premium can be cut in half simply by compounding individual returns before averaging them, some consideration should be given to whether the magnitude of the true premium is really all that large. The various explanations for the premium offered so far would become more plausible if the premium is actually smaller than has been previously reported.

This paper investigates why the mean return computational method can be such a significant choice in some empirical research. The reason seems to be that individual asset returns are not as well-behaved as we might like. Individual assets do not trade continuously and there are significant trading costs. In some empirical studies, the effect of these factors might be safely ignored; but when the object of investigation is related to trading volume (and thus to trading frequency and trading costs), there can be measurement problems. Firm size is related to trading volume and it is used as an example throughout the paper. Other variables related to size and to trading, such as dividend yield, price/earnings ratio, and beta, could also present similar empirical difficulties. Section 2 gives a brief theoretical discussion of mean return computational methods and section 3 presents details of the empirical results for small firm premia.

2. Compounding and the bias in mean return calculation

2.1. Formulae for computing mean returns

To elucidate the differences in mean return computation and explain why they might produce different results, consider a sample of N securities, each having returns observed for T periods. Let R_{it} be the value relative ($1 + $ return), of security i in period t. Suppose also that investment results are reviewed every τ periods. For example, if data were available daily but returns were to be reviewed every month, we would have $\tau \cong 21$ since there are usually about 21 trading days per month.

Two alternative methods of computing the mean equally-weighted return over the review period can be written algebraically as

$$\bar{R}_{AR} = \left[\frac{1}{N \cdot \tau} \sum_i \sum_t R_{it} \right]^{\tau}, \tag{1}$$

$$\bar{R}_{BH} = \frac{1}{N} \sum_i \left[\prod_t R_{it} \right], \tag{2}$$

where the subscripts 'AR' and 'BH' denote 'arithmetic' and 'buy-and-hold', respectively. These labels are intended to portray the sense of the computation method. The first method (1) is simply an arithmetic mean raised to the τth power while the second method gives the actual investment results an investor would achieve from buying equal dollar amounts of N securities and holding the shares for τ periods.

There is also a third possible definition of mean return,

$$\bar{R}_{RB} = \prod_t \left[\frac{1}{N} \sum_i R_{it} \right],$$ (3)

where the subscript 'RB' stands for 'rebalanced'. This would be the actual investment return (ignoring transactions costs) on a portfolio which begins with equal investments in the N securities and *maintains* equal investments by rebalancing at the end of each period, $t = 1, \ldots, \tau$.

To compare results over different review periods, we must choose some typical and familiar calendar interval, say a year, and express the results as percentage returns over that common calendar interval. In the tables below, annualization is accomplished and reported for 'linked' returns; the review period returns within each calendar year are simply multiplied together (or linked) in order to obtain an annual return.[2] Linked annualization includes *every* daily observation in some review period during the year. This assures that in any comparison of the results across review periods, the observed differences are due to review period alone and cannot be ascribed to slightly different sample observations.

The next two subsections investigate some properties of these sample mean returns. Subsection 2.2 derives their expected values under the assumption of temporally independent individual asset returns. Subsection 2.3 then examines the effect of intertemporal dependence.

---

[2]The exact formulae for linked returns can be written as follows. Let $\bar{R}_m(y, \tau)$ denote the mean annualized linked return for year y ($y = 1, \ldots, Y$) using a review period whose length is τ trading days and using method ($m = $ BH, AR, RB), to compute the review period returns. Then,

$$\bar{R}_{BH}(y, \tau) = \prod_{j = (y-1)k_\tau + 1}^{y \cdot k_\tau} \left[\frac{1}{N} \sum_i \prod_{t = (j-1)\tau - 1}^{j \cdot \tau} (R_{it}) \right],$$

$$\bar{R}_{AR}(y, \tau) = \prod_{j = (y-1)k_\tau + 1}^{y \cdot k_\tau} \left[\frac{1}{N \cdot \tau} \sum_i \sum_{t = (j-1)\tau - 1}^{j \cdot \tau} R_{it} \right]^\tau,$$

$$\bar{R}_{RB}(y, \tau) = \prod_{j = (y-1)k_\tau + 1}^{y \cdot k_\tau} \left\{ \prod_{t = (j-1)\tau + 1}^{j \cdot \tau} \left[\frac{1}{N} \sum_i R_{it} \right] \right\},$$

where $k_\tau = T/(Y \cdot \tau)$ is the number of review periods per year and T is the total number of trading days in the entire sample. When returns are reviewed in natural calendar intervals such as months, the review period cannot be a fixed number of trading days and thus τ in the formulae above varies slightly with the actual number of trading days.

2.2. Sample mean return biases with temporal independence

Following Blume (1974), assume that each individual asset return is drawn from a stationary distribution with temporally independent disturbances; that is,

$$\tilde{R}_{it} = \mu_i + \tilde{\varepsilon}_{it}, \quad \forall i, \tag{4}$$

with $E(\tilde{R}_{it}) = \mu_i$, a constant for all t, and where the unexpected return, $\tilde{\varepsilon}_{it}$, satisfies $\text{cov}(\tilde{\varepsilon}_{i,t}, \tilde{\varepsilon}_{i,t-j}) = 0$ for $j \neq 0$.

The expected value of the arithmetic mean (1) can be expressed as

$$E(\bar{R}_{AR}) = E\left[\left(\frac{1}{N}\sum_i \mu_i + \tilde{h}\right)^\tau\right], \tag{5}$$

where

$$\tilde{h} \equiv \frac{1}{N \cdot \tau}\sum_i \sum_t \tilde{\varepsilon}_{it}$$

is the average disturbance on the equally-weighted portfolio over the sample review period τ.

The expected value of the buy-and-hold mean (2) is

$$E(\bar{R}_{BH}) = \frac{1}{N}\sum_i \left[E\prod_t (\mu_i + \tilde{\varepsilon}_{it})\right] = \frac{1}{N}\sum_i (\mu_i^\tau). \tag{6}$$

This follows since the expectation can be taken inside the product with independent returns and since $E(\tilde{\varepsilon}) = 0$, by definition.

The rebalancing method (3) produces a mean return whose expectation is

$$E(\bar{R}_{RB}) = \prod_t \left[\frac{1}{N}\sum_i \mu_i\right] = \left(\frac{1}{N}\sum_i \mu_i\right)^\tau, \tag{7}$$

where, again, the expectation can be taken inside the product because of time independence.

Expressions (5), (6) and (7) imply that the three different mean return definitions do not produce the same results. By Jensen's inequality,

$$E(\bar{R}_{AR}) \geq E(\bar{R}_{RB}),[3]$$

---

[3] Jensen's inequality for a random variable $\tilde{x}$ and a convex function $f(x)$ is $E[f(\tilde{x})] \geq f[E(x)]$. Let $\tilde{x} \equiv (1/N)\sum_i \mu_i + \tilde{h}$; then $f(\tilde{x}) = \tilde{x}^\tau$ is convex since $\tau > 1$. $E(\bar{R}_{AR}) > E(\bar{R}_{RB})$ follows immediately from (5) and (7) since $E(\tilde{h}) = 0$.

with strict inequality if $\text{var}(\tilde{h}) > 0$, and

$$E(\bar{R}_{\text{BH}}) \geqq E(\bar{R}_{\text{RB}}),^4$$

with strict inequality if $N > 1$ and at least two assets have different returns. Since we generally have some randomness $[\text{var}(\tilde{h}) > 0]$, and many securities, $(N > 1)$, the rebalanced method generally should produce lower mean returns than either the arithmetic or the buy-and-hold method, provided that returns are temporally independent.

The relation between the buy-and-hold and arithmetic means is more complex; and, indeed, neither is invariably smaller than the other. The larger the cross-sectional dispersion of individual expected returns, the larger $E(\bar{R}_{\text{BH}})$ relative to $E(\bar{R}_{\text{AR}})$. But there is an offsetting influence: the larger the intertemporal dispersion of unexpected returns $(\tilde{h})$, the larger $E(\bar{R}_{\text{AR}})$ relative to $E(\bar{R}_{\text{BH}})$.[5] Their relation in a given sample depends, therefore, on the characteristics of the underlying individual returns.

2.3. Time series dependence and its effect on estimated expected returns

The effect of serial dependence is seen most easily by examining expected mean returns when the review period is doubled, say from daily to bi-daily or from bi-weekly to monthly. Assume first that returns are collected for the shorter review period and then let $\tau = 2$ (a doubling of the period). Over the doubled review period, the three mean returns are

$$\bar{R}_{\text{AR}} = \left[\frac{1}{N} \sum_i \left(\mu_i + \frac{\varepsilon_{i1} + \varepsilon_{i2}}{2} \right) \right]^2, \tag{8}$$

[4]Define $f(\mu_i) = \mu_i^\tau$, a convex function for $\tau > 1$. With $1/N$ used as a (pseudo) probability, $E(\bar{R}_{\text{BH}}) \geqq E(\bar{R}_{\text{RB}})$ follows immediately from (6) and (7). (Cf. footnote 3.) Strict inequality holds if at least two μ_i's are different. [This result was noted by Cheng and Deets in (1971).]
The inequality above grows with the cross-sectional dispersion in μ_i, ceteris paribus. To prove this, expand μ_i^τ in a Taylor series about $\bar{\mu} \equiv (1/N) \sum_i \mu_i$; the second-order term is a positive function of the cross-sectional variance in μ_i. If μ_i were cross-sectionally normally distributed, the variance *alone* would determine the size of the inequality.
[5]This can be confirmed by using a Taylor series expansion of $E(\bar{R}_{\text{AR}})$. Define $\bar{\mu} = (1/N) \sum_i \mu_i$; then

$$E(\bar{R}_{\text{AR}}) = \bar{\mu}^\tau E \left[1 + \frac{\tilde{h}^2}{2} (\tau)(\tau - 1)\bar{\mu}^{-2} + \frac{\tilde{h}^3}{3!}(\tau)(\tau - 1)(\tau - 2)\bar{\mu}^{-3} + \ldots + \tilde{h}^\tau \bar{\mu}^{-\tau} \right].$$

Jensen's inequality (see footnote 4 above), implies that $E(\bar{R}_{\text{BH}}) > \bar{\mu}^\tau$ with the inequality being larger the larger the cross-sectional variance in μ_i. But the term in brackets just above shows that $E(\bar{R}_{\text{AR}})$ increases with the higher moments of $\tilde{h}$ (since $\bar{\mu}$ is strictly positive). For example, the second term in brackets involves the variance of $\tilde{h}$. Conceivably, this term could more than offset the cross-sectional variance in μ_i. If the unexpected arithmetic portfolio return h happens to be normally-distributed, the expression above simplifies to $E(\bar{R}_{\text{AR}}) = \bar{\mu}^\tau[1 + k \cdot \text{var}(\tilde{h})]$ with the constant $k > 0$. In this case, there is a simple and direct tradeoff between the cross-sectional variance in expected return, μ_i, and the variance of the unexpected portfolio return, $\tilde{h}$.

$$\bar{R}_{\text{BH}} = \frac{1}{N} \sum_i \left[(\mu_i + \varepsilon_{i1})(\mu_i + \varepsilon_{i2}) \right], \tag{9}$$

$$\bar{R}_{\text{RB}} = \left[\frac{1}{N} \sum_i (\mu_i + \varepsilon_{i1}) \right] \left[\frac{1}{N} \sum_i (\mu_i + \varepsilon_{i2}) \right], \tag{10}$$

where $R_{it} = \mu_i + \varepsilon_{it}$ is the observed return on individual stock i ($i = 1, \ldots, N$) in period t, and μ_i is i's single-period (i.e., shorter review period) expected return.

For notational convenience, define the cross-sectional averages

$$\bar{\mu} = \frac{1}{N} \sum_i \mu_i \quad \text{and} \quad \bar{\varepsilon}_t = \frac{1}{N} \sum_i \varepsilon_{it}.$$

Then the three mean returns have expected values,

$$E(\bar{R}_{\text{AR}}) = \bar{\mu}^2 + \tfrac{1}{2}(\sigma_{\bar{\varepsilon}}^2 + \sigma_{\bar{\varepsilon}_1, \bar{\varepsilon}_2}), \tag{11}$$

$$E(\bar{R}_{\text{BH}}) = \frac{1}{N} \sum_i \mu_i^2 + \frac{1}{N} \sum_i \sigma_{\varepsilon_{i1}, \varepsilon_{i2}}, \tag{12}$$

$$E(\bar{R}_{\text{RB}}) = \bar{\mu}^2 + \sigma_{\bar{\varepsilon}_1, \bar{\varepsilon}_2}, \tag{13}$$

where σ_x^2 is the variance of x and $\sigma_{x, y}$ is the covariance of x and y.

Even with serial dependence, the expected arithmetic mean still exceeds the expected rebalanced mean in all circumstances since,

$$E(\bar{R}_{\text{AR}} - \bar{R}_{\text{RB}}) = \tfrac{1}{2}(\sigma_{\bar{\varepsilon}}^2 - \sigma_{\bar{\varepsilon}_1, \bar{\varepsilon}_2}) > 0. \tag{14}$$

Comparing the buy-and-hold means and the rebalanced means, we have

$$E(\bar{R}_{\text{BH}} - \bar{R}_{\text{RB}}) = \sigma_{\mu_i}^2 + \left(\frac{1}{N} \sum_i \sigma_{\varepsilon_{i1}, \varepsilon_{i2}} - \sigma_{\bar{\varepsilon}_1, \bar{\varepsilon}_2} \right).$$

With no serial dependence in the ε's, the term in parentheses is zero and the BH mean would exceed the RB mean by the cross-sectional variance in expected individual returns.

However, with negative serial dependence in unexpected individual returns (ε_{i1} and ε_{i2}) or positive dependence in portfolio returns ($\bar{\varepsilon}_1$ and $\bar{\varepsilon}_2$), the rebalanced mean would become larger; enough such dependence could conceivably render it larger that the buy-and-hold mean. Since the expected arithmetic mean exceeds the expected rebalanced mean, it too could be larger than the BH mean with enough serial dependence of the right type.

There is some reason to anticipate just this type of serial dependence because of the intertemporal characteristics of individual returns. Scholes and Williams (1977, pp. 313–314) explain that because of non-synchronous trading individual assets display first-order *negative* serial dependence while diversified portfolios display *positive* dependence. A difference in the sign of serial dependence between individual assets and portfolios is relevant here because buy-and-hold (BH) means are mainly affected by individual asset serial dependence [see (12)], while the arithmetic (AR) and rebalanced (RB) means are affected by portfolio serial dependence [see (11) and (13)]. The Scholes/Williams explanation implies that BH means would tend to fall as review period lengthens while the AR and RB means would tend to rise.

There is also negative serial dependence induced in very short-term returns because of the institutional arrangement of trading. Neiderhoffer and Osborne (1966) pointed out that negative serial dependence should be anticipated when a market maker is involved in most transactions (because successive transactions are conducted at either the bid or the asked price).[6]

First-order negative serial dependence in individual returns has the effect of widening the disparity between the buy-and-hold mean and the arithmetic and rebalanced means as the review period lengthens. This follows from the fact that a doubling of the review period introduces serial covariance terms in *addition* to those already present. However, the marginal effect of lengthening the review period should probably diminish as the review period becomes longer; the effect on measured mean return should be greater when changing from, say, a daily to a weekly review period than from a monthly to an annual period. The exact impact of serial dependence can, of course, only be determined empirically and we now turn to an examination of the data.

3. The empirical small firm premium

3.1. Results

In the previous section, we found that the computational formula for sample mean returns can affect the estimated expected return. The buy-and-hold (BH) mean (2) gives an unbiased estimate of the holding period return on a realistic portfolio. The rebalanced (RB) mean (3), gives an unbiased estimate of return for its strategy but it is not realistic if the period is short since rebalancing is so costly. Except under a fortuitous combination of circumstances, the arithmetic (AR) mean (3) gives a biased estimate of *both* the rebalanced and the buy-and-hold investment returns.

---

[6]A paper by Blume and Stambaugh (1983), which came to my attention after the first version of this paper was written, investigates this explanation for serial dependence in detail. They find empirical results very similar to those reported here. See also Cohen et al. (1979).

Although the arithmetic and rebalanced methods of calculating the mean return probably do not portray realistic investment experience, the small-firm premium is calculated as the *difference* between the two mean returns and one might hope that the improper portrayal in these methods would cancel. Unfortunately, this is not likely for several reasons. The intertemporal variance in the portfolio disturbance, $\tilde{h}$, and the cross-sectional variance in individual security expected returns, μ_i, will not be the same in samples of large and small firms. The disturbance, $\tilde{h}$, will almost certainly have a larger variance for portfolios of small firms while the cross-sectional variances of μ_i within large- and small-firm portfolios could conceivably differ in either direction. Furthermore, serial dependence has an effect which is stronger for stocks with lower trading volumes and thus with less synchronous trading and with larger bid/ask spreads.

Empirical evidence is reported in table 1. Small Firm Premia (AMEX–NYSE) are given for the 19 complete calendar years, 1963–1981, according to the method of computation and the 'review' period. As explained earlier, the 'review' period refers to the rebalancing interval for buy-and-hold returns. For example, with a monthly review period, an equal allocation is made to stocks listed on the first day of the month and the original positions are held until the end of the month. This is repeated for each calendar month of the sample. The daily rebalancing method uses the same available returns, but it re-initializes equal positions every day during the month. The arithmetic method simply averages the same available returns during the month.

In order to compare results across the different review periods, returns are annualized by linking together the review period returns obtained during the calendar year.[7] Thus, there are 19 annual observations (one for each calendar year, 1963–81), regardless of the review period.[8] Means and t-statistics are calculated from the 19 annual returns differences between exchanges; t-

---

[7]See footnote 2 for exact computational formulae.

[8]Daily and bi-daily returns are over trading day intervals, while weekly and longer returns use actual calendar intervals. In the weekly case, the first week of the year ends on the same day of the week as the last trading day of the previous year, say Thursday for a given year. Then weekly returns are computed from Thursday to Thursday during that year. If the year does not terminate on a Thursday trading day, the last 'weekly' return of the year is over the remaining fraction of a calendar week. This method of year-end padding was used to ensure that every daily return during a year was included, regardless of the review period. Only the bi-daily, weekly, and bi-weekly returns are subject to such padding because the other intervals are evenly divisible into years.

Weekly returns are not always for five trading day intervals. During 1968, the exchanges were closed on Wednesdays for part of the year so that a week was composed of only four trading days. Holidays are also a problem for weekly returns; if the calendar week ended on a holiday, the return was computed through the next trading day. Then the subsequent week's return covered four trading days. Bi-weekly returns were treated identically to weekly returns with respect to year-end padding, holidays, and exchange closings.

Table 1

The small firm premium as measured by the difference in returns between American Exchange and New York Exchange listed stocks, 1963–1981 (basic data are daily, January 2, 1963 — December 31, 1981).

Review period[a] (number of review periods in sample)	Return computation method[b]		
	Buy-and-hold (BH)	Arithmetic (AR)	Daily rebalancing (RB)
	AMEX–NYSE mean return differential (% per annum)[c]		
Daily (4767)	14.9 (3.16) [7.76]	14.9 (3.16) [7.76]	14.9 (3.16) [7.76]
Bi-daily (2389)	12.3 (2.64) [5.58]	14.9 (3.16) [7.06]	14.8 (3.15) [7.01]
Weekly (992)	9.81 (2.16) [3.35]	14.8 (3.15) [5.64]	14.7 (3.14) [5.62]
Bi-weekly (498)	8.27 (1.84) [2.46]	14.9 (3.14) [5.09]	14.7 (3.13) [5.07]
Monthly (228)	7.06 (1.58) [1.82]	14.9 (3.14) [4.40]	14.7 (3.11) [4.38]
Quarterly (76)	6.42 (1.43) [1.67]	15.0 (3.15) [3.88]	14.8 (3.12) [3.85]
Annual (19)	7.45 (1.53) [1.53]	15.1 (3.10) [3.10]	14.9 (3.07) [3.07]

[a]For the daily and bi-daily cases, one- and two-trading-day intervals were used respectively. For all other cases, actual calendar intervals were used. (In the weekly and bi-weekly cases, a residual interval was necessary to fill out each calendar year). All returns were compounded to an annual basis by linking successive observations within each year (see footnote 2 of the text).

[b]The computation method follows expressions (1), (2) and (3) of the text. For interested readers, the author will gladly supply a mimeographed sheet containing details on the treatment of delisting and listing securities. The main feature of the treatment of new listings and delistings was to assure that all three mean return methods employed exactly the same sample observations.

[c]t-statistics based on the 19 annual (linked) observations are in parentheses; t-statistics based on the review period returns as independent observations are given in brackets. To understand the difference in the two reported t-statistics, consider the example of the daily review period of which there are 4767 in the sample. The t-statistic in brackets is calculated from these 4767 (daily) observations (mean *daily* return divided by standard error of mean daily return). The t-statistic in brackets is calculated from 19 annual observations; each annual observation having been calculated by linking together approximately 250 (4767/19) daily observations observed during that year. In calculating the review-period-based t-statistics for the weekly and bi-weekly cases, ten days were omitted; these ten days were the reminders of partial weeks at year end. It turned out that in 10 years of the 19, the year was exactly 52 weeks plus one trading day long. An earlier version of the paper, available on request, details the effect of omitting these single-day partial weeks. N.B. This is an issue only for the bracketed t-statistics. The linked annual returns include *every* sample day.

statistics are also given based on review period returns taken as independent observations.[9]

The results most like actual investment experience are those in the first column, buy-and-hold returns. Most actual portfolios pursue a buy-and-hold strategy within a given review period with only minor modifications induced by new information about particular individual issues. The results are frequently expressed on an annual percentage basis by comparing wealth levels at the ends of successive years, i.e., after linking sub-year results.

The review period seems to have little effect on the AR and RB means. The annual average difference in returns between AMEX and NYSE issues is about fifteen percent. But for the BH means, the review period has a large impact. Monthly and longer review periods give an AMEX–NYSE return differential of only around seven percent (and the t-statistic does not indicate an overwhelming probability that the differential is even positive). The drop in the BH mean with lengthening review period is statistically significant and so is the difference between the BH and the other means.[10]

---

[9]Note that the t-statistics in these tables are based on the assumption that the annual returns (t-statistics in parentheses) and review period returns (t-statistics in brackets) are temporally independent. The results indicate that the AR and RB returns are, in fact, close to independent while there is negative serial dependence in the BH returns. This implies that the t-statistics for the BH means are actually *understated*.

[10]A statistical test of the significance of the review period can be conducted by considering each year's mean difference, AMEX–NYSE, as an independent observation. Let $D_{m,y,\tau}$ be the difference for year y, review period τ, and the method m ($m = $ BH, AR, RB). Then the time series mean of $D_{m,y,\tau} - D_{m,y,\tau'}$ ($\tau \neq \tau'$) can be tested for significance under the presumption that the years constitute independent observations. t-statistics for the AR and RB means, for all combinations of τ and τ', never indicated significance. Of the 42 combinations (21 for each mean AR and RB) none exceeded 2.0, five exceeded 1.5, and 28 were less than 1.0. In contrast, the t-statistics for the BH mean comparisons across review periods are given below:

Review period τ'	Review period τ					
	Daily	Bi-daily	Weekly	Bi-weekly	Monthly	Quarterly
Bi-daily	6.21					
Weekly	6.75	6.82				
Bi-weekly	7.67	8.37	10.8			
Monthly	8.11	8.89	11.3	9.82		
Quarterly	8.10	7.68	8.65	6.49	3.27	
Annual	5.08	4.42	2.81	1.04	−0.532	−1.67

All BH means are significantly different across-review periods except the annual mean versus the bi-weekly, monthly and quarterly means. Note that these table entries are not statistically independent of one another (they were all calculated from the same underlying data).

A similar procedure can be employed to test the statistical significance of mean computational method. The difference $D_{m,y,\tau} - D_{m',y,\tau}$ ($m \neq m'$) forms another time series across years. Based on 19 annual observations, t-statistics for the significance of this difference from zero are as follows:

Given that the BH results in table 1 are most likely to portray actual investment experience, we now turn to the interesting econometric question: What explains the observed pattern of means? To aid in answering this question, the mean returns for each exchange are presented separately in table 2. Notice that the pattern is not predicted by the expected values of the mean returns derived in section 2.2 under the assumption of temporally independent returns. With serial independence, the BH expected mean should be greater than the RB expected mean. The empirical results in table 2 show, however, that serial dependence must be present since $\bar{R}_{BH}$ falls below $\bar{R}_{RB}$ as the review period lengthens.

The arithmetic (AR) mean is larger than the rebalanced (RB) mean as was expected with or without serial dependence. However, these two means are very close and this suggests that serial dependence in *portfolio* returns is not much of an influence [Cf. eq. (14)]. Indeed, the strikingly different behavior of the BH means from the other two means indicates that negative serial dependence in individual securities is the dominant influence on the results.

In order to be certain that the AMEX–NYSE comparison measures the small firm effect properly, table 3 is presented. It contains results for the annual review period and for portfolios classified directly by size. Firm size was calculated as market capitalization (market price times number shares), at the end of each year, 1962–1980. Firms were assigned to fractiles based on market capitalization and their returns were calculated for the following year according to three mean return methods, BH, AR, and RB.

Not surprisingly, the results are consistent with the AMEX corresponding to lower size quintiles and the NYSE to higher quintiles. The overall implication is identical: viz., the estimated small firm premium is much smaller and less significant when mean returns are computed with the buy-

Review period τ	$m = $ AR, $m' = $ BH	$m = $ RB, $m' = $ BH	$m = $ AR, $m' = $ RB
	t-statistic for difference		
Bi-daily	6.82	6.30	1.47
Weekly	7.33	6.80	1.59
Bi-weekly	8.14	7.59	1.74
Monthly	8.44	7.90	2.17
Quarterly	8.21	7.69	2.72
Annual	5.85	5.48	3.16

No statistic was computed in the daily case because all three means are identical by construction in that case. Notice that the BH means are significantly smaller than the other two means for all review periods.

Although the difference between the AR and RB small firm premium is very small (cf. table 1), the AR mean premium is always larger and is significantly larger for monthly, quarterly and annual review periods. This is predicted by eq. (14); the AR mean grows with review period relative to the RB mean.

Table 2
Mean returns on NYSE and AMEX listed securities, 1963–1981.[a]

Review period	Buy-and-hold (BH)		Arithmetic (AR)		Daily rebalancing (RB)	
	NYSE	AMEX	NYSE	AMEX	NYSE	AMEX
	Mean returns (% per Annum)					
Daily	17.24	32.09	17.24	32.09	17.24	32.09
	(2.94)	(3.29)	(2.94)	(3.29)	(2.94)	(3.29)
	[5.09]	[7.72]	[5.09]	[7.72]	[5.09]	[7.72]
Bi-daily	16.93	29.23	17.53	32.42	17.24	32.09
	(2.89)	(3.03)	(2.98)	(3.31)	(2.94)	(3.29)
	[4.59]	[6.25]	[4.76]	[6.96]	[4.68]	[6.88]
Weekly	16.38	26.19	17.79	32.61	17.26	31.99
	(2.80)	(2.78)	(3.02)	(3.34)	(2.94)	(3.28)
	[4.47]	[5.32]	[4.81]	[6.44]	[4.68]	[6.32]
Bi-weekly	15.86	24.14	17.95	32.83	17.29	32.08
	(2.72)	(2.58)	(3.05)	(3.36)	(2.95)	(3.28)
	[4.29]	[4.66]	[4.71]	[5.85]	[4.58]	[5.74]
Monthly	15.34	22.39	18.07	32.96	17.34	32.08
	(2.65)	(2.42)	(3.07)	(3.36)	(2.95)	(3.28)
	[3.11]	[3.08]	[3.67]	[4.54]	[3.51]	[4.41]
Quarterly	15.01	21.42	18.17	33.17	17.38	32.19
	(2.63)	(2.33)	(3.09)	(3.38)	(2.96)	(3.29)
	[2.73]	[2.62]	[3.22]	[3.84]	[3.09]	[3.73]
Annual	15.18	22.63	17.96	33.07	17.16	32.03
	(2.69)	(2.39)	(3.11)	(3.36)	(2.98)	(3.27)
	[2.69]	[2.39]	[3.11]	[3.36]	[2.98]	[3.27]

[a]See footnotes to table 1.

and-hold method than when means are computed with the AR and RB methods.

3.2. Implications for previous research and for the 'risk-adjusted' small firm premium

The implications of these findings for previously-published estimates of the small firm premium are: if the basic data were very short-term and arithmetic or rebalanced means were used, the estimated premium overstates the reward investors can expect from a buy-and-hold position in small firms. Papers by Reinganum (1981a, b, 1982) and Roll (1981) used daily data and arithmetic mean returns. Reinganum's (1982) paper gives monthly and quarterly returns but these were computed with the daily rebalancing method since the author states that '... these holding period returns are created by compounding the daily *portfolio* returns' (p. 34, emphasis added).

Table 3

Mean returns and small firm premia for portfolios classified by size[a] at year-end, 1963-1981, annual review period.

Size quintile	Return computation method[b]		
	Buy-and-hold (BH)	Arithmetic (AR)	Daily rebalancing (RB)
	Mean return (% per annum)[c]		
Smallest	27.9 (2.42)	46.0 (3.68)	44.9 (3.61)
2	21.1 (2.51)	27.6 (3.15)	26.6 (3.04)
3	17.1 (2.41)	20.7 (2.86)	19.7 (2.73)
4	14.6 (2.53)	16.9 (2.89)	16.1 (2.75)
Largest	10.8 (2.50)	12.2 (2.85)	11.5 (2.68)
Small firm premium, smallest–largest quintile (% per annum)			
	17.1 (1.88)	33.9 (3.47)	33.4 (3.46)
Small firm premium, smallest–largest decile (% per annum)			
	22.8 (2.07)	49.1 (3.84)	48.3 (3.83)

[a]Firms are included in the kth size fractile if the closing price times the number of outstanding shares is ranked in that fractile among all listed AMEX and NYSE firms.

[b]The computation method follows expressions (1), (2) and (3) of the text. An unpublished appendix (available from the author) contains details on the treatment of listing and delisting.

[c]t-statistics based on 19 annual observations are in parentheses.

Papers with monthly returns are apparently much less subject to mean return estimation problems. Tables 1 and 2 show that there is little additional discrepancy between the BH and other means in going from monthly to annual data. The well-known paper by Banz (1981) used monthly data as did earlier papers on the closely-related stock price effect [Blume and Husic (1973), Bachrach and Galai (1979)]. Thus, it seems unlikely that the results presented in those papers will be much affected by the problem investigated here. In a more recent paper, Reinganum, (1983) used the buy-and-hold method and found results close to those reported above. Reinganum did not, however, contrast the buy-and-hold with other mean returns.

It is important to ascertain whether the *risk-adjusted* small firm premium is attributable solely to econometric problems. Is underestimation of risk for small firms [Roll (1981), Reinganum (1982)], combined with overestimation of expected returns, sufficient to induce the observed risk-adjusted premium; or is the premium really evidence of a misspecified capital asset pricing model (CAPM), perhaps because of omitted factors in the single index CAPM?

This is tantamount to asking whether the implicit CAPM market risk premium $\hat{p}$ ($\hat{p} \equiv \hat{E}(R_{small} - R_{large})/(\hat{\beta}_{small} - \hat{\beta}_{large})$), is in a reasonable range. $\hat{p}$ was computed by Reinganum (1983) as 37.5 percent per annum using (a) buy-and-hold means on the smallest and largest deciles of NYSE and AMEX stocks, (b) Dimson's (1979) aggregated coefficient betas, (c) the value-weighted C.R.S.P. index and (d) daily data for 1963–1980. The return on the value-weighted index during this period was only about 9.5 percent, so $\hat{p}$ is grossly too large, thereby indicating a substantial risk-adjusted small firm premium.

The main problem with such a test was described some time ago [Roll (1977)]. Even if we make the dubious assumption that the value-weighted C.R.S.P. index is ex-ante mean/variance efficient, there is no necessity in the generalized Black (1972) C.A.P.M. that $E(\hat{p}) = E(R_M - R_F)$. Instead, the model requires that $E(\hat{p}) = E(R_M - R_Z)$ where Z is M's 'zero-beta' portfolio. Depending upon M's position on the efficient frontier, $E(R_Z)$ can be negative and large.

To illustrate the difference in inferences that can be obtained with a different index, I recomputed $\hat{p}$ using (a) buy-and-hold annual means on the smallest and largest deciles of NYSE and AMEX stocks, (b) simple OLS beta coefficients estimated from annual returns,[11] (c) the *equally*-weighted C.R.S.P. index, and (d) annual data for 1963–1981.

The beta estimates (t-statistics) were $\beta_{small} = 1.78$ (5.59), $\beta_{large} = 0.598$ (8.60). Using the estimated premium $E(R_{small} - R_{large}) = 22.8\%$ from table 3, we have $\hat{p} = 19.3$ percent. The actual ex post return on this market index was 15.3 percent, so $\hat{p}$ is still somewhat too high (thus indicating a risk-adjusted small-firm premium). Nevertheless, the discrepancy between a $\hat{p}$ of 19.3 and a market return of 15.3 is much less aberrant than the difference Reinganum (1983) reports between $\hat{p} = 37.5$ and $\bar{R}_M = 9.5$ percent.

It still seems that investigation of the observed small firm premium in the context of a more general asset pricing model would be a worthwhile endeavor; but estimation problems in expected returns and in simple risk parameters can explain much of the apparent anomaly.

---

[11]Instead of the Dimson aggregated coefficient betas, I used betas from annual data because of the now well-documented annual seasonal [Keim (1983), Roll (1983)], which has the potential to induce biases into any betas, including the Dimson type, when they are computed from non-yearly data.

5. Conclusion

Computing mean returns in order to estimate investment experience is not as easy as it sounds. Common stock data have serial dependence which, though seemingly slight, substantially affects the estimates obtained under alternate mean return computational methods. Investment experience is best portrayed by buy-and-hold portfolio returns but scholars often use arithmetic or rebalanced portfolio returns because they are easier to compute.

Perhaps this makes little difference for some studies; but if serial dependence differs systematically with the item being investigated, the computational method can be quite material.

For the small firm premium, as measured by the difference in mean returns of American Exchange and New York Exchange listed stocks, the buy-and-hold mean return difference is only about $7\frac{1}{2}$ percent per annum (for 1963–81) while the rebalanced and arithmetic methods produce annual return differences *with the same stocks and time periods* of over 14 percent. The annual difference in returns between the smallest and largest size quintiles (deciles) is about 34 (49.1) percent using the rebalanced and arithmetic methods and about 17 (22.8) percent using the buy-and-hold method.

The annual small-firm premium is only marginally significant at usual significance levels if mean returns are measured with the buy-and-hold method.

References

Bachrach. Benjamin and Dan Galai, 1979, The risk return relationship and stock prices, Journal of Financial and Quantitative Analysis 14, June, 421–441.

Banz, Rolf W., 1981, The relationship between return and market value of common stocks, Journal of Financial Economics 9, March, 3–18.

Black, Fischer, 1982, Capital market equilibrium with restricted borrowing, Journal of Business 45, July, 444–454.

Blume, Marshall, 1974, Unbiased estimators of long-run expected rates of return, Journal of the American Statistical Association 69, Sept., 634–638.

Blume, Marshall and Frank Husic, 1973, Price, beta, and exchange listing, Journal of Finance 28, May, 283–299.

Blume, Marshall and Robert F. Stambaugh, 1983, Biases in computed returns: An application to the size effect, Journal of Financial Economics, this issue.

Cheng, Pao L. and M. King Deets, 1971, Statistical biases and security rates of return, Journal of Financial and Quantitative Analysis 6, June, 977–994.

Cohen, Kalman T., Steven F. Maier, Robert A. Schwartz and David K. Whitcomb, 1979, On the existence of serial correlation in an efficient securities market, TIMS Studies in the Management Sciences 11, 151–168.

Dimson, Elroy, 1979, Risk measurement when shares are subject to infrequent trading, Journal of Financial Economics 7, June, 197–226.

Keim, Donald, 1983, Size related anomalies and stock return seasonality: Further empirical evidence, Journal of Financial Economics 12, June, 13–32.

Niederhoffer, Victor and M.F.M. Osborne, 1966, Market making and reversal on the stock exchange, Journal of the American Statistical Association 61, Dec., 898–916.

Reinganum, Marc R., 1981, The arbitrage pricing theory: Some empirical results, Journal of Finance 36, May, 313–321.

Reinganum, Marc R., 1981, Misspecification of capital asset pricing: Empirical anomalies based on earnings yields and market values, Journal of Financial Economics 9, March, 19–46.

Reinganum, Marc R., 1982, A direct test of Roll's conjecture on the firm size effect, Journal of Finance 37, March, 27–36.

Reinganum, Marc R., 1983, Portfolio strategies based on market capitalization, Journal of Portfolio Management 9, Winter, 29–36.

Roll, Richard, 1977, A critique of the asset pricing theory's tests, Journal of Financial Economics 4, March, 129–176.

Roll, Richard, 1981, A possible explanation of the small firm effect, Journal of Finance 36, Sept., 879–888.

Roll, Richard, 1983, Vas ist Das? The turn of the year effect and the return premia of small firms, Journal of Portfolio Management 9, Winter, 18–28.

Scholes, Myron and Joseph Williams, 1977, Estimating betas from nonsynchronous data, Journal of Financial Economics 5, Dec., 309–327.