*Portfolio Choice and the Kelly Criterion**

Edward O. Thorp

UNIVERSITY OF CALIFORNIA AT IRVINE

I. Introduction

The Kelly (–Breiman–Bernoulli–Latané or capital growth) criterion is to maximize the expected value $E \log X$ of the logarithm of the random variable X, representing wealth. Logarithmic utility has been widely discussed since Daniel Bernoulli introduced it about 1730 in connection with the St. Petersburg game [3, 28]. However, it was not until certain mathematical results were proved in a limited setting by Kelly in 1956 [14] and then in an expanded and much more general setting by Breiman in 1960 and 1961 [5, 6] that logarithmic utility was clearly distinguished *by its properties* from other utilities as a guide to portfolio selection. (See also Bellman and Kalaba [2], Latané [15], Borch [4], and the very significant paper of Hakansson [11].)

Suppose for each time period ($n = 1, 2, \ldots$) there are k investment opportunities with results per unit invested denoted by the family of random variables $X_{n,1}, X_{n,2}, \ldots, X_{n,k}$. Suppose also that these random variables have only finitely many distinct values, that for distinct n the families are independent of each other, and that the joint probability distributions of distinct families (as subscripted) are identical. Then Breiman's results imply that portfolio strategies Λ which maximize $E \log X_n$, where X_n is the wealth at the end of the nth time period, have the following properties:

Property 1 (Maximizing $E \log X_n$ asymptotically maximizes the rate of asset growth.) If, for each time period, two portfolio managers have the same family of investment opportunities, or investment universes, and one uses a strategy Λ^* maximizing $E \log X_n$ whereas the other uses an "essentially different" [i.e., $E \log X_n(\Lambda^*) - E \log X_n(\Lambda) \to \infty$] strategy Λ, then $\lim X_n(\Lambda^*)/X_n(\Lambda) \to \infty$ almost surely (a.s.).

Property 2 The expected time to reach a fixed preassigned goal x is, asymptotically as x increases, least with a strategy maximzing $E \log X_n$.

The qualification "essentially different" conceals subtleties which are not generally appreciated. For instance, Hakansson [11], which is very close in

* This research was supported in part by the Air Force Office of Scientific Research under Grant AF-AFOSR 1870A. An expanded version of this paper will be submitted for publication elsewhere.

† Reprinted from the 1971 Business and Economics Statistics Section Proceedings of the American Statistical Association.

method to this article, and whose conclusions we heartily endorse, contains some mathematically incorrect statements and several incorrect conclusions, mostly from overlooking the requirement "essentially different." Because of lack of space we only indicate the problem here: If X_n is capital after the nth period, and if $x_j = X_j/X_0$, even though $E \log x_j > 0$ for all j, it need not be the case that $P(\lim x_j = \infty) = 1$. In fact, we can have (just as in the case of Bernoulli trials and $E \log x_j = 0$; see Thorp [26]) $P(\limsup x_j = \infty) = 1$ and $P(\liminf x_j = 0) = 1$ (contrary to Hakansson [11, p. 522, Eq. (18) and following assertions]). Similarly, when $E \log x_j < 0$ for all j we can have these alternatives instead of $P(\lim x_j = 0) = 1$ (contrary to Hakansson [11, p. 522, Eq. (17) and the following statements; footnote 1 is also incorrect]).

We note that with the preceding assumptions, there is a fixed fraction strategy Λ which maximizes $E \log X_n$. A fixed fraction strategy is one in which the fraction of wealth $f_{n,j}$ allocated to investment $X_{n,j}$ is independent of n.

We emphasize that Breiman's results can be extended to cover many if not most of the more complicated situations which arise in real-world portfolios. Specifically, the number and distribution of investments can vary with the time period, the random variables need not be finite or even discrete, and a certain amount of dependence can be introduced between the investment universes for different time periods.

We have used such extensions in certain applications (e.g., Thorp [25; 26, p. 287]).

We consider almost surely having more wealth than if an "essentially different" strategy were followed as the desirable objective for most institutional portfolio managers. (It also seems appropriate for wealthy families who wish mainly to accumulate and whose consumption expenses are only a small fraction of their total wealth.) Property 1 tells us that maximizing $E \log X_n$ is a recipe for approaching this goal asymptotically as n increases. This is to our mind the principal justification for selecting $E \log X$ as the guide to portfolio selection.

In any real application, n is finite, the limit is not reached, and we have $P(X_n(\Lambda^*)/X_n(\Lambda) > 1+M) = 1-\varepsilon(n, \Lambda, M)$, where $\varepsilon \to 0$ as $n \to \infty$, $M > 0$ is given, Λ^* is the strategy which maximizes $E \log X_n$, and Λ is an "essentially different" strategy. Thus in any application it is important to have an idea of how rapidly $\varepsilon \to 0$. Much work needs to be done on this in order to reduce $E \log X$ to a guide that is useful (not merely valuable) for portfolio managers. Some illustrative examples for $n = 6$ appear in the work of Hakansson [11].

Property 2 shows us that maximizing $E \log X$ also is appropriate for individuals who have a set goal (e.g., to become a millionaire).

Appreciation of the compelling properties of the Kelly criterion may have been impeded by certain misunderstandings about it that persist in the literature of mathematical economics.

The first misunderstanding involves failure to distinguish among kinds of utility theories. We compare and contrast three types of utility theories:

(1) *Descriptive*, where data on observed behavior are fitted mathematically. Many different utility functions might be needed, corresponding to widely varying circumstances, cultures, or behavior types.[1]

(2) *Predictive*, which "explains" observed data: Hypotheses are formulated from which fits for observed data are deduced; hopefully future data will also be found to fit. Many different utility functions may be needed, corresponding to the many sets of hypotheses that may be put forward.

(3) *Prescriptive* (also called *normative*), which is a guide to behavior, i.e., a recipe for optimally achieving a stated goal. It is not necessarily either descriptive or predictive nor is it intended to be so.

We use logarithmic utility in this last way, and much of the misunderstanding of it comes from those who think it is being proposed as a descriptive or a predictive theory. The $E \log X$ theory is a prescription for allocating resources so as to (asymptotically) maximize the rate of growth of assets. We assert that this is the appropriate goal in important areas of human endeavor and the theory is therefore important.

Another "objection" voiced by some economists to $E \log X$ and, in fact, to all unbounded utility functions, is that it does not resolve the (generalized) St. Petersburg paradox. The rebuttal is blunt and pragmatic: The generalized St. Petersburg paradox does not arise in the real world because any one real-world random variable is bounded (as is any finite collection). Thus in any real application the paradox does not arise.

To insist that a utility function resolves the paradox is an artificial requirement, certainly permissible, but obstructive and tangential to the goal of building a theory that is also a practical guide.

II. Samuelson's Objections to Logarithmic Utility

Samuelson [21, pp. 245–246; 22 pp. 4–5] says repeatedly, authorities (Williams [30], Kelly [14], Latané [15], Breiman [5, 6]) "have proposed a

[1] Information on descriptive utility is sparce; how many writers on the subject have even been able to determine for us their own personal utility?

drastic simplification of the decision problem whenever T (the number of investment periods)[2] is large.

"*Rule* Act in each period to maximize the geometric mean or the expected value of $\log x_t$.

"The plausibility of such a procedure comes from the recognition of the following valid asymptotic result.

"*Theorem* Acting to maximize the geometric mean at every step will, if the period is 'sufficiently long,' 'almost certainly' result in higher terminal wealth and terminal utility than from any other decision rule....

"From this indisputable fact, it is apparently tempting to believe in the truth of the following false corollary:

"*False Corollary* If maximizing the geometric mean almost certainly leads to a better outcome, then the expected value utility of its outcomes exceeds that of any other rule, provided T is sufficiently large."

Samuelson then gives examples to show that the corollary is false. We heartily agree that the corollary is false. In fact, Thorp [26] had already shown this for one of the utilities Samuelson uses, for he noted that in the case of Bernoulli trials with probability $\frac{1}{2} < p < 1$ of success, one should commit a fraction $w = 1$ of his capital at each trial to maximize expected final gain EX_n ([26, p. 283]; the utility is $U(x) = x$) whereas to maximize $E \log X_n$ he should commit $w = 2p - 1$ of his capital at each trial [26, p. 285, Theorem 4].

The statements which we have seen in print supporting this "false corollary" are by Latané [15, p. 151, footnote 13] as discussed by Samuelson [21, p. 245, footnote 8] and Markowitz [16a, pp. ix and x]. Latané may not have fully supported this corollary for he adds the qualifier "... (in so far as certain approximations are permissible) ...."

That there were (or are?) adherents of the "false corollary," seems puzzling in view of the following formulation. Consider a T-stage investment process. At each stage we allocate our resources among the available investments. For each sequence A of allocations which we choose, there is a corresponding terminal probability distribution $F_T{}^A$ of assets at the completion of stage T. For each utility function $U(\cdot)$, consider those allocations $A^*(U)$ which

[2] Parenthetical explanation added since we have used n.

maximize the expected value of terminal utility $\int U(x)\,dF_T{}^A(x)$. Assume sufficient hypotheses on U and the set of $F_T{}^A$ so that the integral is defined and that furthermore the maximizing allocation $A^*(U)$ exists. Then Samuelson says that $A^*(\log)$ is not in general $A^*(U)$ for other U. This seems intuitively evident.

Even more seems strongly plausible: that if U_1 and U_2 are inequivalent utilities, then $\int U_1(x)\,dF_T{}^A(x)$ and $\int U_2(x)\,dF_T{}^A(x)$ will in general be maximized for different $F_T{}^A$. (Two utilities U_1 and U_2 are equivalent if and only if there are constants a and b such that $U_2(x) = aU_1(x)+b$, $a > 0$; otherwise U_1 and U_2 are inequivalent.) In this connection we have proved [27a].

Theorem Let U and V be utilities defined and differentiable on $(0, \infty)$, with $0 < U'(x)$, $V'(x)$, and $U'(x)$ and $V'(x)$ strictly decreasing as x increases. Then if U and V are inequivalent, there is a one-period investment setting such that U and V have distinct optimal strategies.[3]

All this is in the nature of an aside, for Samuelson's correct criticism of the "false corollary" does not apply to our use of logarithmic utility. Our point of view is this: If your goal is Property 1 or Property 2, then a recipe for achieving either goal is to maximize $E \log X$. It is these properties which distinguish log for us from the prolixity of utility functions in the literature. Furthermore, we consider these goals appropriate for many (*but not all*) investors. Investors with other utilities, or with goals incompatible with logarithmic utility, will of course find it inappropriate for them.

Property 1 imples that if Λ^* maximizes $E \log X_n(\Lambda)$ and Λ' is "essentially different," then $X_n(\Lambda^*)$ tends almost certainly to be better than $X_n(\Lambda')$ as $n \to \infty$. Samuelson says [21, p. 246] after refuting the "false corollary": "Moreover, as I showed elsewhere [20, p. 4], the ordering principle of selecting between two actions in terms of which has the greater probability of producing a higher result does not even possess the property of being transitive. ... we could have w^{***} better than w^{**}, and w^{**} better than w^*, and also have w^* better than w^{***}."

For some entertaining examples, see the discussion of nontransitive dice by Gardner [9]. [Consider the dice with equiprobable faces numbered as follows: $X = (3,3,3,3,3,3)$, $Y = (4,4,4,4,1,1)$, $Z = (5,5,2,2,2,2)$. Then $P(Z > Y) = \frac{5}{9}$, $P(Y > X) = \frac{2}{3}$, $P(X > Z) = \frac{2}{3}$.] What Samuelson does not tell

[3] We have since generalized this to the case where $U''(x)$ and $V''(x)$ are piecewise continuous.

us is that the property of producing a higher result *almost certainly*, as in Property 1, *is* transitive. If we have $w^{***} > w^{**}$ almost certainly, and $w^{**} > w^{*}$ almost certainly, then we must have $w^{***} > w^{*}$ almost certainly.

One might object [20, p. 6] that in a real investment sequence the limit as $n \to \infty$ is not reached. Instead the process stops at some finite N. Thus we do not have $X_N(\Lambda^*) > X_n(\Lambda')$ almost surely. Instead we have $P(X_n(\Lambda^*) > X_n(\Lambda')) = 1-\varepsilon_N$ where $\varepsilon_N \to 0$ as $N \to \infty$, and transitivity can be shown to fail.

This is correct. But then an approximate form of transitivity does hold: Let X, Y, Z be random variables with $P(X > Y) = 1-\varepsilon_1$, $P(Y > Z) = 1-\varepsilon_2$. Then $P(X > Z) \geqq 1-(\varepsilon_1+\varepsilon_2)$. To prove this, let A be the event $X > Y$, B the event $Y > Z$, and C the event $X > Z$. Then $P(A)+P(B) = P(A \cup B)+P(A \cap B) \leqq 1+P(A \cap B)$. But $A \cap B \subset C$ so $P(C) \geqq P(A \cap B) = P(A)+P(B)-1$, i.e., $P(X > Z) \geqq 1-(\varepsilon_1+\varepsilon_2)$.

Thus our approach is not affected by the various Samuelson objections to the uses of logarithmic utility.

Markowitz [16a, p. ix and x] says "... in 1955–1956, I concluded... that the investor who is currently reinvesting everything for 'the long run' should maximize the expected value of the logarithm of wealth." (This assertion seems to be regardless of the investor's utility and so indicates belief in the false corollary.) "Mossin [18] and Samuelson [20] have each shown that this conclusion is not true for a wide range of [utility] functions .... The fascinating Mossin–Samuelson result, combined with the straightforward arguments supporting the earlier conclusions, seemed paradoxical at first. I have since returned to the view of Chapter 6 (concluding that: for large T, the Mossin–Samuelson man acts absurdly, ..." Markowitz says here, in effect, that alternate utility functions (to $\log x$) are absurd. This position is unsubstantiated and unreasonable.

He continues "...like a player who would pay an unlimited amount for the St. Petersburg game...." If you agree with us that the St. Petersburg game is not realizable and may be ignored when fashioning utility theories for the real world, then his continuation "...the terminal utility function must be bounded to avoid this absurdity; ..." does not follow.

Finally, Markowitz says "...and the argument in Chapter 6 applies when utility of terminal wealth is bounded." If he means by this that the false corollary holds if we restrict ourselves to bounded utility functions, then he is mistaken. Mossin [18] already showed that the optimal strategies for $\log x$ and x^γ/γ, $\gamma \neq 0$, are a fixed fraction for these and only these utilities. Thus any bounded utility besides x^γ/γ, $\gamma < 0$, will have optimal strategies which are *not* fixed fraction, hence not optimal for $\log x$. Samuelson [22] gives counterexamples including the bounded utilities x^γ/γ, $\gamma < 0$. Also the counterexamples satisfying the hypotheses of our theorem include many bounded utilities [e.g., $U(x) = \tan^{-1} x$, $x > 0$, and $V(x) = 1 - e^{-x}$, $x \geqq 0$].

III. An Outline of the Theory of Logarithmic Utility

The simplest case is that of Bernoulli trials with probability p of success, $0 < p < 1$. The unique strategy which maximizes $E \log X_n$ is to bet at trial n the fixed fraction $f^* = p - q$ of total current wealth X_{n-1}, if $p > \frac{1}{2}$ (i.e., if the expectation is positive), and to bet nothing otherwise.

To maximize $E \log X_n$ is equivalent to maximizing $E \log[X_n/X_0]^{1/n} \equiv G(f)$, which we call the (exponential) rate of growth (per time period). It turns out that for $p > \frac{1}{2}$, $G(f)$ has a unique positive maximum at f^* and that there is a critical fraction f_c, $0 < f^* < f_c < 1$, such that $G(f_c) = 0$, $G(f) > 0$ if $0 < f < f_c$, $G(f) < 0$ if $f_c < f \leqq 1$ (we assume "no margin"; the case with margin is similar). If $f < f_c$, $X_n \to \infty$ a.s.; if $f = f_c$, $\limsup X_n = +\infty$ a.s. and $\liminf X_n = 0$ a.s.; if $f > f_c$, $\lim X_n = 0$ a.s. ("ruin").

The Bernoulli trials case is particularly interesting because it exhibits many of the features of the following more general case. Suppose we have at each trial $n = 1, 2, \ldots$ the k investment opportunities $X_{n,1}, X_{n,2}, \ldots, X_{n,k}$ and that the conditions of Property 1 (Section I) are satisfied. This means that the joint distributions of $\{X_{n,i_1}, X_{n,i_2}, \ldots, X_{n,i_j}\}$ are the same for all n, for each subset of indices $1 \leqq i_1 < i_2 < \cdots < i_j \leqq k$. Furthermore $\{X_{m,1}, \ldots, X_{m,k}\}$ and $\{X_{n,1}, \ldots, X_{n,k}\}$ are independent, and all random variables $X_{i,j}$ have finite range. Thus we have in successive time periods repeated independent trials of "the same" investment universe.

Since Breiman has shown that there is for this case an optimal fixed fraction strategy $\Lambda^* = (f_1^*, \ldots, f_k^*)$, we will have an optimal strategy if we find a strategy which maximizes $E \log X_n$ in the class of fixed fraction strategies.

Let $\Lambda = (f_1, \ldots, f_k)$ be any fixed fraction strategy. We assume that $f_1 + \cdots + f_k \leqq 1$ so there is no borrowing, or margin. The margin case is similar (the approach resembles that of Schrock [23]). Using the concavity of the logarithm, it is easy to show (see below) that the exponential rate of growth $(1/n) E \log X_n(\Lambda) = G(f_1, \ldots, f_k)$ is a concave function of $(f_1, \ldots, f_k)$, just as in the Bernoulli trials case. The domain of $G(f)$ in the Bernoulli trials case was the interval $[0, 1)$ with $G(f) \downarrow -\infty$ as $f \to 1$. The domain in the present instance is analogous. First, it is a subset of the k-dimensional simplex $S_k = \{(f_1, \ldots, f_k) : f_1 + \cdots + f_k \leqq 1; f_1 \geqq 0, \ldots, f_k \geqq 0\}$.

To establish the analogy further, let $R_j = X_{n,j} - 1$, $j = 1, \ldots, k$, be the return per unit on the ith investment opportunity at an arbitrary time period n.

Let the range of R_j be $\{r_{j,1};\dots;r_{j,i_j}\}$ and let the probability of the outcome $[R_1 = r_{1,m_1}$ and $R_2 = r_{2,m_2}$ and so on, up to $R_k = r_{k,m_k}]$ be $p_{m_1 m_2,\dots,m_k}$. Then

$$
\begin{aligned}
&E \log X_n/X_{n-1} \\
&\quad = G(f_1,\dots,f_k) \\
&\quad = \sum \{p_{m_1,\dots,m_k} \log(1+f_1 r_{1,m_1}+\cdots+f_k r_{k,m_k})\colon\ 1 \leqq m_1 \leqq i_1;\dots;1 \leqq m_k \leqq i_k\},
\end{aligned}
$$

from which the concavity of $G(f_1,\dots,f_k)$ can be shown. Note that $G(f_1,\dots,f_k)$ is defined if and only if $1+f_1 r_{1,m_1}+\cdots+f_k r_{k,m_k} > 0$ for each set of indices $m_1,\dots,m_k$. Thus the domain of $G(f_1,\dots,f_k)$ is the intersection of all these open half-spaces with the k-dimensional simplex S_k. Note that the domain is convex and includes all of S_k in some neighborhood of the origin. Note too that the domain of G is all of S_k if (and only if) $R_j > -1$ for all j, i.e., if there is no probability of total loss on any investment. The domain of G includes the interior of S_k if $R_j \geqq -1$. Both domains are particularly simple and most cases of interest are included.

If $f_1,\dots,f_k$ are chosen so that $1+f_1 r_{1,m_1}+\cdots+f_k r_{k,m_k} \leqq 0$ for some $m_1,\dots,m_k$, then $P(f_1 X_{n,1}+\cdots+f_k X_{n,k} \leqq 0) = \varepsilon > 0$ for all n and ruin occurs with probability 1.

Computational procedures for finding an optimal fixed fraction strategy (generally unique in our present setting) are based on the theory of concave (dually, convex) functions [29] and will be presented elsewhere. (As Hakansson [11, p. 552] has noted, "...the computational aspects of the capital growth model are [presently] much less advanced" than for the Markowitz model.) A practical computational approach for determining the $f_1,\dots,f_k$ to good approximation is given in [15a].

The theory may be extended to more general random variables and to dependence between different time periods. Most important, we may include the case where the investment universe changes with the time period, provided only that there be some mild regularity condition on the $X_{i,j}$, such as that they be uniformly a.s. bounded. (See the discussion by Latané [15], and the generalization of the Bernoulli trials case as applied to blackjack betting by Thorp [26].) The techniques rely heavily on those used to so generalize the law of large numbers.

Transaction costs, the use of margin, and the effect of taxes can be incorporated into the theory. Bellman's dynamic programming method is used here.

The general procedure for developing the theory into a practical tool imitates Markowitz [16]. Markowitz requires as inputs estimates of the expectations, standard deviations, and covariances of the $X_{i,j}$. We require joint probability distributions. This would seem to be a much more severe requirement, but in practice does not seem to be so [16, pp. 193–194, 198–201).

Among the actual inputs that Markowitz [16] chose were (1) past history [16, Example, pp. 8–20], (2) probability beliefs of analysts [16, pp. 26–33], and (3) models, most notably regression models, to predict future performance from past data [16, pp. 33, 99–100]. In each instance one can get enough additional information to estimate $E\log(X_n/X_{n-1})$.

There are, however, two great difficulties which all theories of portfolio selection have, including ours and that of Markowitz. First, there seems to be no established method for generally predicting security prices that gives an edge of even a few percent. The random walk is the best model for security prices today (see Cootner [7] and Granger and Morgenstern [10]).

The second difficulty is that for portfolios with many securities the volume of inputs called for is prohibitive: For 100 securities, Markowitz requires 100 expectations and 4950 covariances; and our theory requires somewhat more information. Although considerable attention has been given to finding condensed inputs that can be used instead, this aspect of portfolio theory still seems unsatisfactory.

In Section V we show how both these difficulties were overcome in practice by an institutional investor. That investor, guided by the Kelly criterion, then outperformed for the year 1970 every one of the approximately 400 Mutual Funds listed by the S & P stock guide!

But first we relate our theory to that of Markowitz.

IV. Relation to the Markowitz Theory; Solution to Problems Therein

The most widely used guide to portfolio selection today is probably the Markowitz theory. The basic idea is that a portfolio P_1 is superior to a portfolio P_2 if the expectation ("gain") is at least as great, i.e., $E(P_1) \geqq E(P_2)$, and the standard deviation ("risk") is no greater, i.e., $\sigma(P_1) \leqq \sigma(P_2)$, with at least one inequality. This partially orders the set $\mathscr{P}$ of portfolios. A portfolio such that no portfolio is superior (i.e., a maximal portfolio in the partial ordering) is called *efficient*. The goal of the portfolio manager is to determine the set of efficient portfolios, from which he then makes a choice based on his needs.

This is intuitively very appealing: It is based on standard quantities for the securities in the portfolio, namely expectation, standard deviation, and covariance (needed to compute the variance of the portfolio from that of the component securities). It also gives the portfolio manager "choice."

As Markowitz [16, Chapter 6] has pointed out, the optimal Kelly portfolio is approximately one of the Markowitz efficient portfolios under certain circumstances. If $E = E(P)$ and $R = P-1$ is the return per unit of the portfolio P, let $\log P = \log(1+R) = \log((1+E)+(R-E))$. Expanding in Taylor's series about $1+E$ gives

$$\log P = \log(1+E) + \frac{R-E}{1+E} - \frac{1}{2}\frac{(R-E)^2}{(1+E)^2} + \text{higher order terms.}$$

Taking expectations and neglecting higher order terms gives

$$E \log P = \log(1+E) - \frac{1}{2}\frac{\sigma^2(P)}{(1+E)^2}.$$

This leads to a simple pictorial relationship with the Markowitz theory. Consider the E–σ plane, and plot $(E(P), \sigma(P))$ for the efficient portfolios. The locus of efficient portfolios is a convex nondecreasing curve which includes its endpoints (Fig. 1).

Then constant values of the growth rate $G = E \log P$ approximately satisfy

$$G = \log(1+E) - \frac{1}{2}\frac{\sigma^2(P)}{(1+E)^2}.$$

This family of curves is illustrated in Fig. 1 and the (efficient) portfolio which maximizes logarithmic utility is approximately the one which lies on the greatest G curve. Because of the convexity of the curve of efficient portfolios and the concavity of the G curves, the (E, σ) value where this occurs is unique.

The approximation to G breaks down badly in some significant practical settings, including that of the next section. But for portfolios with large numbers of "typical" securities, the approximation for G will generally provide an (efficient) portfolio which approximately maximizes asset growth. This solves the portfolio manager's problem of which Markowitz-efficient portfolio to choose. Also, if he repeatedly chooses his portfolio in successive time periods by this criterion, he will tend to maximize the rate of growth of his assets, i.e., maximize "performance." We see also that in this instance the problem is reduced to that of finding the Markowitz-efficient portfolios plus the easy step of using Fig. 1. Thus if the Markowitz theory can be applied in practice in this setting, so can our theory. We have already remarked on the ambiguity of the set of efficient portfolios, and how our theory resolves them. To illustrate further that such ambiguity represents a defect in the Markowitz theory, let X_1 be uniformly distributed over [1, 3], let X_2 be uniformly distributed over [10, 100], let $\text{cor}(X_1, X_2) = 1$, and suppose these are the only securities. Then X_1 and X_2 are both efficient with $\sigma_1 < \sigma_2$ and $E_1 < E_2$ so the Markowitz theory does not choose between them. Yet "everyone" would choose X_2 over X_1 because the worst outcome with X_2 is far

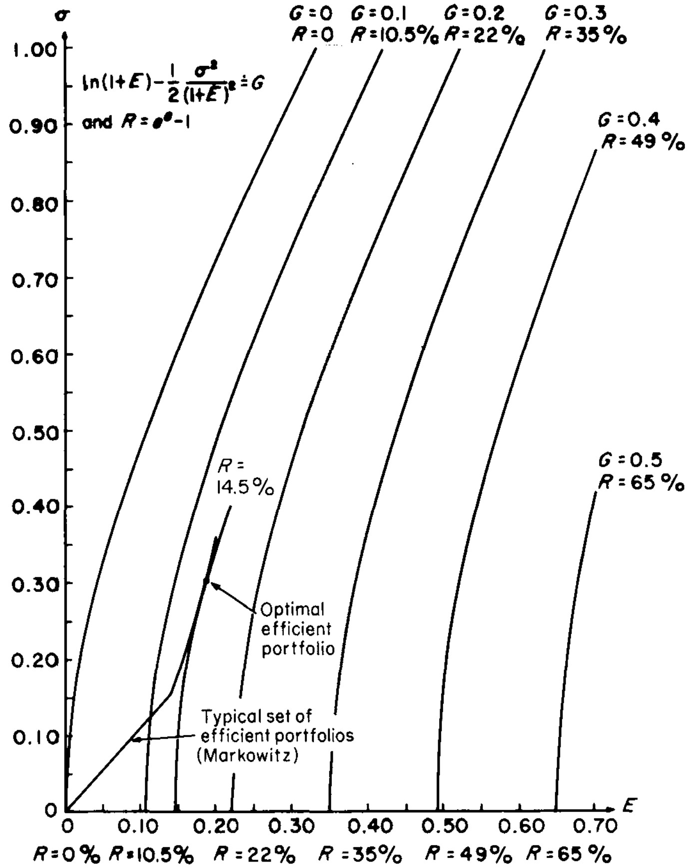

Fig. 1. Growth rate G (return rate R) in the E–σ plane assuming the validity of the power series approximation.

better than the best outcome from X_1. (We presented this example in Thorp [26]. Hakansson [11] presents further examples and extended analysis. He formalizes the idea by introducing the notion of stochastic dominance: X stochastically dominates Y if $P(X \geqq Y) = 1$ and $P(X > Y) > 0$. It is easy to see that

Lemma An $E \log X$ optimal portfolio is never stochastically dominated.

Thus our portfolio theory does not have this defect.)

There are investment universes $(X_1, \ldots, X_n)$ such that a unique portfolio P maximizes $E \log P$, yet P is not efficient in the sense of Markowitz. Then choosing P in repeated independent trials will outperform any strategy limited to choosing efficient portfolios. In addition, the optimal Kelly strategy gives positive growth rate, yet some of the Markowitz-efficient strategies give negative growth rate and ruin after repeated trials. Thorp [26] gave such an example and another appears in the work of Hakansson [11]. See also Hakansson [11, pp. 553–554] for further discussion of defects in the Markowitz model.

V. The Theory in Action: Results for a Real Institutional Portfolio

The elements of a practical profitable theory of convertible hedging were published by Thorp and Kassouf [27]. Thorp and Kassouf indicated an annualized return on investments of the order of 25% per year. Since then the theory has been greatly extended and refined [26] with most of these new results thus far unpublished.

The historical data which have been used to develop the theory include for warrants monthly observations for about 20 years, averaging about 20 warrants, or about 4800 observations, plus weekly observations for three years on an average of about 50 warrants, another 7500 observations. Each of these more than 12,000 observations is an n-tuple including price of common, price of warrant, dividend, dilution, time to expiration, and several other quantities.

The studies also used weekly data for three years on an average of 400 convertible bonds and 200 convertible preferreds, or about 90,000 observations. Including still other data, well over 100,000 observations have been incorporated into the study.

A convertible hedge transaction generally involves *two* securities, one of which is convertible into the other. Certain mathematical price relationships exist between pairs of such securities. When one of the two is underpriced, compared to the other, a profitable convertible hedge may be set up by buying the relatively underpriced security and selling short an appropriate amount of the relatively overpriced security.

The purpose of selling short the overpriced security is to reduce the risk in the position. Typically, one sells short in a single hedge from 50 to 125% as much stock (in "share equivalents") as is held long. The exact proportions depend on the analysis of the specific situation; the quantity of stock sold short is selected to minimize risk. The risk (i.e., change in asset value with fluctuations in market prices) in a suitable convertible hedge should be much less than in the usual stock market long positions.

TABLE I

PERFORMANCE RECORD

Date	Chg. to Date (%)[a]	Growth rate to date[b]	Elapsed time (months)	Closing DJIA[c]	DJIA Chg. (%)[a]	Starting even with DJIA[c]	Gain over DJIA (%)[a]
11-3-69	0.0	—	0	855	0.0	855	0.0
12-31-69	+4.0	+26.8	2	800	−6.3	889	+10.3
9-1-70	+14.0	+17.0	10	758	−11.3	974	+25.3
12-31-70	+21.0	+17.7	14	839	−1.8	1034	+22.8
3-31-71	+31.3	+21.2	17	904	+5.8	1123	+25.5
6-30-71	+39.8	+22.3	20	891	+4.2	1196	+35.6
9-30-71	+49.4	+23.3	23	887	+3.7	1278	+45.7
12-31-71	+61.3	+24.7	26	890	+4.1	1379	+57.2
3-31-72	+69.5	+24.4	29	940	+9.9	1449	+59.6
6-30-72	+75.0	+23.3	32	934	+9.3	1496	+65.7
9-30-72	+78.9	+22.1	35	960	+12.3	1526	+66.6
12-31-72	+84.5	+21.3	38	1020	+19.3	1577	+65.2
3-31-73	+90.0	+20.7	41	957	+11.9	1625	+78.1
6-30-73	+91.6	+19.4	44	891	+4.2	1638	+87.4
9-30-73	+100.3	+19.4	47	947	+10.7	1713	+89.6
12-31-73	+102.9	+17.7	52	851	−0.5	1734	+103.4

[a] Round to nearest 0.1%.
[b] Compound growth rate, annualized.
[c] Round to nearest point; DJIA = Dow Jones industrial average.

The securities involved in convertible hedges include common stock, convertible bonds, convertible preferreds, and common stock purchase warrants. Options such as puts, calls, and straddles may replace the convertible security. For this purpose, the options may be either written or purchased.

The reader's attention is directed to three fundamental papers on the theory of options and convertibles which have since been published by Black and Scholes [3a, 3b] and by Merton [16b].

The theory of the convertible hedge is highly enough developed so that the probability characteristics of a single hedge can be worked out based on an assumption for the underlying distribution of the common. (Sometimes even this can almost be dispensed with! See Thorp and Kassouf [27, Appendix C].) A popular and plausible assumption is the random walk hypothesis: that the future price of the common is log-normally distributed about its current price, with a trend and a variance proportional to the time. Plausible estimates of these parameters are readily obtained. Furthermore, it turns out that the return from the hedge is comparatively insensitive to changes in the estimates for these parameters.

Thus with convertible hedging we fulfill two important conditions for the practical application of our (or any other) theory of portfolio choice: (1) We have identified investment opportunities which are markedly superior to the usual ones. Compare the return rate of 20–25% per year with the long-term rate of 8% or so for listed common stocks. Further, it can be shown that the risks tend to be much less. (2) The probability inputs are available for computing $G(f_1, \ldots, f_n)$.

On November 3, 1969, a private institutional investor decided to commit all its resources to convertible hedging and to use the Kelly criterion to allocate its assets. Since this article was written, this institution has continued to have a positive rate of return in every month; on 12–31–73 the cumulative gain reached +102.9% and the DJIA equivalent reached 1734, whereas the DJIA was at 851, off 0.5%. The performance record is shown in Table I.

The market period covered included one of the sharpest falling markets as well as one of the sharpest rising markets (up 50% in 11 months) since World War II. The gain was +16.3% for the year 1970, which outperformed all of the approximately 400 Mutual Funds listed in the S & P stock guide. Unaudited figures show that gains were achieved during every single month.

(*Note added in proof:* Gains continued consistent in 1974. On 12–31–74 the cumulative gain reached +129%, the DJIA equivalent was 1960, and the DJIA reached +218%. Proponents of efficient market theory, please explain.)

The unusually low risk in the hedged positions is also indicated by the results for the 200 completed hedges. There were 190 winners, 6 break-evens, and 4 losses. The losses as a percentage of the long side of the specific investment ranged from 1 to 15%.

A characteristic of the Kelly criterion is that as risk decreases and expectation rises, the optimal fraction of assets to be invested in a single situation may become "large." On several occasions, the institution discussed above invested up to 30% of its assets in a single hedge. Once it invested 150% of its assets in a single arbitrage. This characteristic of Kelly portfolio strategy is not part of the behavior of most portfolio managers.

To indicate the techniques and problems, we consider a simple portfolio with just one convertible hedge. We take as our example Kaufman and Broad common stock and warrants. A price history is indicated in Fig. 2.

The figure shows that the formula $W = 0.455S$ is a reasonable fit for $S \leqq 38$ and that $W = S - 21.67$ is a reasonable fit for $S \geqq 44$. Between $S = 38$ and $S = 44$ we have the line $W = 0.84x - 15.5$. For simplicity of calculation we

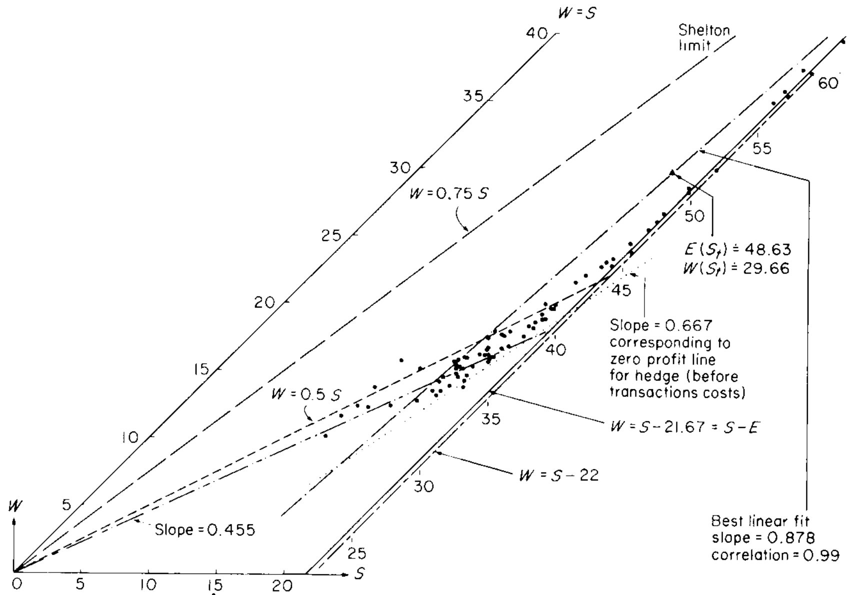

Fig. 2. Price history of Kaufman and Broad common S versus the warrants W. The points moved up and to the right until they reached the neighborhood of (38, 17). At this point a 3 : 2 hedge (15,000 warrants long, 11,200 common short) was instituted. As the points continued to move up and to the right during the next few months, the position was closed but in stages with the final liquidation at about (58, 36). Terms of warrant: one warrant plus \$21.67 → one share common stock until 3–1–74. Full protection against dilution. The company has the right to reduce the exercise price for temporary periods; 750,000 warrants and 5,940,000 common outstanding. Common dividends Q.05ex 10–26–70.

replace this in our illustrative analysis by $W = 0.5S$ if $S \leqq 44$ and $W = S - 22$ if $S \geqq 44$. The lines are also indicated in Fig. 2.

Past history at the time the hedge was instituted in late 1970 supported the fit for $S \leqq 38$. The conversion feature of the warrant ensures $W \geqq S - 21.67$ until the warrant expires. Thus $W = S - 21.67$ for $S \geqq 44$ underestimates the price of the warrant in this region. Extensive historical studies of warrants [12, 13, 24, 27] show that the past history fit would probably be maintained until about two years before expiration, i.e., until about March 1972. Thus it is plausible to assume that for the next 1.3 years, the S may be roughly approximated by $W = 0.5S$ for $S \leqq 44$ and $W = S-22$ for $S \geqq 44$.

Next we assume that S_t, the stock price at time $t > 0$ years after the hedge was initiated, is log-normally distributed with density

$$f_{S_t}(x) = (x\sigma\sqrt{2\pi})^{-1} \exp[-(\ln x - \mu)^2/2\sigma^2],$$

mean $E(S_t) = \exp(\mu+\sigma^2/2)$, and standard deviation

$$\sigma(S_t) = E(S_t)(\exp(\sigma^2)-1)^{1/2}.$$

The functions $\mu \equiv \mu(t)$ and $\sigma \equiv \sigma(t)$ depend on the stock and on the time t. If t is the time in years until S_t is realized, we will see below that it is plausible to assume $\mu_t = \log S_0 + mt$ and $\sigma_t{}^2 = a^2 t$, where S_0 is the present stock price and m and a are constants depending on the stock.

Then $E(S_t) = S_0 \exp[(m+a^2/2)t]$ and a mean increase of 10% per year is approximated by setting $m+a^2/2 = 0.1$. If we estimate a^2 from past price changes, we can solve for m. In the case of Kaufman and Broad it is plausible to take $\sigma \doteq 0.45$, from which $a^2 = \sigma^2 \doteq 0.20$. This yields $m = 0$.

The standard deviation is then

$$\sigma(S_t) = S_0 \exp[(m+a^2/2)t][\exp(a^2 t)-1]^{1/2} \doteq 0.52 S_0.$$

It is generally agreed that the serial correlation of stock-price changes is very weak, and that changes in a stock-price series are approximately independent if the time intervals are nonoverlapping [7, 10]. If the changes in unit time for a stock were bounded independent identically distributed random variables, the central limit theorem would lead to the normal approximation with the mean and variance of a change proportional to the time. But this has difficulties. For instance, the change is bounded below because stock prices are nonnegative. Also, the magnitude of the price change per unit time is in fact dependent on the current price, increasing as the price increases.

A more realistic model which eliminates these difficulties and seems more plausible, is to assume that the price changes are proportional to the current price. This leads to the hypothesis that the changes in the logarithms of the prices are bounded identically distributed independent random variables, i.e., that $\log S_t - \log S_0$ is normally distributed with $\mu(t) = mt$ (hence $\log S_t$ has mean $\log S_0 + mt$) and $\sigma(t)^2 = a^2 t$. This is a sketch of the thinking behind our assumptions in the present example. For a detailed discussion, see Osborne [19] and Ayres [1a].

It is by no means established that the log-normal model is the appropriate one for stock-price series [10, Chapter 7]. However, once we clarify certain general principles by working through our example on the basis of the log-normal model, it can be shown that the results are substantially unchanged by choosing instead any distribution that roughly fits observation!

For a time of one year, a computation shows the return $R(S)$ on the stock to be $+10.5\%$, the return $R(W)$ on the warrant to be $+34.8\%$, $\sigma(S) = 0.52$, $\sigma(W) = 0.92$, and the correlation coefficient $\text{cor}(S, W) = 0.99$. The difference in $R(S)$ and $R(W)$ shows that the warrant is a much better buy than the common. Thus a hedge of long warrants and short common has a substantial positive expectation. The value $\text{cor}(S, W) = 0.99$ shows that a hedge corresponding to the best linear fit of W to S has a standard deviation of approximately $(1-0.99)^{1/2} = 0.1$, which suggests that $\sigma(P)$ for the optimal hedged portfolio is probably going to be close to 0.1. The high return and low risk for the hedge will remain, it can be shown, under wide variations in the choice of m and a.

To calculate the optimal mix of warrants long to common short we maximize $G(f_1, f_2) = E\log(1+f_1 S+f_2 W)$. The detailed computational procedures are too lengthy and involved to be presented here. We hope to publish them elsewhere.

The actual decision made by our institutional investor took into consideration other positions already held, some of which might have to be closed out to release assets, and also those which were currently candidates for investment. The position finally taken was to short common and buy warrants in the ratio of three shares to four. The initial market value of the long side was about 14% of assets and the initial market value of the short side was about 20% assets. The profit realized was, in terms of the initial market value of the long side, about 20% in six months. This resulted from a move in the common from about 40 to almost 60.

VI. Concluding Remarks

As we remarked above, we do not propose logarithmic utility as descriptive of actual investment behavior, nor do we believe any one utility function could suffice. It would be of interest, however, to have empirical evidence showing areas of behavior which are characterized adequately by logarithmic utility. Neither do we intend logarithmic utility to be predictive; again, it would be of interest to know what it does predict.

We only propose the theory to be normative or prescriptive, and only for those institutions, groups or individuals whose overriding current objective is maximization of the rate of asset growth. Those with a different "prime directive" may find another utility function which is a better guide.

We remark that $E \log X$ has in our experience been a valuable qualitative guide and we suggest that this could be its most important use. Once familiarity with its properties is gained, our experience suggests that many investment decisions can be guided by it without complex supporting calculations.

It is interesting to inquire into the sort of economic behavior is to be expected from followers of $E \log X$. We find that insurance is "explained," i.e., that even though it is a negative expectation investment for the insured and we assume both insurer and insured have the same probability information, it is often optimal for him (as well as for the insurance company) to insure [3]. It usually turns out that insurance against large losses is indicated and insurance against small losses is not. (Do not insure an old car for collision, take $200 deductible, not $25, etc.)

We find that if all parties to a security transaction are followers of $E \log X$, they will often find it mutually optimal to make securities transactions. This may be true whether the transactions be two party (no brokerage), or three party (brokerage), and whether or not they have the same probability information about the security involved, or even about the entire investment universe.

Maximizing logarithmic utility excludes portfolios which have positive probability of total loss of assets. Yet it can be argued that an impoverished follower of $E \log X$ might in some instances risk "everything." This agrees with some observed behavior, but is not what we might at first expect in view of the prohibition against positive probability of total loss. Consider each individual as a piece of capital equipment with an assignable monetary value. Then if he risks and loses all his cash assets, he hasn't really lost everything [3].

All of us behave as though death itself does not have infinite negative utility. Since the risk of death, although generally small, is ever present, a negative infinite utility for death would make all expected utilities negative infinite and utility theory meaningless. In the case of logarithmic utility as applied to the extended case of the (monetized) individual plus all his resources, death has a finite, though large and negative, utility. The value of this "death constant" is an additional arbitrary assumption for the enlarged theory of logarithmic utility.

In the case of investors who behave according to $E \log X$ (or other utilities unbounded below), it might be possible to discover their tacit "death constants."

Hakansson [11, p. 551] observes that logarithmic utility exhibits decreasing absolute risk aversion in agreement with deductions of Arrow [1] and others on the qualities of "reasonable" utility functions. Hakansson says, "What the relative risk aversion index [given by $-xU''(x)/U'(x)$] would look like for a meaningful utility function is less clear.... In view of Arrow's conclusion that '...broadly speaking, the relative risk aversion must hover around 1, being, if anything, somewhat less for low wealths and somewhat higher for high wealths...' the optimal growth model seems to be on safe ground." As he notes, for $U(x) = \log x$, the relative risk aversion is precisely 1. However, in both the extension to valuing the individual as capital equipment, and the further extension to include the death constant, we are led to $U(x) = \log(x+c)$, where c is positive. But then the relative risk-aversion index is $x/(x+c)$, which behaves strikingly like Arrow's description. See also the discussion of $U(x) = \log(x+c)$ by Freimer and Gordon [8, pp. 103, 112].

Morgenstern [17] has forcefully observed that assets are random variables, not numbers, and that economic theory generally does not incorporate this. To replace assets by *numbers* having *the same expected utility* in valuing companies, portfolios, property, and the like, allows for comparisons when asset values are given as random variables. We, of course, think logarithmic utility will often be the appropriate tool for such valuation.

ACKNOWLEDGMENT

I wish to thank James Bicksler for several stimulating and helpful conversations.

REFERENCES

1. ARROW, K. J., *Aspects of the Theory of Risk-Bearing*. Yrjö Jahnssonin Säätiö, Helsinki, 1965. Reprinted in K. J. Arrow, *Essays in the Theory of Risk-Bearing*. Markham, Chicago and North-Holland, London (1970).
1a. AYRES, H. F., "Risk Aversion in the Warrant Markets." S. M. Thesis, MIT, Cambridge, Massachusetts; *Industrial Management Review* **5**: 1 (1963), 45–53; reprinted by Cootner [7, pp. 479–505].
2. BELLMAN, R., and KALABA, R., "On the Role of Dynamic Programming in Statistical Communication Theory." *IRE Transactions of the Professional Group on Information Theory* **IT-3**: 3 (1957), 197–203.
3. BERNOULLI, D., "Exposition of a New Theory on the Measurement of Risk." *Econometrica* **22** (1954), 23–36 (translated by Louise Sommer).
3a. BLACK, F., and SCHOLES, M., "The Valuation of Option Contracts and a Test of Market Efficiency." *Journal of Finance* **27** (1972), 399–417.
3b. BLACK, F., and SCHOLES, M., "The Pricing of Options and Corporate Liabilities." *Journal of Political Economy* **81** (1973), 637–654.
4. BORCH, K. H., *The Economics of Uncertainty*. Princeton Univ. Press, Princeton, New Jersey, 1968.
5. BREIMAN, L., "Investment Policies for Expanding Businesses Optimal in a Long Run Sense." *Naval Research Logistics Quarterly* **7**: 4 (1960), 647–651.

6. Breiman, L., "Optimal Gambling Systems for Favorable Games." *Symposium on Probability and Statistics 4th, Berkeley, 1961*, **1** pp. 65–78.
7. Cootner, P. H. (ed.), *The Random Character of Stock Market Prices.* MIT Press, Cambridge, Massachusetts, 1964.
8. Freimer, M., and Gordon, M. S., "Investment Behavior with Utility a Concave Function of Wealth." In *Risk and Uncertainty* (K. Borch and J. Mossin, eds.), pp. 94–115. St. Martin's Press, New York, 1968.
9. Gardner, M., "Mathematical Games: The Paradox of the Non-Transitive Dice and the Elusive Principle of Indifference." *Scientific American* **223** (December 1970), 110.
10. Granger, C., and Morgenstern, O., *Predictability of Stock Market Prices.* Heath, Lexington Books, Lexington, Massachusetts, 1970.
11. Hakansson, N., "Capital Growth and the Mean-Variance Approach to Portfolio Selection." *Journal of Finance and Quantitative Analysis* **6** (1971), 517–557.
12. Kassouf, S. T., "A Theory and an Econometric Model for Common Stock Purchase Warrants." Ph.D. Thesis, Columbia Univ. New York, 1965; published by Analytic Publ., New York, 1965. [A regression model statistical fit of normal price curves for warrants. There are large systematic errors in the model due to faulty (strongly biased) regression techniques. The average mean square error in the fit is large. Thus, it is not safe to use the model in practice as a predictor of warrant prices. However, the model and the methodology are valuable as a first *qualitative* description of warrant behavior and as a guide to a more precise analysis.]
13. Kassouf, S. T., "An Econometric Model for Option Price with Implications for Investors' Expectations and Audacity." *Econometrica* **37**: 4 (1969), 685–694. [Based on the thesis. The variance of residuals is given as .248 for a standard deviation of about .50 in y, the normalized warrant price, and a standard error of about .34. The mid-range of y varies from 0 to .5 and is never greater, thus the caveat about not using the model for practical predictions!]
14. Kelly, J. L., "A New Interpretation of Information Rate." *Bell System Technical Journal* **35** (1956), 917–926.
15. Latané, H. A., "Criteria for Choice Among Risky Ventures." *Journal of Political Economy* **67** (1959), 144–155.
15a. Maier, S. F., Peterson, D. W., and Vander Weide, J. H., "A Monte Carlo Investigation of Characteristics of Optimal Geometric Mean Portfolios." *Duke Working Paper* #120. Presented to session on Multiperiod Portfolio Theory, American Financial Association Annual Meeting, San Francisco (December, 1974).
16. Markowitz, H. M., *Portfolio Selection: Efficient Diversification of Investments.* Wiley, New York, 1959.
16a. Markowitz, H. M., *Portfolio Selection: Efficient Diversification of Investments*, 2nd printing, Preface. Yale Univ. Press, New Haven, Connecticut, 1970, reprint.
16b. Merton, R. C., "Theory of Rational Option Pricing." *Bell Journal of Economic and Management Science* **4** (1973), 141–183.
17. Morgenstern, O., *On the Accuracy of Economic Observations*, 2nd ed., revised. Princeton Univ. Press, Princeton, New Jersey, 1963.
18. Mossin, J., "Optimal Multiperiod Portfolio Policies." *Journal of Business* **41** (1968), 215–229.
19. Osborne, M. F. M., "Brownian Motion in the Stock Market." *Operations Research* **7** (1959), 145–173, reprinted in Cootner [7, pp. 100–128].

20. Samuelson, P. A., "Risk and Uncertainty: A Fallacy of Large Numbers." *Scienta (Milan)* **57** (1963), 153–158.
21. Samuelson, P. A., "Lifetime Portfolio Selection by Dynamic Stochastic Programming." *Review of Economics and Statistics* **51** (1969), 239–246.
22. Samuelson, P. A., "The 'Fallacy' of Maximizing the Geometric Mean in Long Sequences of Investing or Gambling." *Proceedings of the National Academy of Sciences of the United States* **68** (1971), 2493–2496.
23. Schrock, N. W., "The Theory of Asset Choice: Simultaneous Holding of Short and Long Positions in the Futures Market." *Journal of Political Economy* **79**: 2 (1971), 270–293.
24. Shelton, J. P., "The Relation of the Price of a Warrant to its Associated Common Stock." *Financial Analysts Journal* **23**: 3 (1967), 143–151; **23**: 4 (1967), 88–99.
25. Thorp, E., "A Winning Bet in Nevada Baccarat." *Journal of the American Statistical Association* **61** Pt. I (1966), 313–328.
26. Thorp, E., "Optimal Gambling Systems for Favorable Games." *Review of the International Statistical Institute* **37**: 3 (1969), 273–293.
27. Thorp, E., and Kassouf, S., *Beat the Market.* Random House, New York, 1967.
27a. Thorp, E., and Whitley, R. J., "Concave Utilities are Distinguished for Their Optimal Strategies." *Colloquia Mathematica Societatis Janos Bolyai* **9**. European Meeting of Statisticians, Budapest (Hungary), 1972, in press.
28. Todhunter, I., *A History of the Mathematical Theory of Probability*, 1st edition. Cambridge, 1865, as reprinted by Chelsea, New York, 1965. (See pp. 213 ff. for details on Daniel Bernoulli's use of logarithmic utility.)
29. Wagner, H. M., *Principles of Operations Research, with Application to Managerial Decisions*. Prentice-Hall, Englewood Cliffs, New Jersey, 1969.
30. Williams, J. B., "Speculation and the Carryover." *Quarterly Journal of Economics* **50** (1936), 436–455.