UNBIASED ESTIMATORS OF LONG RUN EXPECTED RATES OF RETURN

by

Marshall E. Blume*

Working Paper #19-72b

RODNEY L. WHITE CENTER FOR FINANCIAL RESEARCH

The Wharton School

University of Pennsylvania

Philadelphia, PA 19104

Abstract

This paper documents the biases in using sample arithmetic or geometric means of one-period returns to assess long run expected rates of return. The formulae developed are applicable to other compound growth processes. For types of distributions of one period returns likely to be encountered for bonds and stocks, numerical values for these biases are given. Then four unbiased estimators of long run expected rates of return are developed and their relative efficiency examined.

I. Introduction

In a variety of financial decisions, an individual or firm must assess the long run expected rates of return of some investment vehicle. As one example, a professor whose institution invests in TIAA/CREF on his behalf would certainly try to assess the magnitude of his retirement fund in determining his current schedule of savings. As another example, an actuary in calculating premiums for a life insurance policy would need to make some assumption about long run expected rates of return. Such persons as these would typically base their assessments of future expected rates of return upon past experience.

Assume, for instance, that this past experience consists of T monthly relatives, defined as the ratio of the value at the end of the month to the value at the end of the previous month. Now, assume that one wishes to determine the expected increase in value of this asset if it were to be held N months, where this increase is measured by the ratio of the terminal value to the initial value -- a so-called N-period relative. If it can be assumed that the relatives in each single period approximate identically distributed independent normal variates, the expected N-period relative is given by the population expected one-period relative to the N^{th} power.

In practice, one does not know the population statistic and therefore must make an estimate. Some might be tempted to estimate the expected N-period relative by raising the arithmetic average of the T one-period relatives to the N^{th} power. As long as N exceeds one, this procedure will yield an upward biased estimate. Others would take the geometric mean of the T observations and raise this number to the N^{th} power to derive an estimate of the expected N-period relative. This estimate is downward biased if N is less than T.

The paper develops formulae for the magnitude of these biases which, when evaluated at reasonable values for the stock market, show that the biases are sometimes substantial. More generally, these formulae can be used to calculate their magnitude for any compound process.

An unbiased estimate of the expected N-period relative for $N<T$ will therefore be between the arithmetic mean raised to the N^{th} power and the geometric mean raised to the N^{th} power.[1] Finally, this paper will propose and evaluate various unbiased estimators of the expected N-period relative for data like those found in the bond and stock markets.

II. The Bias in the Arithmetic Mean

Let R_t represent an one-period relative or one plus the interest rate. Further,assume that R_t is an independent, normally distributed random variate with positive μ and non-zero $\sigma(R)$ -- stationary over time. It is convenient to define a new random variable ε_t as

2.1 $$R_t = \mu + \varepsilon_t.$$

The random variable ε_t is thus independently and normally distributed with mean zero and a standard deviation the same as R_t.

The expected N-period relative, denoted by $E(W_N)$, is given by

2.2 $$E(W_N) = E\left(\prod_{t=1}^{N} R_t\right)$$

$$= E\left[\prod_{t=1}^{N} (\mu + \varepsilon_t)\right]$$

Because of independence, 2.2 becomes

2.3 $$E(W_N) = \mu^N$$

Equation 2.3 shows that the population expected N-period relative is the population expected one-period relative raised to the N^{th} power.

From a sample of T observations, R_t, $t = 1, \ldots, T$, an unbiased estimate of the expected one-period return is

2.4 $$A = \mu + \left(\sum_{t=1}^{T} \varepsilon_t\right)/T,$$

where A denotes the arithmetic mean of one-period relatives. Raising 2.4 to the N^{th} power in the spirit of 2.3 and letting

2.5 $$h = \left(\sum_{t=1}^{T} \varepsilon_t\right)/T,$$

one obtains the following estimate of $E(W_N)$:

2.6 $$A^N = (\mu + h)^N$$

It follows directly from 2.6 that the estimator A^N is asymptotically unbiased and consistent. Since h is an average of normally distributed and independent random variables, h is itself a normal variate. As T approaches infinity for fixed values of N, the variance of h will approach zero and therefore the probability limit of A^N is μ^N.

Although A^N is asymptotically unbiased and consistent, it is upward biased for finite T and N greater than one. Applying expected value operators to 2.6 yields

2.7 $$E(A^N) = E[(\mu + h)^N]$$

Jensen's inequality shows that the term on the right is equal to or greater than μ^N, so that the arithmetic estimate is upward biased.

To measure the magnitude of the bias, $E(A^N)$ was evaluated[2] for values of μ from 1.00 to 1.01 and $\sigma(R)$ from 0.03 to 0.15. The values assigned to N and T ranged up to 100. These ranges are roughly the ranges one might encounter in empirical work with monthly relatives for bonds and common stocks. A comparison of the estimated expected N-period relatives with the corresponding population statistic discloses that the biases are in many cases substantial. For instance, for $E(R) = 1.01$ and $\sigma(R) = 0.15$, the expected 40-period relative estimated from 80 observations is 1.8416 compared to the population statistic of 1.4888.

III. The Bias in the Geometric Mean

From a sample of T observations, the sample geometric mean is calculated as

3.1 $$G = [\prod_{t=1}^{T} (R_t)]^{1/T}$$

where G denotes geometric mean and where it is now assumed that every R_t exceeds zero. An estimate of the expected N-period relative is given by

3.2 $$G^N = [\prod_{t=1}^{T} (R_t)]^{N/T}$$

If every R_t must exceed zero, R_t cannot be normally distributed as was assumed in the previous section. Nonetheless, for monthly relatives of stocks or bonds in which μ will be somewhat greater than 1.0 and $\sigma(R_t)$ around 0.15 or less, the distribution of R_t may closely approximate a normal distribution.[3]

The estimate of the expected N-period relative given by the geometric mean in 3.2 is downward biased when N is less than T. The demonstration of this bias follows from first substituting 2.1 into 3.2, which gives 3.3:

3.3 $$G^N = [\prod_{t=1}^{T} (\mu + \varepsilon_t)]^{N/T}$$

Defining Y as $[\prod_{t=1}^{T} (\mu + \varepsilon_t)] - \mu^T$ and taking expected values, 3.3 becomes

3.4 $$E(G^N) = E\,[(\mu^T + Y)^{N/T}]$$

Since $\mu + \varepsilon_t$ must be assumed positive for the use of the geometric mean, since a positive variable raised to the N/T power is a concave function for N less than T, and since E(Y) equals zero, the following inequality holds

3.5 $$E(G^N) < (\mu^T)^{N/T} = \mu^N$$

providing at least one ε_t is non-zero.

If N equals T, the geometric mean raised to the N^{th} power provides an unbiased estimator, but this is not surprising in that this estimator is merely one drawing from the distribution of N-period relatives. A single drawing from a distribution is of course an unbiased estimator of the mean. If N is greater than T, the estimator of $E(W_N)$ provided by the geometric mean is biased upwards. Further, the estimate provided by the geometric mean is not consistent.[4]

To measure the magnitude of the bias in the geometric mean, $E[G^N]$ was evaluated numerically[5] for the same values of N, T, E(R), and $\sigma(R)$ as for the arithmetic mean. These biases are sometimes substantial. For instance, for E(R) = 1.01 and $\sigma(R)$ = 0.15, the expected 40-period relative estimated from 80 observations is 1.1880 compared to the population statistic of 1.4888. It may be recalled that the corresponding estimate provided by the arithmetic mean was 1.8416.

The analytical and the numerical results of this section and the previous one show that estimators of the expected N-period relative derived either from arithmetic means or geometric means of T observations may be substantially biased for distributions of relatives for common stocks and bonds. More specifically for N less than T and N greater than one -- a case of importance for empirical work, the arithmetic estimate of $E(W_N)$ will be upward biased while the geometric estimate will be downward biased. Thus, an unbiased estimate of $E(W_N)$ will be between the arithmetic and geometric estimates. The remainder of this paper explores methods of obtaining unbiased estimates of $E(W_N)$.

Before proceeding, it may be worthwhile to record an explicit comparison for the case in which N equals one since a large number of empirical studies of stock market returns are based upon this case.[6] The arithmetic mean provides an unbiased and consistent estimate of the expected one-period relative, while the geometric mean provides a biased and inconsistent estimate. Further, a formula to be developed in the next section can be used to show that the geometric mean has a larger sample variance than the arithmetic mean.[7] It therefore appears that if one can assume that the relatives, R_t, are distributed by independent, stationary, normal distributions, the arithmetic mean provides a superior estimate of the expected one-period relative compared to that provided by the geometric mean.

IV. Unbiased Estimates

This section proposes four different methods of obtaining unbiased estimates of the expected N-period relative for N less than T.[8] The next section uses Monte Carlo techniques to obtain an insight into the distributional properties of these unbiased estimators as well as the generally biased estimators provided by the arithmetic and geometric means discussed above.

The first type of estimator will be dubbed the "simple unbiased" estimator. This estimator is appropriate where the number of observations in the sample, T, is an integral multiple of the number of periods, N, for which the expected relative is calculated. To calculate this estimate, multiply the first N relatives together, the second N relatives, and so on until the T one-period relatives are exhausted. Then, average these products or N-period relatives, T/N in number, to obtain an unbiased estimate of the expected N-period relative. The reader should note that this procedure makes no assumptions about the independence of the distributions of the one-period relatives.

The second type of estimator, discussed in [5], will be called the "overlapped unbiased" estimator. This estimator proceeds by calculating N-period relatives, T-N+1 in number, by multiplying the first through the N^{th} one-period relatives together, the second through the $(N + 1)^{st}$ one-period relatives together, and so on. These overlapped relatives are then averaged to obtain an unbiased estimate. Intuitively, some investigators might anticipate that this estimator would be more efficient than the previous one in that it incorporates somehow more information. Nonetheless, it is easy to construct a counter example which shows that it may be less efficient.[9] Indeed, the Monte Carlo simulations in the next section will show for data likely to be observed in the stock market that the "overlapped unbiased" estimator is probably less efficient than the "simple unbiased" estimator.

The third type of estimator will be termed a "weighted unbiased" estimator because it is calculated as a weighted average of the biased estimators provided by the arithmetic and geometric means. Using formulae developed earlier in the paper, the footnote [10] shows that an approximately unbiased estimator of $E(W_N)$ is given by the weighted average:

4.1 $$\hat{E}(W_N) \simeq \frac{T-N}{T-1} A^N + \frac{N-1}{T-1} G^N$$

The coefficients of A^A and of G^N in 4.1 sum to one and can be used to form a weighted average of the estimates of μ provided by the arithmetic and geometric means.[11] These weights which are functions of T and N make intuitive sense. When N equals 1, all the weight is given to the arithmetic mean. When N equals T, all the weight is given to the geometric mean. As

N drops from T, more and more weight is given to the arithmetic mean and less to the geometric mean. Since the arithmetic mean is consistent while the geometric mean is not, the weighting is sensible.

The fourth type of estimator adjusts A^N with an appropriate adjustment factor. This estimator will be termed the "adjusted unbiased" estimator. For monthly data from the bond or stock markets, μ is likely to be in the interval from 1.00 to 1.01 while $\sigma(R)$ may range as high as 0.15. For these ranges of μ and $\sigma(R)$ and for $N \leq 80$, $T \leq 100$, and $N \leq T$, the following regression which does not include μ as an independent variable fits the bias calculations extremely well:[12]

$$4.2 \qquad \ln \left\{ \frac{[E(A^N)]^{1/N} - \mu}{\mu} \right\} = -0.9174 + 1.9958 \ln \sigma(R) + 1.0441 \ln N - 0.9989 \ln T, \quad \bar{R}^2 = .9990$$

The values of $[E(A^N)]^{1/N}$, implicitly defined by 4.2, differ in absolute values from their true values by 1.1 percent on average and by 4.2 percent at most.[13] Using the sample estimate to measure $\sigma(R)$, 4.2 implies for any particular N and T a value of the ratio of $E(A^N)$ to μ^N. Dividing this ratio into A^N should yield an approximately unbiased estimator of $E(W_N)$. The next section will examine the degree of approximation introduced by using a sample value of $\sigma(R)$ instead of the population value.

V. A Monte Carlo Analysis

To examine the empirical properties of these various estimators of the expected N-period relative, 80,000 randomly distributed unit normal variates were calculated using the procedure found in [14]. These variates, reexpressed so as to have appropriate values of E(R) and σ(R), were partitioned sequentially into 1000 separate samples of 80 observations to correspond to a T of 80.

For each sample, A^N, G^N, and the four unbiased estimators discussed above were obtained for N equals 5, 10, 20, and 40. The adjusted unbiased estimator was calculated using both the estimated and population values of σ(R). Although in any application only the estimated value could be used, a comparison of the two estimates will indicate the magnitude of the error introduced by using an estimate rather than the population value.[14]

Table 1 gives descriptive statistics of the distributions of the various estimates for two cases: (a) μ = 1.00 and σ(R) = 0.03 and (b) μ = 1.01 and σ(R) = 0.15. A comparison of the arithmetic and geometric means of T one-period relatives raised to the N^{th} power to the population statistic, μ^N, show that the estimates are biased in the anticipated directions. The simple unbiased estimate, as well as the overlapped estimate,are very close to the population statistic as would be expected. The averages for the three remaining estimators show that any errors introduced into the estimates becuase of the approximations used in deriving the formulae do not create any substantial biases. Further, a comparison of the two estimates provided by the adjusted unbiased estimator show that little error is introduced in using an estimate of σ(R) instead of the population value in calculating the adjustment factor.

The figures in Table 1 additionally suggest that the overlapped unbiased estimator is markedly less efficient (say as measured by the standard

TABLE 1

Sample Distributions of Estimators of $E(V_N)$ for T = 80

N	Estimators	μ = 1.00, σ(R) = 0.03 Average	Standard Error	Fractiles 0.05	0.50	0.95	μ = 1.01, σ(R) = 0.15 Average	Standard Error	Fractiles 0.05	0.50	0.95
5		1.0000					1.0510				
	ARITHMETIC	1.0005	0.0193	0.9734	1.0008	1.0275	1.0564	0.0871	0.9181	1.0553	1.2005
	GEOMETRIC	0.9983	0.0191	0.9712	0.9986	1.0250	0.9991	0.0847	0.8618	0.9983	1.1424
	SIMPLE UNBIASED	1.0005	0.0193	0.9733	1.0008	1.0277	1.0545	0.0893	0.9149	1.0523	1.2038
	OVERLAPPED UNBIASED	1.0005	0.0192	0.9722	1.0007	1.0292	1.0544	0.0898	0.9094	1.0524	1.2078
	WEIGHTED UNBIASED	1.0004	0.0191	0.9733	1.0007	1.0274	1.0535	0.0869	0.9160	1.0528	1.1970
	ADJ. UNB. (EST S(R))	1.0004	0.0192	0.9733	1.0007	1.0274	1.0532	0.0868	0.9159	1.0526	1.1964
	ADJ. UNB. (POP S(R))	1.0004	0.0192	0.9732	1.0007	1.0274	1.0532	0.0868	0.9153	1.0521	1.1968
10		1.0000					1.1046				
	ARITHMETIC	1.0014	0.0345	0.9474	1.0016	1.0558	1.1237	0.1844	0.8430	1.1137	1.4411
	GEOMETRIC	0.9970	0.0343	0.9432	0.9973	1.0507	1.0054	0.1699	0.7427	0.9966	1.3050
	SIMPLE UNBIASED	1.0011	0.0346	0.9465	1.0009	1.0560	1.1138	0.1929	0.8257	1.1043	1.4581
	OVERLAPPED UNBIASED	1.0009	0.0357	0.9438	1.0012	1.0584	1.1115	0.1962	0.8132	1.1020	1.4401
	WEIGHTED UNBIASED	1.0009	0.0345	0.9470	1.0013	1.0552	1.1102	0.1827	0.8331	1.1023	1.4227
	ADJ. UNB. (EST S(R))	1.0009	0.0345	0.9470	1.0013	1.0552	1.1096	0.1821	0.8335	1.1016	1.4210
	ADJ. UNB. (POP S(R))	1.0009	0.0344	0.9469	1.0011	1.0553	1.1096	0.1821	0.8324	1.0997	1.4231
20		1.0000					1.2202				
	ARITHMETIC	1.0041	0.0672	0.8976	1.0033	1.1147	1.2966	0.4303	0.7106	1.2402	2.0769
	GEOMETRIC	0.9952	0.0667	0.8896	0.9946	1.1040	1.0397	0.3553	0.5516	0.9933	1.7030
	SIMPLE UNBIASED	1.0022	0.0673	0.8956	1.0009	1.1156	1.2411	0.4508	0.6344	1.1813	2.0488
	OVERLAPPED UNBIASED	1.0016	0.0736	0.8863	1.0010	1.1239	1.2332	0.4802	0.6136	1.1519	2.0897
	WEIGHTED UNBIASED	1.0019	0.0671	0.8960	1.0015	1.1121	1.2348	0.4119	0.6743	1.1805	1.9732
	ADJ. UNB. (EST S(R))	1.0020	0.0670	0.8961	1.0015	1.1121	1.2313	0.4087	0.6747	1.1786	1.9664
	ADJ. UNB. (POP S(R))	1.0020	0.0671	0.8957	1.0012	1.1124	1.2310	0.4085	0.6747	1.1775	1.9718
40		1.0000					1.4888				
	ARITHMETIC	1.0127	0.1347	0.8057	1.0066	1.2425	1.8661	1.3247	0.5050	1.5382	4.3135
	GEOMETRIC	0.9949	0.1324	0.7915	0.9892	1.2188	1.2070	0.9818	0.3042	0.9866	2.9001
	SIMPLE UNBIASED	1.0040	0.1342	0.7964	0.9999	1.2305	1.5147	1.2007	0.3528	1.2255	3.6189
	OVERLAPPED UNBIASED	1.0040	0.1567	0.7691	0.9929	1.2720	1.5282	1.4104	0.2860	1.1472	3.8228
	WEIGHTED UNBIASED	1.0039	0.1336	0.7994	0.9992	1.2306	1.5407	1.1045	0.4115	1.2691	3.5893
	ADJ. UNB. (EST S(R))	1.0040	0.1335	0.7997	0.9994	1.2307	1.5093	1.0699	0.4054	1.2464	3.5213
	ADJ. UNB. (POP S(R))	1.0040	0.1335	0.7988	0.9980	1.2319	1.5070	1.0698	0.4078	1.2422	3.4836

deviation or the 0.05 and 0.95 fractiles) than the other unbiased estimators. In addition, the simple unbiased estimator appears somewhat less efficient than both the weighted unbiased estimator and the adjusted unbiased estimator, where the adjustment factor is estimated with the sample value of $\sigma(R)$. Finally, the reader may note that the sample distributions are skewed to the right with the skewness more pronounced for the case in which $E(R) = 1.010$ and $\sigma(R) = 0.150$.

VI. Conclusion

The theoretical and empirical results of this paper suggest that one should proceed very cautiously in using arithmetic or geometric means of one-period relatives to assess the expected N-period relatives. More explicitly, an estimate of the expected N-period relative derived by raising either of these statistics to the N^{th} power would usually be biased.

If one can assume that the one-period relatives are distributed by an independent normal process, the paper has shown for data like that which might be encountered in the stock or bond markets an average of overlapped data may be much less efficient than merely a simple average of non-overlapped data. The paper then went on to suggest two non-linear methods of assessing unbiased estimates which appear somewhat more efficient than a simple average of N-period relatives: (1) a weighted unbiased estimator and (2) an adjusted unbiased estimator. Although there is little difference in efficiency between the weighted unbiased estimator and the adjusted estimator, the weighted unbiased estimator is probably safer to use than the other. One could easily visualize types of departures from stationary independent normal distributions which might lead to absurd estimates from the adjusted unbiased estimator but not from the weighted unbiased estimator. Yet, if one cannot assume independence of successive one-period relatives of if there is even a slight chance that these relatives are dependent, the simple average of N-period relatives would appear preferable to the non-linear estimators which even under ideal conditions yield only a modest increase in efficiency.

Footnotes

*Associate Professor of Finance at the University of Pennsylvania. The author wishes to thank the Rodney L. White Center for Financial Research of the Wharton School for financial support, as well as Professors Stephen Ross and Randolph Westerfield whose helpful comments were greatly appreciated.

[1]The author has been familiar for some time with the biases in using the geometric mean to estimated expected one-period or long run rates of return. Research on a recent paper, co-authored with Irwin Friend, [8] suggested that the arithmetic mean may yield substantially biased estimates of expected long run rates of return. This paper is in part an outgrowth of this research.

Lawrence Fisher [7] observed in empirical data the biases associated with the use of one-period geometric and arithmetic means to estimate long run rates of return. He however did not give a full explanation of this phenomenon.

Cheng and Deets [5] have previously shown that the geometric mean raised to the N^{th} power is downward biased for N less than T and upward biased for N greater than T. They however did not calculate the magnitude of the bias. In the process of deriving a formula for such a calculation and in extending the theoretical results to a comparison of the relative efficiencies of arithmetic and geometric estimates of one-period expected returns, this paper will again demonstrate these biases but in a more concise way.

[2]The computational formula for $E(A^N)$ was derived by first expanding 2.6 with the Binomial expansion and taking expected values. Noting that h is normally distributed, all odd moments in the expansion can be set to zero and $E(h^i)$ for even i replaced by $\frac{i!}{2^{i/2}(\frac{i}{2})!}(\sigma^2(h))^{i/2}$. After changing the index of summation and setting σ^2 (h) to $\sigma^2(\varepsilon)/T$, the resulting formula is

$$E(A^N) = \mu^N + \sum_{i=1}^{n} \binom{N}{2i} \mu^{N-2i} \frac{2i!}{2^i\ i!} \left(\frac{\sigma^2\ (\varepsilon)}{T}\right)$$

where n is the largest integer equal to or less than (N/2).

[3]Some investigators (e.g., [6], [11]) prefer log normal distributions on the assumption that the distribution of R_t would be skewed to the right. However, [3] finds no evidence of asymmetry in the distributions of R_t for

monthly data. In fact for monthly data, there may well be no distinguishable empirical differences whether $\ell n\ R_t$ or R_t is used. For longer periods, asymmetry will become more pronounced (cf. [1]).

[4]This statment is based upon the following: Taking the probability limit of 3.2, one obtains

$$\text{plim } G^N = \exp \ \{\text{plim } \frac{N}{T} \sum_{t=1}^{T} \log R_t\}$$

The term in the braces is N E [log R] , which is less than N log μ for non-degenerate distributions since the logarthmic function is concave. Thus

$$\text{plim } G^N < \exp\ (N \log \mu\ ) = \mu^N$$

Of course, for fixed N, taking the probability limit implies that $T > N$.

[5]A formula to calculate these values was obtained by taking expected values of 3.3 and rearranging terms to yield

$$E\ [G^N\ ] = \mu^N\ E\left[\prod_{t=1}^{T} (1 + \frac{\varepsilon_t}{\mu})^{\frac{N}{T}}\right]$$

Using the exponential function, the above becomes

$$E[G^N] = \mu^N\ E\left\{\prod_{t=1}^{T} \exp\left[\frac{N}{T} \log\ (1 + \frac{\varepsilon_t}{\mu})\right]\right\}$$

On the assumption that ε_t/μ is less than one in absolute value, the logarithmic function can be expanded in an infinite series as

$$E[G^N] = \mu^N\ E\left\{\prod_{t=1}^{T} \exp\left[\frac{N}{T} \sum_{j=1}^{\infty} \frac{(-1)^{j+1}}{j} (\frac{\varepsilon_t}{\mu})^j\right]\right\}$$

The expansion of the exponential function yields

$$E\ [G^N] = \mu^N\ E\left\{\prod_{t=1}^{T} \left[\sum_{i=o}^{\infty} \left\{\frac{N}{T} \sum_{j=1}^{\infty} \frac{(-1)^{j+1}}{j} \left(\frac{\varepsilon_t}{\mu}\right)^j\right\}^i \Big/ i!\right]\right\}$$

Noting that ε_k and ε_l, $k \neq l$; are independent and the formula for the i^{th} moment about the mean for a normal distribution, one can calculate the desired numbers to any degree of accuracy by tedious but straightforward numerical calculations for any specific values of the parameters.

[6]A representative sample of such articles includes [2], [4], [6], and [10]. Each of these articles contain bibliographies which point to a large number of other articles.

[7]This statement about efficiency is based upon a lengthy algebraic calculation for which an outline follows: Remove the expected value operators from the last equation in footnote 5 and set N equal to one. Dropping from the expansion of this equation all terms of degree greater than two and all terms involving cross products of ε_k and ε_l, $k \neq l$, one obtains

$$G \approx \mu \left\{ 1 + \frac{1}{T} \sum_{t=1}^{T} \left[\frac{\varepsilon_t}{\mu} - \frac{(T-1)\,\varepsilon_t^{\,2}}{2T\,\mu^2} \right] \right\}$$

Subtracting E(G) given by the expression in footnote 10 from the above, squaring, and taking expected values, one obtains

$$E\left\{G - E(G)\right\}^2$$

$$\approx \frac{\sigma^2(\varepsilon_t)}{T} + \frac{T-1}{2T\,\mu^2}\; E\left\{\left[\left(\sum_{t=1}^{T} \varepsilon_t^{\,2}\right) - \sigma^2(\varepsilon)\right]^2\right\}$$

Since the first term on the right is the variance for the arithmetic mean, the geometric mean should have a somewhat larger variance than the arithmetic mean.

[8]The techniques are designed for N greater than one although the first three reduce to the arithmetic mean when N equals one.

[9]A counter example using four one-period relatives R_t, $t=1, \ldots, 4$, to estimate the expected two-period relative is as follows: Since the R_t's are independent and stationary, the variance of the simple unbiased estimator will be $\sigma^2(R_1R_2)/2$, while the variance of the overlapped unbiased estimator is $\sigma^2(R_1R_2)/3 + [4 \text{ Cov } (R_1R_2, R_2R_3)/9]$. Assuming $E(R_t)$ equals 1.0, $\sigma^2(R_1R_2)$ can be rewritten as $[E(R_1^{\,2})]^2-1$ and cov (R_1R_2, R_2R_3) as $E(R_1^{\,2})-1$. It is easily verified that for $E(R_1^{\,2})$ greater than 1.0 and less than 5/3, the variance of the simple unbiased estimator is less than the variance of the overlapped estimator. For example, if $E(R_1^{\,2}) = 4/3$, the variance of the simple average will be 21/54 compared to 22/54 for the overlapped average. It might be noted that Cheng and Deets [5] showed that the expected sample variance of an average calculated from a single drawing of two overlapped two-period relatives

is smaller than that calculated from two non-overlapped two-period relatives. Though correct, this observation about expected sample variances has no obvious implications for the efficiency of these two estimators which should be judged by a comparison of the population variances of the estimators and not by the variances calculated from single samples. The situation they analyzed is similar to a regression with autocorrelated residuals, for which it is well known that the expected standard error calculated from a single sample is downward biased.

[10] To derive this weighted average, the expected value of the estimate of the N-period relative can be approximated from the formula in footnote 1 by dropping all terms involving moments of greater than the second order. The resulting approximation is

$$E\,[A^N] \approx \mu^N + \frac{N(N-1)\sigma^2(\varepsilon)\mu^{N-2}}{2T}$$

$$= \mu^N \left[1 + (N-1)\frac{N\,\sigma^2(\varepsilon)}{2T\,\mu^2}\right]$$

For the geometric mean, a similar type of approximation can be derived from the last equation in footnote 5 by expanding it and then by dropping terms involving moments of ε greater than the second order. The resulting approximation is

$$E\,[G^N] \approx \mu^N \left[1 - \frac{N\sigma^2(\varepsilon)}{2\mu^2} + \frac{N^2\sigma^2(\varepsilon)}{2T\,\mu^2}\right] = \mu^N\left[1 - (T - N)\frac{N\sigma^2(\varepsilon)}{2T\mu^2}\right]$$

Solving one of these equations for $[N\sigma^2(\varepsilon)]/[2T\mu^2]$ and substituting the resulting expression into the other, one obtains by solving for μ^N the basis for the expression in the text. Although this development involves substituting approximations into approximations -- a treacherous procedure, the absolute bias in the weighted estimator for $1 < N < T$ will always be less than the absolute maximum of the biases in A^N or G^N.

[11] This technique is not the same as employed in [7].

[12] This regression was fitted for E(R) equal to 1.000, 1.005, and 1.010, for $\sigma(R)$ equal to 0.030, 0.060, 0.100, and 0.150 and for N running from 10 to 80 and from 10 to 100 in increments of 10.

[13] Including in μ as an explanatory variable increased the value of $\bar{R}^2$ to 0.99994.

[14] Similarly, estimates of $E(W_N)$ were obtained for N equals 5, 10, and 20 and T equals 40 from the first half of each of the 1000 samples. For reasons of space, these are not presented. They, however, give substantially the same conclusions.

REFERENCES

[1] Arditti, Fred D., "Risk and the Required Return on Equity," Journal of Finance, 22 (March 1967), 19-36.

[2] Black, Fisher; Jensen, Michael and Scholes, Myron, "The Capital Asset Pricing Model: Some Empirical Tests," in Michael C. Jensen, ed., Studies in the Theory of Capital Markets, New York: Praeger Publishing, 1973.

[3] Blume, Marshall, "Portfolio Theory: A Step Towards Its Application," Journal of Business, 43 (April 1970), 152-173.

[4] ________ and Friend, Irwin, "A New Look at the Capital Asset Pricing Model," Journal of Finance, 28 (March 1973), 19-33.

[5] ________ and ________, "Risk, Investment Strategy and the Long-Run Rates of Return," The Review of Economics and Statistics, forthcoming.

[6] Cheng, Pao L. and Deets, M. King, "Statistical Biases and Security Rates of Return," Journal of Financial and Quantitative Analysis, 6 (June 1971), 977-994.

[7] Fama, Eugene, et al., "Adjustment of Stock Prices to New Information," International Economic Review, 10 (February 1969), 1-21.

[8] Fisher, Lawrence, "Some New Stock Market Indexes," Journal of Business, 39 (January 1966), 191-225.

[9] Hoel, Paul G., Introduction to Mathematical Statistics, New York: John Wiley and Sons, Inc., 1958.

[10] Jensen, Michael, "Risk, the Pricing of Capital Assets, and the Evaluation of Investment Portfolios," Journal of Business, 42 (April 1969), 167-247.

[11] Osborne, M. F. M., "Brownian Motion in the Stock Market," Operations Research, 7 (March-April 1959), 145-173.

[12] Rie, Daniel, "Single Parameter Risk Measures and Multiple Sources of Risk: A Reexamination of the Data Based on Changes in Determinants of Price Over Time," Rodney L. White Center for Financial Research, Working Paper No. 14-72.

[13] Samuelson, Paul, "Lifetime Portfolio Selection by Dynamic Stochastic Programming," The Review of Economics and Statistics, 51 (August 1969), 239-246.

[14] "System/360 Scientific Subroutine Package: Version III: Programmer's Manual," International Business Machines Corporation (August 1970).