Generalized Mean-Variance Tradeoffs for Best Perturbation Corrections to Approximate Portfolio Decisions

Paul A. Samuelson; Robert C. Merton


Stable URL:
http://links.jstor.org/sici?sici=0022-1082%28197403%2929%3A1%3C27%3AGMTFBP%3E2.0.CO%3B2-A


Your use of the JSTOR archive indicates your acceptance of JSTOR’s Terms and Conditions of Use, available at http://www.jstor.org/about/terms.html. JSTOR’s Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at http://www.jstor.org/journals/afina.html.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

JSTOR is an independent not-for-profit organization dedicated to creating and preserving a digital archive of scholarly journals. For more information regarding JSTOR, please contact support@jstor.org.
GENERALIZED MEAN-VARIANCE TRADEOFFS FOR BEST PERTURBATION CORRECTIONS TO APPROXIMATE PORTFOLIO DECISIONS*

PAUL A. SAMUELSON  
ROBERT C. MERTON

For investors with utilities "near" to logarithmic utility of terminal wealth, it is shown how use of the pair of parameters \( E(\log W) \) and \( \text{Variance}(\log W) \) can form an efficiency tradeoff frontier useful locally for people of different risk-return tolerances. Exact perturbation approximations are given in power series, and this two-parameter-tradeoff device is generalized to investors sharing the same general degree of risk tolerance, even if it is not near to \( \log W \). Standard Markowitz mean-variance is seen to be one such special case. Another case is that of Hakansson's average-compound returns, whose expected value and variance—absurd and irrelevant as they are to those with utilities of terminal wealth not close to \( \log W \)—can be used as asymptotic surrogates for our locally-sufficient pair of parameters.

I. Solutions to Portfolio Decision Problem

A. Local Perturbation Technique

Exact solutions to the standard portfolio decision problem involves

\[
\max_{\{w_i\}} \int_0^\infty U[W, \Sigma_i w_i Z_j] P(dZ_1, \ldots, dZ_n) = U[w^*_2, \ldots, w^*_n]
\]

where \( U[ \ ] \) is a strictly-concave function of terminal wealth, \( W \), \( P(Z_1, \ldots, Z_n) \) is the joint probability distribution of the non-negative random variables that depict each terminal outcome in dollars from one dollar invested in the \( j \)-th security or program, \( E[ \ ] \) is the operator for mathematical expectation, \( W_o \) is given initial wealth, \( (w_{o,1}, \ldots, w_n) \) are fractions of wealth invested in each security, and \( w_i = 1 - \sum_j w_j \) is the amount invested in an available safe security.

Although general properties of the exact solution are known, its computation can be lengthy. Therefore, approximate solutions are legitimate objects of research.

Mistakenly, some have proposed replacing the general problem by \( U[W] = \log W \), in the fallacious belief that such a maximizing strategy will give a good approximation to the true solution (or perhaps even "better" it) when there are many periods of independent reinvesting. See Williams [12], Latane

* This paper is a spin-off of a longer study [8] by the authors dealing with fallacious log-normal approximations in portfolio decision making. We owe thanks to the National Science Foundation for financial support.
[5], Kelly [4], Aucamp [1], Breiman [2], and their critique by Samuelson [10].

Some others have been attracted by the criterion of maximum expected average compound return discussed by Hakansson [3], in which the general $U[W]$ can be replaced by $U[W] = W^{1/N}$ or by $W^{a/N}$, where $N$ is an integer greater than or equal to two (and usually much greater) and $a \leq 1$, $a \neq 0$. (The case of $a = 0$ can be assimilated under the log $W$ strategy already mentioned.)

However, it should be obvious that anyone whose true $U[W]$ is far away from $W^{r/\epsilon}$, as for example if it were $2W_{1/2} - W^{-1}$, $(1 + W)/(2 + W)$, or $W^{r}/\gamma$ for $|\gamma| > 0$, will be badly served by the average compound return criterion—no matter how numerous the periods of his independent reinvestings! And supplementing its mean by its variance will not, in the general situation, remedy the failure of such a “maximum growth” policy.

Fallacies are fallacies. And they must be recognized as such. Nevertheless, we can ask the following question that will be of limited interest in some special cases.

Suppose there are two or more people, facing the same investment options, each with $U$ functions near to $W^{r/\epsilon}$ where $|\epsilon|$ is a small number. Thus, man B has $W^{1/10}/(1/9)$, man C has $-W^{-1/100}/(1/100)$, . . . . If by luck, some man A has solved the problem for log $W$, which is clearly “near” the other men’s solutions, we should be able to develop a perturbation technique whereby supplementing A’s $E\{\log W\}$ by Variance $\{\log W\}$ will give men B and C close approximations to their exact solutions.

How close is “close?” For any man we can write the optimal portfolio proportions as

$$w_{1}(\epsilon) = w_{1}(0) + \epsilon w_{1}(0) + O(\epsilon^2), \quad (i = 2, \ldots, n).$$

Then $w_{1}(0)$ is the max $E\{\log W\}$ solution. And the best estimate for the correction-term coefficient $w_{1}(\epsilon)$ is calculable from $\text{Var}\{\log W\}$.

This means that we can validly generalize the classical absolute-mean-variance analysis of Markowitz [7], which works with arithmetic (rather than logarithmic) values $E\{W\}$ and $\text{Var}\{W\}$ and which must be viewed with caution when probabilities are not “compact” in the sense of Samuelson [9]. Now, we prepare once and for all a local efficiency tradeoff frontier between $\mu = E\{\log W\}$ and $\sigma^2 = \text{Var}\{\log W\}$, which can be used by all men like B and C to find their own best “risk-return” point to a high degree of approximation.

Warning: $\mu$ and $\sigma^2$ are “asymptotically sufficient” parameters for the problem only in the non-general case where $\epsilon$ in $W^{r/\epsilon}$ is small enough so that the exact first correction in (2) is not swamped by the neglected remainder in the perturbation expansion.

If, however, $\epsilon$’s being small is a legitimate empirical hypothesis, then there are also other equivalent ways of arriving at the $w_{1}(0)$ correction term. Thus, Hakansson’s frontier of $E\{\text{average compound return}\}$ and $\text{Var}\{\text{average compound return}\}$ can equivalently be employed for this purpose—even by one who (like us) regards preoccupation with average compound return as gratuitous or fallacious.
B. Sketch of Derivation

In this preliminary report, we shall only heuristically present the argument. Replace \( W^*/\epsilon \) by its equivalent for maximization purposes

\[
U[\epsilon;W] = (W^* - 1)/\epsilon
\]

\[
\lim_{\epsilon \to 0} U[\epsilon;W] = \log W
\]

\[
\lim_{\epsilon \to 0} \frac{\partial U[\epsilon;W]}{\partial \epsilon} = \frac{1}{2} \log^2 W.
\]

Then, by Taylor's expansion,

\[
U[\epsilon;W] = \log W + \frac{1}{2} \epsilon \log^2 W + O(\epsilon^2)
\]

and

\[
E\{U[\epsilon;W]\} = \mu + \frac{1}{2} \epsilon (\sigma^2 + \mu^2) + O(\epsilon^2)
\]

\[
= u[\epsilon;w_2(\epsilon), \ldots, w_n(\epsilon)]
\]

\[
= u[\epsilon;w] \text{ for short.}
\]

Then,

\[
\text{Max } u[\epsilon;w] = u[\epsilon;w^* (\epsilon)]
\]

is defined by the solution of the implicit equations satisfying the first-order maximum conditions.

\[
0 = u_w[\epsilon;w^*] = u_w[0;w^*(0)] + \epsilon u_{ww}[0;w^*(0)]w^{*'}(0) + O(\epsilon^2),
\]

where \( u_w[\epsilon, w^*] \) stands for the row vector \( [\partial u[\epsilon, w^*]/\partial w_2, \ldots, \partial u[\epsilon, w^*]/\partial w_n] \) and \( u_{ww}[\epsilon, w^*] \) stands for the Hessian matrix of \( u[\epsilon, w^*] \)'s second partial derivatives, and similarly for the column vectors \( w^*(0) \) and \( w^{*'}(0) \).

We calculate recursively new terms such as \( w^{*'}(0) \) from relations of the type

\[
u_w[0,w^*(0)] = 0.
\]

\[
u_{ww}[0,w^*(0)]w^{*'}(0) = -\partial u[0,w^*(0)]/\partial \epsilon.
\]

For higher order terms, such as \( w^{*''}(0) \) or \( w^{*(n+1)}(0) \), we have similar linear relations

\[
u_{ww}[0,w^*(0)]w^{*''}(0) = \text{a function of the partials of } u[0,w^*(0)]
\]

\[\text{and of } w^*(0) \text{ and } w^{*'}(0).\]

\[
u_{ww}[0,w^*(0)]w^{*(n+1)}(0) = \text{a function of the partials of } u[0,w^*(0)]
\]

\[\text{and of } w^*(0), w^{*'}(0), \ldots, w^{*(n-1)}(0), w^{*(n)}(0).\]

So long as \( \epsilon \) is small, maximizing expected utility can be done neglecting terms in \( \epsilon^2 \) and higher powers, i.e., the exact optimal solution \( \{w^*(\epsilon)\} \) will satisfy (7) which can be rewritten as
when \( U[\varepsilon; W] \) is as defined in (4), while the (approximate) optimal solution \( \{w^*(\varepsilon)\} \) obtained by maximizing \( [\mu + \frac{1}{2}\varepsilon(\sigma^2 + \mu^2)] \) will satisfy

\[
0 = \frac{\partial \mu(w^*)}{\partial w_i} + \frac{1}{2}\varepsilon(\partial \sigma^2/\partial w_i + 2\mu \partial \mu/\partial w_i), \quad i = 2, \ldots, N. \tag{10}
\]

However, maximization of \( [\mu + \frac{1}{2}\varepsilon(\sigma^2 + \mu^2)] \) is not the only way of obtaining a valid approximation to \( \{w^*(\varepsilon)\} \). Namely, if \( w^{tf}(\varepsilon) \) is the solution to

\[
\text{Max}[\mu + \frac{1}{2}\varepsilon \sigma^2], \tag{11}
\]

then it will satisfy

\[
0 = \frac{\partial \mu [w^{tf}(\varepsilon)]}{\partial w_i} + \frac{1}{2}\varepsilon \partial \sigma^2/\partial w_i, \quad i = 2, \ldots, n. \tag{12}
\]

But, (10) can be rewritten as

\[
0 = \frac{\partial \mu(w^t)}{\partial w_i} + \frac{\varepsilon}{2(1 + \mu \varepsilon)} \partial \sigma^2/\partial w_i, \quad i = 2, \ldots, n \tag{13}
\]

Therefore, \( w^t_i(\varepsilon) - w^{tf}_i(\varepsilon) = O(\varepsilon^2) \) and \( w^*_i(\varepsilon) - w^{tf}_i(\varepsilon) = O(\varepsilon^2) \), \( i = 2, \ldots, n \). Relation (11) as a criterion of interest because it is a linear logarithmic mean-variance tradeoff and because all feasible solutions to (11) can be generated by an efficient frontier in logarithmic mean-variance space, i.e., man B and man C will each find their (approximate) optimal solutions on the frontier defined by

\[
\text{Max} \mu(w_2, \ldots, w_n) \tag{14}
\]

s.t. \( \sigma^2(w_2, \ldots, w_n) = \text{constant} \).

Men with negative \( \varepsilon \) will wish to be slightly less risk-taking than man A with his maximum-growth (i.e., max \( \mu \) alone) solution. Those with positive \( \varepsilon \) will want to be a bit less cautious—i.e., both branches of the frontier around A are definitely efficient.

C. **Graphical Depiction**

Figure 1 shows a rough heuristic sketch of the \((\mu, \sigma^2)\) efficiency tradeoff frontier. As noted, large \( \sigma^2 \) as well as small provide efficient points. Men like B will want to be on the declining branch. A*, B*, C* show the exact best strategies for the three men. B and C are our perturbation approximations, differing from true B* and C* by terms of \( O(\varepsilon^2) \). The tangency at A* of the exact curve to the tradeoff frontier constitutes the essence of our perturbation theorem.

Figure 2 shows the story in the strategy space. The equilateral triangle provides a barycentric diagram: the sum of the distances from any point in

1. \( \text{Max} \mu = \text{Max} E[\log W] \) exists and is finite. Hence, \( \mu \varepsilon = O(\varepsilon) \).
the triangle to the three sides is constant, taken to be unity. Then in a three-security case—say $w_1$ for fraction in cash, and $w_2$ and $w_3$ for fractions in two risky securities—the distance from the base (side 1) and the other two sides (sides 2 and 3) respectively depict $w_1 = 1 - w_2 - w_3$, $w_2$, $w_3$. The locus C A*B represents Figure 1's local approximations. Again tangency at A* epitomizes our power-series first approximations. (Note that the $w^*_1(0)$ +
ew*(1/0) approximation, which is very close to our C A*B locus, is not quite identical to it. But, of course, up to O(e^2) there is no significant difference.

D. Higher-Order Conditions

Remark: The same optimal solution, w*(\epsilon), is derived from any monotone stretching of u](\epsilon;w] such as g(\epsilon;w) = G(u[\epsilon;w],e) with \partial G/\partial u > 0. Note that g(\epsilon;w] will generally not have exactly the same coefficients in its power-series expansion in powers of \epsilon (even though the solution to (7) and (8) for it must yield the same coefficients for [w*(0),w*(0), . . . ]). This will explain why it happens that both Max[\mu + 1/2\epsilon(\sigma^2 + \mu^2)] and the different expression Max[\mu + 2\epsilon\sigma^2] in (11) can each yield the correct first-order approximating strategies—namely the same correct [w*(0),w*(0), . . . ] coefficients.²

Warning: Only in a near neighborhood of A’s maximum will e^2 terms be ignorable. To bring in a new second-order correction term w*′′(1/0), we obviously need the third moment of skewness of logW, \mu_3, since (5) becomes

\[ U[\epsilon;W] = \log W + \frac{1}{2} \epsilon \log W + \frac{1}{12} \epsilon^2 \log W + O(\epsilon^3). \]  

(15)

Now we need a tradeoff frontier in the three dimensions of logarithmic mean, variance, and skewness. And so it goes for fourth moment corrections to make the remainder O(\epsilon^4). Every new power in the expansion involves a moment of one higher order, in analogy to Samuelson [9].

II. Example

To illustrate that w*(1), w*(1/0), and w*+̃(1) agree to 0(\epsilon^2), we consider the simple case of one risky and one riskless asset used in Merton and Samuelson [8]. The return per dollar on the riskless asset is R and the return per dollar on the risky asset is denoted by the Bernoulli-distributed random variable y where \text{Prob}(y = \lambda) = \text{Prob}(y = \delta) = \frac{1}{2} and \lambda > R > \delta > 0. The optimal portfolio proportion in the risky asset was derived in [8] to be

\[ w*(\epsilon) = (A - 1)R/[(\lambda - R) + A(R - \delta)] \]  

(16)

where

\[ A \equiv [(R - \delta)/(\lambda - R)]. \]

Further, it was shown in [8] that the optimal solution to (11), w*+̃(\epsilon), satisfies

\[ 0 = \left( \frac{\lambda - R}{R - \delta} \right) \left[ 1 + \frac{\epsilon}{2} \log B \right] - B \left[ 1 - \frac{\epsilon}{2} \log B \right] \]  

(17)

2. Actually, whenever a(w) + eb(w) + O(\epsilon^2) is replaced by a(w) + e[a(w),b(w)] + O(\epsilon^2), we must get the same first-order conditions for w*(0), since \{\delta[.] / \partial a\} \{\delta a(w) / \delta w\} will vanish by virtue of the second factor's vanishing. So the presence or absence of a term like \mu^2 in \frac{1}{2}\epsilon(\sigma^2 + \mu^2) makes no formal difference. Of course, the proper w*′′(0) cannot be found from \mu + \frac{1}{2}\epsilon\sigma^2 if the proper 0(\epsilon^3) terms are not taken into account.
where
\[ B(\epsilon) = \frac{w^\dagger(\epsilon)(\lambda - R) + R}{w^\dagger(\epsilon)(\delta - R) + R}. \]

From (16), \[ w^*(0) = R(\lambda + \delta - 2R)/(2(R - \delta)(\lambda - R)), \]
and \( w^\dagger(0) = w^*(0). \) Since from (4), \( E\{U[0,W]\} = E\{\log W\} = \mu. \)
Differentiating (16), we have
\[ w^*(0) = \log \left[ \frac{\lambda - R}{R - \delta} \right] \frac{\lambda - \delta}{2(\lambda + \delta - 2R)} w^*(0). \]
(18)

Applying the implicit function theorem, we have
\[ \frac{w^{\dagger\dagger}(0)}{w^\dagger(0)} = \frac{B(0)}{B(0) - 1} \frac{B'(0)(R - \delta)}{[(\lambda - R) + B(0)(R - \delta)]} = B'(0) \left[ \frac{1}{(\lambda + \delta - 2R)} - \frac{1}{2(\lambda - R)} \right](R - \delta), \]
(19)
and from (17),
\[ B'(0) = \left( \frac{\lambda - R}{R - \delta} \right) \log \frac{\lambda - R}{R - \delta}. \]
(20)

Combining (19) and (20), we have
\[ w^{\dagger\dagger}(0) = \frac{(\lambda - \delta)}{2(\lambda + \delta - 2R)} \log \left[ \frac{\lambda - R}{R - \delta} \right] w^\dagger(0), \]
(21)
and using the result that \( w^*(0) = w^\dagger(0) \) with (18) and (21), we have \( w^{\dagger\dagger}(0) = w^\star'(0). \) Hence,
\[ w^*(\epsilon) - w^\dagger(\epsilon) = w^*(0) + \epsilon w^\star'(0) + O(\epsilon^2) \]
\[ - w^\dagger(0) - \epsilon w^{\dagger\dagger}(0) - O(\epsilon^2) = O(\epsilon^2), \]
(22)
and \( w^\dagger(\epsilon) \) is a valid approximation to \( w^*(\epsilon) \) to \( O(\epsilon^2) \).

To complete the illustration, we now show that \( w^\dagger(\epsilon) \) is an equally good approximation. \( w^\dagger(\epsilon) \) satisfies
\[ \text{Max} \left[ \mu + \frac{\epsilon}{2}(\sigma^2 + \mu^2) \right] = \text{Max} \left\{ \frac{1}{2} \log[w(\lambda - R) + R] \right. \]
\[ + \frac{1}{2} \log[w(\delta - R) + R] \]
\[ + \frac{\epsilon}{4} \left( \log^2[w(\lambda - R) + R] + \log^2[w(\delta - R) + R] \right) \right\}. \]
(23)
The first-order condition satisfied by \( w^\dagger(\epsilon) \) is
\[ 0 = \frac{\lambda - R}{w^\dagger(\lambda - R) + R} + \frac{\delta - R}{w^\dagger(\delta - R) + R} + \epsilon \frac{(\lambda - R)\log[w^\dagger(\lambda - R) + R]}{w^\dagger(\lambda - R) + R} \]
\[ + \frac{\epsilon}{w^\dagger(\delta - R) + R}. \]
(24)
Clearly, \( w^t(0) = w^t(0) = w(0) \). Using this fact and the implicit function theorem, we have

\[
\left( \frac{dw^t}{d\varepsilon} \right)_{\varepsilon=0} = \frac{C(\lambda - R)[w^t(\delta - R) + R] \log[C]}{(\lambda - R)^2 + C^2(\delta - R)^2} \tag{25}
\]

where \( C = [w^t(\lambda - R) + R]/[w^t(\delta - R) + R] \). Using \( w^t(0) = w^*(0) = (\lambda + \delta - 2R)R/[2(R - \delta)(\lambda - R)] \) and hence that \( C(0) = (\lambda - R)/(R - \delta) \), we simplify (25) to

\[
w^t(0) = \frac{R(\lambda - \delta)}{4(\lambda - R)(R - \delta)} \log \left( \frac{\lambda - R}{R - \delta} \right) \tag{26}
\]

Comparing (26) with (21) and (18), we have \( w^t(0) = w^t(0) = w^*(0) \). Hence, \( w^t(\varepsilon) \) and \( w^t(\varepsilon) \) are equally good approximations to \( w^*(\varepsilon) \) with the error in either case of \( O(\varepsilon^2) \).

### III. Generalization

Lest anyone think that there is something important about the neighborhood in which the \( \gamma \) in \( W^\gamma/\gamma \) is small, we could write down equivalent perturbations around any \( \gamma \) neighborhood, as, e.g., \( \gamma = -3.14 \).

Thus, in general

\[
W^{\gamma + \varepsilon}\gamma = W^\gamma \left[ 1 + \varepsilon \log W + O(\varepsilon^2) \right], \tag{27}
\]

and our asymptotically sufficient pair of parameters become \( E(W^\gamma/\gamma) \) and \( E(W^\gamma \log W/\gamma) \).

Even more generally, we need not consider constant-relative-risk averters with \( U \) functions in the class \( W^\gamma/\gamma \). All we need is

\[
U[\varepsilon;W] = U[0;W] + \varepsilon W[0;W]/\partial \varepsilon + O(\varepsilon^2) \tag{28}
\]

and the pair of parameters, \( E\{U[0;W]\} \) and \( E\{\partial U[0;W]/\partial \varepsilon\} \).

From this general viewpoint, we derive the original Markowitz (or for that matter Marschak [6] and Tobin [11]) mean-variance analysis by writing

\[
U[\varepsilon;W] = W - \varepsilon bW^2 + O(\varepsilon^2) \tag{29}
\]

However, for this approximation to be justified \( U \) must almost be linear while our earlier approximations required "near" log-linear. Hence, the earlier approximations are generally more relevant for concave utility maximizers. Warning: the approximation in (28) should not be confused with the justification of the use of mean-variance for "compact probabilities" which involves expansions like

\[
U[\varepsilon;W] = \varepsilon(W - bW^2) + O(\varepsilon^2). \tag{30}
\]
IV. Digression on Average Compound Return

We may briefly relate the Hakansson criterion of expected average compound return to our perturbation analysis. In the case of identical probability distributions in each period, serially independent of each other, the utility functions $W^\gamma/\gamma$ will all give rise to "myopic" and uniform strategies. In that case, the Hakansson criterion (except for an additive constant) is defined as the $1/e$ power of $E\{1 + \varepsilon U[\varepsilon; W]\}$ where $\varepsilon = 1/N$ and $N$ is the number of periods.

Hence, it becomes

$$h[\varepsilon; w] = (1 + \varepsilon u[\varepsilon; w])^{1/\varepsilon} = \exp\left[\mu + \frac{1}{2} \varepsilon \sigma^2 + 0(\varepsilon^2)\right] = \varepsilon^\mu \left[1 + \frac{1}{2} \varepsilon \sigma^2 + 0(\varepsilon^2)\right].$$

(31)

Note that $h[\varepsilon; w]$, $u[\varepsilon; w]$, and $g[\varepsilon; w] = \log h[\varepsilon; w] = \mu + \frac{1}{2} \varepsilon \sigma^2 + 0(\varepsilon^2)$ have first-order coefficients different from one another.

Contemplating $g[\varepsilon; w]$, we can give an affirmative answer to Hakansson’s open question of whether, for $N$ large and for $\varepsilon$ small, his maximizing $\mu + \frac{1}{2} \varepsilon \sigma^2$ yields a good approximation to his efficiency frontier. Note that no reference is made in our derivation to the treacherous assumption of a log-normal surrogate for the probability distribution of $W/W_o$ (i.e., of $W_N/W_o$).\(^3\)

Up until now we have had no need to make explicit use of the variance of expected average compound return. Again with no recourse to log-normal surrogates, we calculate this as

$$V[\varepsilon; w] = [E(W^2 \varepsilon)]^{1/\varepsilon} - [E(W \varepsilon)]^{2/\varepsilon} = (1 + 2\varepsilon u[2\varepsilon; w])^{1/\varepsilon} - (1 + \varepsilon u[\varepsilon; w])^{2/\varepsilon} = \exp\left[2\mu + \varepsilon \sigma^2 + 0(\varepsilon^2)\right] \left(\exp\left[\sigma^2 \varepsilon + 0(\varepsilon^2)\right] - 1\right) = \varepsilon^{\sigma^2 \varepsilon} + 0(\varepsilon^2).$$

To keep this variance from vanishing (relative to the expected value) as $\varepsilon \to 0$, we work with the normalized variance $\eta[\varepsilon; w] = V[\varepsilon; w]/\varepsilon$. It is apparent that $\eta[0; w]$ can provide the added information needed for the first-degree $\varepsilon$ terms in $u[\varepsilon; w]$, and $w^*$ ( ). Hence, we can plot an efficiency frontier in the $(h, \eta)$ pair, maximizing $h[0; w]$ for given values of $\eta[0; w]$.

Remark: Although $1/N < 0$ makes no immediate sense, there is no reason why we should not work in $h[\varepsilon; w]$ with negative $\varepsilon$. What is useful about $h[\varepsilon; w]$ is not its average-compound-return interpretation, but rather that it is a legitimate surrogate for $u[\varepsilon; w]$.

Remark: In the usual indifference curve analysis of non-stochastic demand theory, when the utility indicator is stretched and renumbered, the utility

---

3. See Merton and Samuelson [8] for a discussion of the fallacies associated with such surrogates. Not even the Law of Large Numbers is involved in our derivations, since these are as valid for one-period as for many-period decision making.
contours themselves remain invariant. By contrast, in our diagram of first-order approximation, every stretching of the utility criterion—from \( u[ \cdot ] \) to \( \{ 1 + cu[ \cdot ] \}^{1/c} \) or \( \log(1 + eu[ \cdot ])/e \)—definitely alters the shapes of the indifference contours in the diagram. But, and this is crucial, the proper tangency points of local optimality are *invariants* under the stretchings!

Figure 3 shows how the two parameters, expected-value-of average compound-return-minus-one and its variance, here called respectively \( h_1[\varepsilon; w_2, \ldots] \) and \( h_2[\varepsilon; \ldots]/|\varepsilon| \).
. . ., wₙ] and h₂[ε; w₂, . . . , wₙ] in honor of Hakansson, can be used as a surrogate for our local-approximation technique even by those who have no use for the average-compound-return criterion for its own sake. The horizontal axis plots the normalized variance h₂[ε; w]/ε for reasons of convenience that will become apparent; the topology of the frontier is quite unchanged by this inessential monotone stretching. Actually, for 1/N small, the unnormalized variance shrinks to zero and would compress the diagram's scale to an unmanageable and inconvenient degree. (It also betrays people into thinking that there is something mandatory, or good, about pretending that men like B and C should forget that they have γ's a finite and irreducible distance away from A's γ = 0.) Further, as will be seen, with this normalization, the tradeoff frontiers approach a unique locus as 1/N → 0, one which depends only on the locally sufficient parameters μ(w) = \(E(\log W)\) and \(σ^2(w) = \text{Var}(\log W)\). More exactly, \(h_{1}[1/N; w] \rightarrow \exp[μ(w)],\) \(N h_{2}[1/N; w] \rightarrow \exp[2μ(w)]σ^2(w)\) as 1/N → 0.

Figure 3 shows several \((h_1, h_2/ε)\) loci for various values of 1/N. In every case, the points C*, A*, B* represent the exact \(w^*(ε)\) solutions corresponding to the same C*, A*, B* points of Figure 2 and Figure 1. In every case, the C, A, B points represent first-order approximations. Note that contrary to the mistaken notion that minimum variance of average compound returns (for a fixed mean) is a desirable thing, the declining branch of the frontier is efficient (for someone like B, for example).

Moreover, to debunk effectively the notion that expected average compound return is desirable for its own sake, Figure 3 plots an efficiency tradeoff locus for 1/N < 0. Indeed, man C can get a much more accurate approximation—indeed an exact evaluation—if we set 1/N = γ_c = −.01. Thus, C and C* coincide on the lowest curve shown, and both A and B are on the quite-admissible declining branch.

A. Confused Defense of Average Return in Terms of Ranking Good Guessers About Future Probabilities

The general criterion of average return has a certain plausibility about it. If in the last five years one mutual fund has had algebraic percentage gains, +10, +20, −10, 0, +25, and another has had +12, +21, −10, 0, +5, most people will prefer the first to the second since the first "dominates" the second in gains. Unconsciously, they will infer that those who run the first fund probably have better inductions about the probability distributions that we and Hakansson take as exactly (1) known. But, probably, when ranking a third fund that scores close to +9 per cent in each of the five years, they will prefer this last to either of the other two funds, even though on a five-year basis it has no gain advantage over the first fund. Ask them, "Why do you prefer, out of two funds with the same average compound return for the time horizon you say is crucial to you, the fund with the lower intra-horizon variance of annual return?" A little thought will suggest the answer that an actual sample of decision-makers will give: 'We naturally feel that the fund with the steady yield through thick and thin has an uncanny ability to come up with better guesses about the true [unknown] probabilities we are going
to be running up against in the future. It will likely be better at avoiding losers."

The economist will understand this line of reasoning. But he will realize that it would be a confusion for him to regard it as a valid reason for being interested in the expected average compound return as a maximand in an analytical model where the decision maker is working with prescribed probabilities. We must warn in the strongest terms against confusing the conditions of real life with those of the scores of analytical articles in the financial journals, such as Hakansson's 1971 paper or the present one.4

B. Comment on Hakansson

Since those who have not thought the matter through are prone to find the criterion of average compound return appealing—and indeed are prone to read into it implications far beyond those that Hakansson has ever claimed on its behalf—it is worth digressing further to evaluate its merits, to respond to Hakansson's [3, p. 880] "Comment on Samuelson" to see whether he has cogently rebutted the critique made of the average return concept and to explain wherein post-Hakansson decision-makers are self-deceived.

1. We begin with Hakansson's three-quarter-page comment on the Samuelson [10] critique. This begins with his mention that the median capital, after a large number of periods, N, will be "approximately equal to" (in our notation, for the case of independent, identical probabilities in each period) exp[μN].

How should someone who adheres to the Ramsey-Savage-Neumann axioms of expected utility react to this heuristic enunciation of one facet of the Central Limit Theorem? If his utility function is far from logW_T, he will receive this announcement with deservedly strong indifference. Let us hope that a new generation of thoughtless median maximizers does not arise on the scene!

2. Next Hakansson points out a facet of the Law of Large Numbers—namely, that average expected compound return converges in probability to exp[μN]. Since expμN increases with μ, the unwary will think that the sure-thing principle requires that he pick a higher μ in preference to a lower one. Yet, as our analysis has demonstrated, every rational maximizer of a U(W_T) far from logW_T will greet this convergence in probability result with deserved indifference.

Indeed, the sophisticated probabilist will say "It is precisely because average expected return is so badly normalized as to converge in probability to exp[μN] that it is a confusing and undesirable variable for decision makers to pay attention to." To Hakansson's assertion that μ provides an "intuitively appealing measure[s] of growth," a sophisticated utility maximizer will reply,

4. The reader may ask: "How does one go about reaching the real life problem of inductive inference?" Briefly the answer is "If the statistician believes the array of 10,000 eligible stocks can be regarded as subject to a joint log-normal (or other distribution), he must use Bayesian, maximum likelihood, or other methods for guessing at the future(1) parameters of that distribution. There is, in any case, no warrant for anyone who has as his utility U = W_0 - 5/S or U = W_T^{7/8}(7/8) to use U = logW_T or W_T^{1/8}/(1/N) in his final calculations with those separately estimated probabilities."
"Informed intuition realizes that alleged 'measures of growth' are misleading to a rational expected utility maximizer." This agreed to, Hakansson's further observation, that \( \sigma^2 \) clearly is an indicator of the 'smoothness' of this growth," loses all interest and relevance to portfolio optimizing in the face of given probabilities.

3. Then Hakansson repeats and agrees with the literal correctness of Samuelson's statement that, although for large enough N the twin who maximizes \( \mu \) will "almost certainly" end up with more terminal wealth than his brother who does anything else, that does not mean that the first twin will have as high expected terminal utility as will the twin who acts correctly to maximize the \( \mathbb{E}(U(W_T)) \) that they actually share as goals.

At this point Hakansson goes on to rule out risk tolerances greater than those of Bernoulli \( \log W_T \) maximizers. When Einstein objected to the probability basis for quantum mechanics with the aphorism "God doesn't run a gambling casino," Bohr answered him with the reminder, "Who are we to tell Him how to run His business?" When Hakansson, more modest than Einstein, argues that long-run investors will not want to have certain concave utility functions hitherto deemed admissible, his chain of reasoning deserves careful examination.

Hakansson says: "...there is good reason to believe that few individuals with long-run goals and possessing all the facts about (the implications of) various utility functions would, upon reflection, emerge with preferences consistent with utility functions other than (in our present notation) \( U(W_T) = -(W_T)^{-\beta} \) or \( U(W_T) = \log W_T \)."

Note the sweep of this amazing dictum: \( U(W_T) = W_T^{1/2} \), proposed by Cramer two centuries ago and deemed admissible by Bernoulli and Marshall, is now declared to be "on reflection" inadmissible. And what is the "good reason" offered to persuade one to renounce \( W_T^{1/2} \) or \( W_T^{0.0001} \)? It is Hakansson's belief that one should want, on reflection, to be on that branch of the Hakansson mean-variance frontier (recall our Figure 3) which he defines as the only efficient one.

What we have here is completely a case of petio principii: assuming what needs to be proved—that it is natural to want to minimize the variance of average compound return for the same expected value of that (irrelevant!) animal.

Hakansson is correct that contemplation of the St. Petersburgh Paradox can alert most people to the fact that their utility is strictly concave rather than linear. But it would be a delusion to think that his defense of minimum variance of average returns has equal force for reflective people.

We two are people. We try to be reflective. Yet we reject his criterion. More important, not one of the dozen reflective people we have discussed that matter with were willing, after they understood the intricacies of the issue, to end up agreeing with the contention. And one who has read all of Hakansson's valuable work will not want to bet that the author himself will want to reaffirm: "...the preceding analysis suggests that all risk-averse utility functions other than those with constant relative risk aversion (at least as
great as that of $\log W_T$ are implausible among (rational) long-run investors.”

It is true that casual and systematic empiricism suggests that many people, perhaps most, have risk-aversion more like the utility functions $-W^{-1}$ or $-W^{-2}$ than like $W^{1/2}$ or $W^{7/8}$. But that same body of limited data suggests, we would guess, that more have risk aversions like $-W^{-1}$ or $-W^{-2}$ than like $\log W$ or $-W^{-0.001}$. And such an alleged fact, we have shown in this paper, would, if true, be fatal to would be maximum-growth-cum-smooth-growth strategies. For, as we have shown, someone with $U = -W^{-2}$ will find efficiency frontiers in our $(\mu, \sigma^2)$ or in Hakansson’s $(h_1, h_2)$ space grossly irrelevant for good approximating to his true optimal strategies. Many of Hakansson’s casual readers, we have found, think that somehow $-W^{-2}$ gets validly converted into $-W^{-2/N}$ and hence for large $N$ is brought near to $\log W$ for decision-making purposes. We emphasize that this is a false interpretation of what is valid and useful in Hakansson’s work. We have other objections to average compound return arguments. But enough is enough.

V. Conclusion

The powerful approximation methods developed here are seen to have nothing to do with multiplicity of periods of decision-making. In principle, these expansions are as valid for one-period as for one-thousand-period decision-making. And they do not depend upon the different periods’ probability distributions being independent or even stationary.

REFERENCES