The Annals of Probability
1988, Vol. 16, No. 2, 876–898

ASYMPTOTIC OPTIMALITY AND ASYMPTOTIC EQUIPARTITION PROPERTIES OF LOG-OPTIMUM INVESTMENT

BY PAUL H. ALGOET[1] AND THOMAS M. COVER[2]

Boston University and Stanford University

We ask how an investor (with knowledge of the past) should distribute his funds over various investment opportunities to maximize the growth rate of his compounded capital. Breiman (1961) answered this question when the stock returns for successive periods are independent, identically distributed random vectors. We prove that maximizing conditionally expected log return given currently available information at each stage is asymptotically optimum, with no restrictions on the distribution of the market process.

If the market is stationary ergodic, then the maximum capital growth rate is shown to be a constant almost surely equal to the maximum expected log return given the infinite past. Indeed, log-optimum investment policies that at time n look at the n-past are sandwiched in asymptotic growth rate between policies that look at only the k-past and those that look at the infinite past, and the sandwich closes as $k \to \infty$.

1. Introduction. Suppose an investor starts with an initial fortune $S_0 = 1$. At the beginning of each period t (where t takes on discrete values $0, 1, \dots$), the current capital S_t is distributed over investment opportunities $j = 1, \dots, m$ according to some portfolio $b_t = (b_t^j)_{1 \le j \le m}$, a vector of nonnegative weights summing to 1. Let $X_t^j \ge 0$ denote the return per monetary unit allocated to stock j during period t, and $X_t = (X_t^j)_{1 \le j \le m}$ the vector of returns. The yield per unit invested according to portfolio b_t is the weighted average of the return ratios of the individual stocks, i.e., the inner product

$$(b_t, X_t) = \sum_{1 \le j \le m} b_t^j X_t^j. \tag{1}$$

Given that S_t units are invested at the beginning of period t, the total amount collected at the end of the period when the random outcome X_t is revealed is $S_{t+1} = S_t(b_t, X_t)$. This capital is redistributed at the beginning of the next round, and the compounded capital after n investment periods is

$$S_n = \prod_{0 \le t < n} (b_t, X_t). \tag{2}$$

Received September 1985; revised August 1987.

[1]Partially supported by Joint Services Electronics Program Grant DAAG 29-84-K-0047 and National Science Foundation Grant ECS-82-11568.

[2]Partially supported by National Science Foundation Grant ECS-82-11568 and DARPA contract N00039-84-C-0211.

AMS 1980 *subject classifications.* Primary 90A09, 94A15, 28D20; secondary 60F15, 60G40, 49A50.

Key words and phrases. Portfolio theory, gambling, expected log return, log-optimum portfolio, capital growth rate, ergodic stock market, asymptotic optimality principle, asymptotic equipartition property (AEP), Shannon–McMillan–Breiman theorem.

Portfolio b_t must be chosen on the basis of $\mathscr{F}_t$, a σ-field that embodies what is known at the beginning of period t. It obviously makes a difference whether decisions may depend on the history of an aggregate quantity like the Dow–Jones average, on detailed records of the past, or perhaps on inside information or help of a clairvoyant oracle. Our default assumption is that $\mathscr{F}_t = \sigma(X_0, \dots, X_{t-1})$ is the information contained in the past outcomes. We wish to distinguish an optimum strategy $\{b_t^*\}_{0 \le t < \infty}$ among all nonanticipating strategies $\{b_t\}_{0 \le t < \infty}$ such that b_t is $\mathscr{F}_t$-measurable for all $t \ge 0$.

We are dealing with a sequential version of the portfolio selection problem that has received much attention in the literature (not to speak of financial practice). Economic theory promotes the maximization of subjective expected utility as a guiding principle toward its solution, and this is certainly appropriate if the investor's preferences are sufficiently well elucidated so that they can be captured in a well-defined utility function. But subjective utilities are difficult to assess and many investors may prefer a less elusive and more objective criterion if there is some rationale for its use. The mean–variance analysis of Markowitz (1952, 1959) trades off expected return with risk as quantified by the standard deviation of the return. This approach is mathematically and computationally tractable, but it lacks generality [cf. Samuelson (1967, 1970)] and it fails to single out an optimum among the portfolios located on the efficient frontier. However, its economic foundation becomes more solid when cast in the form of the capital asset pricing model [cf. Sharpe (1985)]. Breiman (1960, 1961) considered a market with m stocks and independent, identically distributed discrete-valued return vectors $X_t = (X_t^j)_{1 \le j \le m}$, and proved asymptotic optimality of the portfolio b^* that attains the maximum expected log return $w^* = \sup_b E\{\log(b, X)\}$. Thorp (1971) exhibited certain optimality properties of the log return as a normative utility function, and Bell and Cover (1980, 1986) proved that log-optimum investment is also competitively optimum, from a game-theoretic point of view. Although some authors [e.g., Samuelson (1967, 1971)] have suggested that the log return should be considered just one among many possible utility functions, we hope to convince the reader of its more fundamental character.

We consider arbitrarily distributed outcomes $\{X_t\}$ and prove that maximizing the conditional expected log return given currently available information at each stage is optimum in the long run. A nonanticipating portfolio $b_t^* = b^*(X_0, \dots, X_{t-1})$ is called log-optimum (for period t) if it attains the maximum conditional expected log return

$$(3)\quad w_t^* = \mathbf{E}\{\log(b_t^*, X_t) | \mathscr{F}_t\} = \sup_{b = b(X_0, \dots, X_{t-1})} \mathbf{E}\{\log(b, X_t) | X_{t-1}, \dots, X_0\}.$$

Such b_t^* also attains the maximum (unconditional) expected log return

$$(4)\qquad W_t^* = \mathbf{E}\{w_t^*\} = \mathbf{E}\{\log(b_t^*, X_t)\} = \sup_{b = b(X_0, \dots, X_{t-1})} \mathbf{E}\{\log(b, X_t)\}.$$

A log-optimum portfolio b_t^* always exists, and is unique if the conditional distribution of X_t given $\mathscr{F}_t$ has full support not confined to a hyperplane in $\mathscr{R}^m$. In any case, the return (b_t^*, X_t) is always uniquely defined, even if b_t^* is not.

The results of Breiman (1960, 1961) and Finkelstein and Whitley (1981) for independent, identically distributed $\{X_t\}$ are enhanced by the following theorem, which proves that b_t^* is optimum to first order in growth exponent.

THEOREM (Asymptotic optimality principle). *Let* $S_n^* = \prod_{0 \le t < n}(b_t^*, X_t)$ *and* $S_n = \prod_{0 \le t < n}(b_t, X_t)$, *respectively, denote the capital growth over n periods of investment according to the log-optimum strategy* $\{b_t^*\}_{0 \le t < \infty}$ *and a competing strategy* $\{b_t\}_{0 \le t < \infty}$. *Then* $\{S_n/S_n^*, \mathscr{F}_n\}_{0 \le n < \infty}$ *is a nonnegative supermartingale converging almost surely to a random variable Y with* $\mathbf{E}\{Y\} \le 1$, *and*

$$\limsup_n n^{-1}\log(S_n/S_n^*) \le 0 \quad a.s. \tag{5}$$

Thus $S_n < \exp(n\varepsilon)S_n^*$ *eventually for large n and arbitrary* $\varepsilon > 0$, *which means that no strategy can infinitely often exceed the log-optimum strategy by an amount that grows exponentially fast.*

The asymptotic optimality principle will be deduced from the Kuhn–Tucker conditions for log-optimality using Markov's inequality and the Borel–Cantelli lemma.

Now suppose $\{X_t\}_{-\infty < t < \infty}$ is a two-sided sequence of return vectors, and $\bar{b}_t^* = b^*(X_{-1}, \dots, X_{-t})$ is a log-optimum portfolio for period 0 based on the t-past $\bar{\mathscr{F}}_t = \sigma(X_{-1}, \dots, X_{-t})$. Portfolio $\bar{b}_t^*$ attains the maximum conditional expected log return for period 0 given $\bar{\mathscr{F}}_t$,

$$\bar{w}_t^* = \mathbf{E}\{\log(\bar{b}_t^*, X_0) | \bar{\mathscr{F}}_t\} = \sup_{b = b(X_{-1}, \dots, X_{-t})} \mathbf{E}\{\log(b, X_0) | X_{-1}, \dots, X_{-t}\}. \tag{6}$$

The maximum expected log return for period 0 given $\bar{\mathscr{F}}_t$ is given by

$$\bar{W}_t^* = \mathbf{E}\{\bar{w}_t^*\} = \mathbf{E}\{\log(\bar{b}_t^*, X_0)\} = \sup_{b = b(X_{-1}, \dots, X_{-t})} \mathbf{E}\{\log(b, X_0)\}. \tag{7}$$

The supremum is taken over a larger set of portfolios as t increases, so that $\bar{W}_t^*$ is monotonically increasing and $\{\bar{w}_t^*, \bar{\mathscr{F}}_t\}_{0 \le t < \infty}$ is a submartingale [strictly speaking only if all $\bar{W}_t^*$ are finite].

The information fields $\bar{\mathscr{F}}_t = \sigma(X_{-1}, \dots, X_{-t})$ increase to a limiting σ-field $\bar{\mathscr{F}}_\infty = \sigma(X_{-1}, X_{-2}, \dots)$. Any accumulation point of $\{\bar{b}_t^*\}$ is a log-optimum portfolio for period 0 based on $\bar{\mathscr{F}}_\infty$, and $\bar{b}_t^* = b^*(X_{-1}, \dots, X_{-t})$ almost surely converges to $\bar{b}_\infty^* = b^*(X_{-1}, X_{-2}, \dots)$ if the log-optimum portfolio for period 0 given $\bar{\mathscr{F}}_\infty$ is unique. Furthermore, $\bar{W}_t^*$ increases to the maximum expected log return given the infinite past,

$$\bar{W}_t^* \nearrow \bar{W}_\infty^* = \mathbf{E}\{\log(\bar{b}_\infty^*, X_0)\} = \sup_{b = b(X_{-1}, X_{-2}, \dots)} \mathbf{E}\{\log(b, X_0)\}. \tag{8}$$

We may use the expanded notation $W_t^* = W^*(X_t | X_{t-1}, \dots, X_0)$ and $\bar{W}_t^* = W^*(X_0 | X_{-1}, \dots, X_{-t})$. Setting $\mathbf{E}\{\log S_n^*\} = W^*(X_0, \dots, X_{n-1})$ yields the chain rule

$$W^*(X_0, \dots, X_{n-1}) = \sum_{0 \le t < n} W^*(X_t | X_{t-1}, \dots, X_0). \tag{9}$$

If $\{X_t\}$ is stationary, then $\overline{W}_t^* = W^*(X_0|X_{-1},\dots,X_{-t})$ is equal to $W_t^* = W^*(X_t|X_{t-1},\dots,X_0)$ and these definitions are equivalent:

$$\begin{aligned}\overline{W}_\infty^* = W^*(X_0|X_{-1},X_{-2},\dots) &= \lim_t \uparrow W^*(X_0|X_{-1},\dots,X_{-t})\\ &= \lim_t \uparrow W^*(X_t|X_{t-1},\dots,X_0)\\ &= \lim_n \uparrow n^{-1}W^*(X_0,X_1,\dots,X_{n-1}).\end{aligned} \tag{10}$$

These identities for maximum capital growth rate generalize those for relative entropy rate in information theory. Indeed, suppose one stock will return m times the amount invested in it, whereas all other stocks return 0. Thus we must gamble against uniform odds, on the identity of the winning stock (indicated by the direction of X_t). Placing proportional bets, $b_t^j = \mathrm{Prob}\{X_t^j \neq 0|X_{t-1},\dots,X_0\}$ is log-optimum, and $W^*(X_t|X_{t-1},\dots,X_0) = \log m - H^*(X_t|X_{t-1},\dots,X_0)$, where $H^*(X_t|X_{t-1},\dots,X_0)$ is the conditional entropy of X_t given $X_{t-1},\dots,X_0$. Now $H^*(X_0|X_{-1},X_{-2},\dots) = \lim_t \downarrow H^*(X_t|X_{t-1},\dots,X_0)$ is the entropy rate of $\{X_t\}$, and

$$W^*(X_0|X_{-1},X_{-2},\dots) = \log m - H^*(X_0|X_{-1},X_{-2},\dots). \tag{11}$$

The following AEP for log-optimum investment in a stationary ergodic market generalizes the Shannon–McMillan–Breiman theorem of information theory.

THEOREM (Asymptotic equipartition property or AEP). *If $\{X_t\}$ is stationary ergodic, then $S_n^* = \prod_{0\le t<n}(b_t^*, X_t)$ grows exponentially fast with constant asymptotic rate almost surely equal to the maximum expected log return given the infinite past, i.e.,*

$$n^{-1}\log S_n^* \to \overline{W}_\infty^* = W^*(X_0|X_{-1},X_{-2},\dots) \quad a.s. \tag{12}$$

Equivalently, $S_n^* = \exp[n(\overline{W}_\infty^* + o(1))]$, where $o(1) \to 0$ a.s. The rate $\overline{W}_\infty^*$ is highest possible.

The AEP is an immediate consequence of the ergodic theorem if $\{X_t\}$ is finite order Markov. A sandwich argument and the asymptotic optimality principle will reduce the proof of the general case to applications of the ergodic theorem.

In the first half of the paper we discuss log-optimum investment for a single period. The Kuhn–Tucker conditions for log-optimality of a portfolio b^* are recalled in Section 2, and in Section 3 we examine log-optimum portfolio selections and the maximum expected log return as functions of the distribution P of the random outcome $X = (X^j)_{1\le j\le m}$ on $\mathscr{R}_+^m$. To simplify the analysis we use a divide-and-conquer approach. Namely, we consider the decomposition $X = (\beta, X)U$, where $\beta = (\beta^j)_{1\le j\le m}$ is a fixed reference portfolio and $U = X/(\beta, X)$ is the scaled outcome in the simplex $\mathscr{U} = \{u = (u^j)_{1\le j\le m} \in \mathscr{R}_+^m: (\beta, u) = 1\}$. The return (b, X) factors as $(\beta, X)(b, U)$, and the maximum expected log return $w^*(P) = \sup_b E_P\{\log(b, X)\}$ decomposes as the sum of a reference level $r(P) = E_P\{\log(\beta, X)\}$ that is affine in P and an extra term $w^*(Q) = \sup_b E_Q\{\log(b, U)\}$ that depends on P only through the marginal distribution Q of U. The term $w^*(Q)$ is nonnegative, bounded and continuous in Q when the space of probability measures on $\mathscr{U}$ is equipped with the weak topology, whereas the irregular term $r(P) = E_P\{\log(\beta, X)\}$ is irrelevant for portfolio selection.

We need the decomposition $w^*(P) = r(P) + w^*(Q)$ to show that the maximum conditional expected log return w_t^* is always attained by an $\mathscr{F}_t$-measurable portfolio b_t^*. Furthermore, the nonnegativity and lower semicontinuity of $w^*(Q)$ are essential in Section 4 when we argue that the maximum expected log return given the t-past converges to the maximum expected log return given the infinite past (i.e., $\overline{W}_t^* \nearrow \overline{W}_\infty^*$ at $t \to \infty$).

The asymptotic optimality principle is proved in Section 5, for an arbitrarily distributed sequence of return vectors. In Section 6 we argue that S_n^* has a well-defined growth rate if $\{X_t\}$ is stationary ergodic, and in Section 7 we examine whether the same is true if the market is stationary, or stationary in an asymptotic sense. Although the ergodic theorem is generally valid for asymptotically mean stationary processes (whose definition is recalled in Section 7), the AEP will hold for an asymptotically mean stationary market only if the investor can recover from transient losses before reaching the asymptotic regime. Finally, in Section 8 we specialize the investment game to gambling on the next outcome of a random process.

2. The Kuhn–Tucker conditions for log-optimality. When managing funds during a given investment period, an investor may diversify his risk by building a portfolio that includes several assets. The allocation of one unit of capital over elementary investment opportunities $j = 1, \dots, m$ is conveniently described by a vector of weights $b = (b^j)_{1 \le j \le m}$. The weights must be nonnegative (since no borrowing is allowed) and sum to 1. Thus a portfolio is a vector b in the unit simplex

$$\mathscr{B} = \left\{ b = (b^j)_{1 \le j \le m} \in \mathscr{R}_+^m \colon b^1 + \cdots + b^m = 1 \right\}. \tag{13}$$

Let $X^j \ge 0$ denote the return per monetary unit invested in stock j, and let $X = (X^j)_{1 \le j \le m}$ denote the vector of returns. Capital invested according to portfolio b will grow by the factor $(b, X) = \sum_{1 \le j \le m} b^j X^j$, that is, the weighted average of the per-unit returns of the individual stocks. Portfolio b must be selected at the beginning of the investment period, before the actual value of the random outcome X is revealed. However, the distribution of X on $\mathscr{R}_+^m$ is assumed to be known.

Let the expected log return of a portfolio b be denoted by

$$w(b) = E\{\log(b, X)\}. \tag{14}$$

We set $w(b) = -\infty$ if the expectation is not well defined in the usual sense.

DEFINITION. A portfolio b^* is called *log-optimum* if no competing portfolio b can improve the expected log return relative to b^*, i.e., if

$$E\left\{\log\left(\frac{(b, X)}{(b^*, X)}\right)\right\} \le 0, \quad \text{for all } b \in \mathscr{B}. \tag{15}$$

Every log-optimum portfolio b^* attains the maximum expected log return

$$w^* = \sup_{b \in \mathscr{B}} E\{\log(b, X)\}. \tag{16}$$

Conversely, if w^* is finite, then every portfolio b^* attaining $w^* = \sup_b w(b)$ is log-optimum. However, condition (15) may single out a unique log-optimum portfolio b^* even if $w(b)$ is infinite for all $b \in \mathscr{B}$.

We recall the Kuhn–Tucker conditions for log-optimality derived in Bell and Cover (1980). Let the expected score vector be defined for each portfolio b as

$$\alpha(b) = E\{X/(b, X)\}. \tag{17}$$

THEOREM 1. *Let $\alpha^* = \alpha(b^*)$ denote the expected score vector for portfolio b^*. Then b^* is log-optimum iff the Kuhn–Tucker conditions $\alpha^{*j} \le 1$ hold for all $1 \le j \le m$, or equivalently, iff*

$$(b, \alpha^*) = E\left\{\frac{(b, X)}{(b^*, X)}\right\} \le 1, \quad \textit{for all } b \in \mathscr{B}. \tag{18}$$

PROOF. For $b \in \mathscr{B}$ and $0 < \bar{\lambda} = 1 - \lambda < 1$ let $b_\lambda = \bar{\lambda} b^* + \lambda b$. Then

$$\frac{(b_\lambda, X)}{(b^*, X)} = \bar{\lambda} + \lambda \frac{(b, X)}{(b^*, X)} = 1 + \lambda Z, \quad \text{where } Z = \frac{(b, X)}{(b^*, X)} - 1.$$

Using a Taylor series expansion we obtain, for any $a > 0$,

$$\begin{aligned}\lambda Z \ge \log(1 + \lambda Z) &\ge \log(1 + \lambda(Z \wedge a)) \\ &= \lambda(Z \wedge a) - \tfrac{1}{2}\theta\lambda^2(Z \wedge a)^2 \quad (\text{for some } 0 < \theta < 1) \\ &\ge \lambda(Z \wedge a) - \tfrac{1}{2}\lambda^2 a^2.\end{aligned}$$

Choosing $a = a(\lambda)$ so that $a(\lambda) \to \infty$ and $\lambda a(\lambda) \to 0$ as $\lambda \searrow 0$, we see that $\lambda^{-1}E\{\log(1 + \lambda Z)\} \to E\{Z\}$ as $\lambda \searrow 0$. But $E\{Z\} = (b, \alpha^*) - 1$, so the right derivative at $\lambda = 0$ of $w(b_\lambda) = E\{\log(b_\lambda, X)\}$ is given by

$$\left.\frac{d}{d\lambda} w(b_\lambda)\right|_{\lambda = 0+} = \lim_{\lambda \searrow 0} \frac{E\{\log(1 + \lambda Z)\}}{\lambda} = E\{Z\} = (b, \alpha^*) - 1. \tag{19}$$

The Kuhn–Tucker conditions assert that b^* is log-optimum iff the directional derivative of the expected log return is nonpositive when moving from b^* to any competing portfolio b (in particular, when moving from b^* to any extreme point of $\mathscr{B}$). The infinitesimal conditions $dw(b_\lambda)/d\lambda|_{\lambda=0+} \le 0$ are necessary for log-optimality of b^*, and they are also sufficient because $w(b)$ is concave in b. □

The set B^* of log-optimum portfolios is never empty [cf. Cover (1984)]. In fact, let $\mathscr{L}$ denote the linear hull of the support of the distribution of X, that is, the smallest linear subspace of $\mathscr{R}^m$ such that $X \in \mathscr{L}$ with probability 1. Then $w(b)$ is strictly concave when restricted to $\mathscr{L}$ and constant along fibers perpendicular to $\mathscr{L}$. It follows that B^* is a polyhedral set (the intersection of $\mathscr{B}$ with a fiber orthogonal to $\mathscr{L}$), and the log-optimum portfolio b^* is unique if X has full support ($\mathscr{L} = \mathscr{R}^m$). The return (b^*, X) and log return $\log(b^*, X)$ are unambiguously defined, independent of the choice of log-optimum portfolio b^* in B^*.

3. Continuity and attainability of the maximum expected log return. We make explicit how various quantities depend on the distribution P of X on $\mathscr{R}_+^m$. Let $w(b, P) = E_P\{\log(b, X)\}$ denote the expected log return of portfolio b, $w^*(P) = \sup_b w(b, P)$ the maximum expected log return and $B^*(P)$ the set of log-optimum portfolios. It is clear that $w^*(P)$ is convex in P, since $w^*(P)$ is the supremum of functions $E_P\{\log(b, X)\}$ that are affine in P.

The direction of the return vector X embodies everything an investor needs to know in order to maximize the expected log return. To justify this claim, we choose a fixed reference portfolio $\beta = (\beta^j)_{1 \le j \le m}$ with $\beta^j > 0$ for all j, and we define the scaled return vector

$$U = u(X), \quad \text{where } u(x) = x/(\beta, x). \tag{20}$$

Thus U is obtained by projecting the return vector X on the simplex

$$\mathscr{U} = \left\{u = (u^j)_{1 \le j \le m} \in \mathscr{R}_+^m : (\beta, u) = 1\right\}. \tag{21}$$

If $X = 0$, then we set $U = u(0) = u_0$ for some arbitrary $u_0 \in \mathscr{U}$.

The distribution Q of $U = u(X)$ on $\mathscr{U}$ is obtained by integrating out the distribution P of X along rays through the origin. All mass accumulated along a ray is collected at the point where the ray crosses the simplex $\mathscr{U}$, except that mass found at $X = 0$ is transferred to $u(0) = u_0$. Thus Q is the image measure of P through u: $\mathscr{R}_+^m \to \mathscr{U}$, and for any Borel subset $A \subseteq \mathscr{U}$ we have

$$Q\{U \in A\} = P\{u(X) \in A\} = P\left(u^{-1}(A)\right). \tag{22}$$

Since $X = (\beta, X)u(X)$, the expected log return may be decomposed as the sum $E_P\{\log(b, X)\} = E_P\{\log(\beta, X)\} + E_P\{\log(b, u(X))\}$, or equivalently,

$$w(b, P) = r(P) + w(b, Q). \tag{23}$$

Here $r(P) = w(\beta, P)$ denotes the expected log return of the reference portfolio β. We interpret $r(P)$ as a reference level for the expected log return, since it is an inherent property of the market over which the investor has no control. Whereas $r(P) = E_P\{\log(\beta, X)\}$ is affine in P, it is also a very irregular function of P, possibly infinite or ill defined. Since our choice of b cannot affect its value, we shall subtract $r(P)$ from the expected log return $w(b, P)$. The remaining quantity $w(b, Q) = E_Q\{\log(b, U)\}$ depends on P only through the marginal distribution Q of the scaled outcome U, and represents the relative improvement in expected log return that results when portfolio b is chosen instead of β. The maximum expected log return can be expressed as the sum

$$w^*(P) = r(P) + w^*(Q), \quad \text{where } w^*(Q) = \sup_{b\in\mathscr{B}} E_Q\{\log(b, U)\}. \tag{24}$$

Maximizing $w(b, P)$ or $w(b, Q)$ are equivalent operations, so that $B^*(P) = B^*(Q)$. Notice that $w^*(Q) = w^*(P) - r(P) \geq 0$, with equality iff the reference portfolio β is log-optimum.

It is an interesting fact that the maximum expected log return $w^*(Q)$ is always attained by some portfolio choice. However, we need a stronger result, namely, the existence of log-optimum portfolios $b^*(Q)$ that depend measurably on Q. To prove the existence of a measurable selection of log-optimum portfolios, we make use of topological properties, including compactness of $\mathscr{B}$ and upper semicontinuity of the expected log return $w(b, Q) = E_Q\{\log(b, U)\}$ in (b, Q).

The space $\mathscr{Q}$ of probability measures on the compact metric space $\mathscr{U}$ is compact and metrizable when equipped with the weak topology [that is the weakest topology on $\mathscr{Q}$ such that $Q \mapsto E_Q\{f(U)\}$ is continuous in $Q \in \mathscr{Q}$ for all bounded continuous functions $f\colon \mathscr{U} \to \mathscr{R}$]. Its Borel σ-field is the smallest σ-field on $\mathscr{Q}$ such that $A \mapsto Q(A)$ is measurable in Q for all Borel subsets $A \subseteq \mathscr{U}$.

THEOREM 2. *The maximum expect log return $w^*(Q) = \sup_{b\in\mathscr{B}} E_Q\{\log(b, U)\}$ is convex, bounded [between 0 and $\max_j(-\log \beta^j)$] and uniformly continuous when the space $\mathscr{Q}$ of probability measures on $\mathscr{U}$ is equipped with the weak topology. The set of log-optimum portfolios $B^*(Q)$ is a nonempty compact convex subset of $\mathscr{B}$ for every distribution Q on $\mathscr{U}$, and a log-optimum portfolio $b^*(Q) \in B^*(Q)$ can be selected for each $Q \in \mathscr{Q}$ so that $b^*(Q)$ is measurable in Q.*

PROOF. Clearly $w^*(Q)$ is convex in Q for the same reason that $w^*(P)$ is convex in P. We argue that $w^*(Q)$ is bounded below and lower semicontinuous on $\mathscr{Q}$, because (β, u) is bounded below on $\mathscr{U}$ and (b, u) is concave in $b \in \mathscr{B}$ and lower semicontinuous in $u \in \mathscr{U}$. We also prove that $w(b, Q)$ is bounded above and upper semicontinuous, using compactness of $\mathscr{B}$ and boundedness above and upper semicontinuity of (b, u) on $\mathscr{B} \times \mathscr{U}$. Boundedness and uniform continuity of $w^*(Q)$ and existence of a measurable selection of log-optimum portfolios $b^*(Q)$ will follow automatically.

First, we argue that $w^*(Q)$ is nonnegative and lower semicontinuous on $\mathscr{Q}$. For $0 \leq \lambda \leq 1$ and $b \in \mathscr{B}$, let $\bar{\lambda} = 1 - \lambda$, $b_\lambda = \bar{\lambda}\beta + \lambda b$, $\mathscr{B}_\lambda = \{b_\lambda\colon b \in \mathscr{B}\}$ and

$$w_\lambda^*(Q) = \sup_{b\in\mathscr{B}_\lambda} E_Q\{\log(b, U)\} = \sup_{b\in\mathscr{B}} E_Q\{\log(b_\lambda, U)\}. \tag{25}$$

Observe that $w_\lambda^*(Q)$ is monotonically increasing in λ, since the supremum is taken over a larger set $\mathscr{B}_\lambda$ as λ increases. Furthermore, $\mathscr{B}_1 = \mathscr{B}$, $\mathscr{B}_0 = \{\beta\}$ and $(\beta, u) = 1$ for all $u \in \mathscr{U}$, so that

$$w^*(Q) = w_1^*(Q) \geq w_\lambda^*(Q) \geq w_0^*(Q) = E_Q\{\log(\beta, U)\} = 0.$$

If $\lambda < 1$, then $\log(b_\lambda, u)$ is bounded below (by $\log \bar{\lambda}$) and lower semicontinuous in u, so that $w(b_\lambda, Q) = E_Q\{\log(b_\lambda, U)\}$ and $w_\lambda^*(Q) = \sup_{b\in\mathscr{B}} w(b_\lambda, Q)$ are lower semicontinuous in Q. On the other hand, the inequality $(b_\lambda, u) \geq \lambda(b, u)$

implies that

$$w_\lambda^*(Q) \le w^*(Q) \le w_\lambda^*(Q) - \log \lambda,$$

and hence $w_\lambda^*(Q) \nearrow w^*(Q)$ as $\lambda \nearrow 1$. Since $w^*(Q)$ is the supremum of lower semicontinuous functions $w_\lambda^*(Q)$, it follows that $w^*(Q)$ is lower semicontinuous as well.

The expected log return $w(b, Q) = E_Q\{\log(b, U)\}$ is bounded above and upper semicontinuous on $\mathscr{B} \times \mathscr{Q}$, since (b, u) is bounded and upper semicontinuous on $\mathscr{B} \times \mathscr{U}$. Since $\mathscr{B}$ is compact, it follows [cf. Bertsekas and Shreve (1978), Proposition 7.33] that $w^*(Q) = \sup_{b \in \mathscr{B}} w(b, Q)$ is bounded above and upper semicontinuous on $\mathscr{Q}$. Furthermore, log-optimum portfolios $b^*(Q) \in B^*(Q)$ can be selected in a measurable fashion for all $Q \in \mathscr{Q}$ by the measurable selection theorem of Kuratowski and Ryll-Nardzewski (1961). The upper bound $w^*(Q) \le \max_j(-\log \beta^j)$ holds since $(b, u) \le \sum_j u^j \le \max_j(1/\beta^j)$ for all $b \in \mathscr{B}$ if u satisfies $(\beta, u) = 1$. □

It is impossible to select a portfolio $b^*(Q) \in B^*(Q)$ for all distributions Q on $\mathscr{U}$ so that $b^*(Q)$ is continuous in Q. However, if $\overline{Q}_n \to \overline{Q}_\infty$ and $\bar{b}_n^* \in B^*(\overline{Q}_n)$ for all n, then any accumulation point $\bar{b}_\infty^*$ of the sequence $\{\bar{b}_n^*\}$ is a point in $B^*(\overline{Q}_\infty)$. Furthermore, $(\bar{b}_n^*, U) \to (\bar{b}_\infty^*, U)$ almost surely under $\overline{Q}_\infty$. These continuity properties of the multivalued correspondence $Q \mapsto B^*(Q)$ follow from the following.

THEOREM 3. *The set $Gr(B^*) = \{(Q, b^*)\colon\ b^* \in B^*(Q)\}$ is closed in $\mathscr{Q} \times \mathscr{B}$. Consequently, any selection of log-optimum portfolios $Q \mapsto b^*(Q) \in B^*(Q)$ is continuous at any $Q \in \mathscr{Q}$ such that $B^*(Q) = \{b^*(Q)\}$ is a singleton set.*

PROOF. Since $\mathscr{B}$ is compact, the theorem will follow from the following claim: If $\overline{Q}_n \to \overline{Q}_\infty$ in $\mathscr{Q}$, $\bar{b}_n^* \to \bar{b}_\infty^*$ in $\mathscr{B}$ and $\bar{b}_n^* \in B^*(\overline{Q}_n)$ for all n, then $\bar{b}_\infty^* \in B^*(\overline{Q}_\infty)$.

To prove the claim we consider the sequence of maximum expected log returns $w^*(\overline{Q}_n) = w(\bar{b}_n^*, \overline{Q}_n)$. It is clear that $w^*(\overline{Q}_n) \to w^*(\overline{Q}_\infty)$ since $\overline{Q}_n \to \overline{Q}_\infty$ in $\mathscr{Q}$ and $w^*(Q)$ is continuous in Q. On the other hand (see the proof of Theorem 2), $w(b, Q)$ is upper semicontinuous on $\mathscr{B} \times \mathscr{Q}$ and hence

$$\limsup_n w\big(\bar{b}_n^*, \overline{Q}_n\big) \le w\Big(\lim_n \bar{b}_n^*, \lim_n \overline{Q}_n\Big) = w\big(\bar{b}_\infty^*, \overline{Q}_\infty\big).$$

The claim $\bar{b}_\infty^* \in B^*(\overline{Q}_\infty)$ and Theorem 3 follow, since

$$w\big(\bar{b}_\infty^*, \overline{Q}_\infty\big) \ge \limsup_n w\big(\bar{b}_n^*, \overline{Q}_n\big) = \lim_n w^*\big(\overline{Q}_n\big) = w^*\big(\overline{Q}_\infty\big) = \sup_b w\big(b, \overline{Q}_\infty\big). \quad □$$

The maximum expected log return $w^*(P)$ is neither bounded nor continuous (for the weak topology) as P ranges over the space of probability measures on $\mathscr{R}_+^m$. But if the support of P is constrained to a closed subset $\mathscr{K}$ of $\mathscr{R}_+^m$, then $w^*(P)$ is lower semicontinuous and bounded below iff $\mathscr{K}$ is bounded away from 0, upper semicontinuous and bounded above iff $\mathscr{K}$ is bounded, and bounded and uniformly continuous iff $\mathscr{X}$ is bounded away from 0. In particular, if $\mathscr{X} = \mathscr{U}$ (i.e., if X is distributed on the simplex $\mathscr{U}$), then $P = Q$ and $w^*(P) = w^*(Q)$ is bounded and continuous.

Some of the conclusions of Theorems 2 and 3 continue to hold if the investor may distribute his funds over a countable set or even a separable metrizable space $\mathscr{A}$ of investment opportunities. Indeed, suppose every realization of the return X is a nonnegative lower semicontinuous function $x(a)$ on $\mathscr{A}$. (This is no restriction if $\mathscr{A}$ is a countable set with the discrete topology.) The average return $(b, x) = \int_{\mathscr{A}} x(a) b(da)$ is then well defined for every portfolio b [i.e., for every normalized measure $b(da)$ on the Borel σ-field of $\mathscr{A}$]. Further assume the existence of a reference portfolio β such that $(\beta, x) > 0$ is strictly positive for any return function $x(a)$ that is not identically 0. [Such β exists if $\mathscr{A}$ is locally compact, and, in particular, if $\mathscr{A}$ is countable.] If P and Q denote the distribution of X and $U = X/(\beta, X)$, then the maximum expected log return $w^*(P)$ admits the decomposition $r(P) + w^*(Q)$, and $w^*(Q)$ is nonnegative and lower semicontinuous by the argument presented in the proof of Theorem 2. If, moreover, $\mathscr{A}$ is compact and the return functions $x(a)$ are continuous and bounded by a fixed constant, then $w^*(Q)$ is bounded and continuous and a measurable selection of log-optimum portfolios $b^*(Q)$ exists by Theorem 2, and $Gr(B^*)$ is closed by Theorem 3.

4. Martingale properties. It will be shown that the maximum expected log return given increasing information fields tends to the maximum expected log return given the limiting σ-field. We assume that the random return vector $X(\omega) \in \mathscr{R}_+^m$ is defined on a perfect probability space $(\Omega, \mathscr{F}, \mathbf{P})$, so that X admits a regular conditional probability distribution given any sub-σ-field of $\mathscr{F}$. See Jiřina (1954) for a proof of this fact, and Ramachandran (1979) for a complete discussion of perfect measures.

THEOREM 4. *Suppose the random vector X is defined on a perfect probability space $(\Omega, \mathscr{F}, \mathbf{P})$, and $\{\bar{\mathscr{F}}_t\}_{0 \le t < \infty}$ is an increasing sequence of sub-σ-fields of $\mathscr{F}$ with limiting σ-field $\bar{\mathscr{F}}_\infty \subseteq \mathscr{F}$.*

(a) *If $\bar{P}_t$ is a regular conditional probability distribution of X given $\bar{\mathscr{F}}_t$, then*

$$\bar{P}_t \to \bar{P}_\infty \quad \textit{weakly a.s.} \tag{26}$$

(b) *If $b^*(\cdot)$ is a measurable selector of log-optimum portfolios, then $\bar{b}_t^* = b^*(\bar{P}_t)$ is an $\bar{\mathscr{F}}_t$-measurable portfolio attaining the maximum conditional expected log return given $\bar{\mathscr{F}}_t$. Moreover, $(\bar{b}_t^*, X) \to (\bar{b}_\infty^*, X)$ a.s., and hence*

$$\log(\bar{b}_t^*, X) \to \log(\bar{b}_\infty^*, X) \quad a.s. \tag{27}$$

If the log-optimum portfolio given $\bar{\mathscr{F}}_\infty$ is unique $[B^*(\bar{P}_\infty) = \{\bar{b}_\infty^*\}]$, then $\bar{b}_t^* \to \bar{b}_\infty^*$ a.s. as well.

(c) *If $w^*(\cdot)$ denotes the maximum expected log return function, then the maximum conditional expected log return given $\bar{\mathscr{F}}_t$ is given by*

$$\bar{w}_t^* = w^*(\bar{P}_t) = \sup_{b \in \bar{\mathscr{F}}_t} \mathbf{E}\{\log(b, X) | \bar{\mathscr{F}}_t\} = \mathbf{E}\{\log(\bar{b}_t^*, X) | \bar{\mathscr{F}}_t\} \quad a.s. \tag{28}$$

Furthermore, $\{\bar{w}_t^*, \bar{\mathscr{F}}_t\}_{0 \le t \le \infty}$ *is a submartingale and*

(29) $$\bar{w}_t^* \to \bar{w}_\infty^* \quad a.s. \left(\text{and in } L^1 \text{ if } \overline{W}_\infty^* < \infty\right).$$

(d) *The maximum expected log return given* $\bar{\mathscr{F}}_t$ *is given by*

(30) $$\overline{W}_t^* = \mathbf{E}\{\bar{w}_t^*\} = \sup_{b \in \bar{\mathscr{F}}_t} \mathbf{E}\{\log(b, X)\} = \mathbf{E}\{\log(\bar{b}_t^*, X)\}.$$

Furthermore,

(31) $$\overline{W}_t^* \nearrow \overline{W}_\infty^*, \quad \text{as } t \to \infty.$$

PROOF. Lévy's martingale convergence theorem for conditional expectations of a bounded continuous (or nonnegative measurable) function $f(x)$ states that

$$\int f\, d\bar{P}_t = \mathbf{E}\{f(X)|\bar{\mathscr{F}}_t\} \to \int f\, d\bar{P}_\infty = \mathbf{E}\{f(X)|\bar{\mathscr{F}}_\infty\} \quad \text{a.s.}$$

This proves (a), and assertion (b) follows in view of Theorem 3. Notice that $\bar{b}_t^* = b^*(\bar{P}_t)$ and $\bar{w}_t^* = w^*(\bar{P}_t)$ are $\bar{\mathscr{F}}_t$-measurable, since $\bar{P}_t$ is measurable on $(\Omega, \bar{\mathscr{F}}_t)$ and both $b^*(\cdot)$ and $w^*(\cdot)$ are measurable functions.

If $0 \le s \le t \le \infty$, then $\bar{\mathscr{F}}_s \subseteq \bar{\mathscr{F}}_t$, so that every $\bar{\mathscr{F}}_s$-measurable portfolio (including $\bar{b}_s^*$) is also $\bar{\mathscr{F}}_t$-measurable. It follows that

$$\mathbf{E}\{\log(\bar{b}_s^*, X)|\bar{\mathscr{F}}_t\} \le \bar{w}_t^* = \sup_{b \in \bar{\mathscr{F}}_t} \mathbf{E}\{\log(b, X)|\bar{\mathscr{F}}_t\}.$$

Taking $\bar{\mathscr{F}}_s$-conditional expectations proves that $\bar{w}_s^* = \mathbf{E}\{\log(\bar{b}_s^*, X)|\bar{\mathscr{F}}_s\} \le \mathbf{E}\{\bar{w}_t^*|\bar{\mathscr{F}}_s\}$, and hence $\{\bar{w}_t^*, \bar{\mathscr{F}}_t\}_{0 \le t \le \infty}$ is a submartingale. The maximum expected log returns $\overline{W}_t^* = \sup_{b \in \bar{\mathscr{F}}_t} \mathbf{E}\{\log(b, X)\}$ increase with t since the supremum is taken over larger and larger sets ($b \in \bar{\mathscr{F}}_s \Rightarrow b \in \bar{\mathscr{F}}_t$). More information does not hurt!

It remains to show that $\overline{W}_t^* \nearrow \overline{W}_\infty^*$ and $\bar{w}_t^* \to \bar{w}_\infty^*$ a.s. (and in L^1 if $\overline{W}_\infty^*$ is finite). For this purpose we choose a reference portfolio β (with $\beta^j > 0$ for all $1 \le j \le m$), and we recall the decomposition $w^*(P) = r(P) + w^*(Q)$ of the maximum expected log return into a reference level $r(P) = E_P\{\log(\beta, X)\}$ and a relative improvement $w^*(Q)$ that only depends on the distribution Q of the scaled return vector $U = u(X) = X/(\beta, X)$.

Let $\bar{Q}_t$ designate a regular conditional probability distribution of $U = u(X)$ given $\bar{\mathscr{F}}_t$, for $0 \le t \le \infty$. Then $\bar{Q}_t \to \bar{Q}_\infty$ weakly almost surely and $\{w^*(\bar{Q}_t), \bar{\mathscr{F}}_t\}_{0 \le t \le \infty}$ is a submartingale. Since $w^*(Q)$ is bounded and continuous in Q, it follows that $w^*(\bar{Q}_t) \to w^*(\bar{Q}_\infty)$ a.s. and in L^1, and $\mathbf{E}\{w^*(\bar{Q}_t)\} \nearrow \mathbf{E}\{w^*(\bar{Q}_\infty)\}$. The sequence $\{r(\bar{P}_t), \bar{\mathscr{F}}_t\}_{0 \le t \le \infty}$ [where $r(\bar{P}_t) = \mathbf{E}\{\log(\beta, X)|\bar{\mathscr{F}}_t\}$] is a martingale [at least if $\log(\beta, X)$ has finite expectation], and the martingale convergence theorem for conditional expectations asserts that

$$r(\bar{P}_t) = \mathbf{E}\{\log(\beta, X)|\bar{\mathscr{F}}_t\} \to r(\bar{P}_\infty) = \mathbf{E}\{\log(\beta, X)|\bar{\mathscr{F}}_\infty\} \quad \text{a.s.}$$

[and in L^1 if $\mathbf{E}\{\log(\beta, X)\}$ is finite]. Since $\bar{w}_t^* = r(\bar{P}_t) + w^*(\bar{Q}_t)$, we may conclude that $\{\bar{w}_t^*, \bar{\mathscr{F}}_t\}_{0 \le t < \infty}$ is a submartingale such that $\bar{w}_t^* \to \bar{w}_\infty^*$ a.s. (and in

L^1 if $\overline{W}_\infty^*$ is finite). The expectations satisfy

$$\overline{W}_t^* = \mathbf{E}\{\log(\beta, X)\} + \mathbf{E}\{w^*(\overline{Q}_t)\} \nearrow \overline{W}_\infty^* = \mathbf{E}\{\log(\beta, X)\} + \mathbf{E}\{w^*(\overline{Q}_\infty)\}.$$

□

The main conclusion of Theorem 4 is that no gap exists between $\lim_t \uparrow \overline{W}_t^*$ and $\overline{W}_\infty^*$. Thus the limit of the expectations $\overline{W}_t^* = \mathbf{E}\{\log(\bar{b}_t^*, X)\}$ coincides with $\overline{W}_\infty^* = \mathbf{E}\{\log(\bar{b}_\infty^*, X)\}$, which is the expectation of the limit $\log(\bar{b}_\infty^*, X) = \lim_t \log(\bar{b}_t^*, X)$.

We have shown that $\overline{w}_t^* \to \overline{w}_\infty^*$ a.s. and $\overline{W}_t^* \nearrow \overline{W}_\infty^*$, using boundedness and continuity of $w^*(Q)$. These convergence theorems also hold for a market with infinitely many investment opportunities, when $w^*(Q)$ is only nonnegative and lower semicontinuous. Indeed, $\{\overline{w}_t^*, \overline{\mathscr{F}}_t\}_{0 \le t \le \infty}$ is still a submartingale, so that $\overline{w}_t^* \le \mathbf{E}\{\overline{w}_\infty^* | \overline{\mathscr{F}}_t\}$ and $\lim_t \uparrow \mathbf{E}\{\overline{w}_t^*\} \le \mathbf{E}\{\overline{w}_\infty^*\}$ and hence, by Lévy's martingale convergence theorem for conditional expectations,

$$\limsup_t \overline{w}_t^* \le \lim_t \mathbf{E}\{\overline{w}_\infty^* | \overline{\mathscr{F}}_t\} = \mathbf{E}\{\overline{w}_\infty^* | \overline{\mathscr{F}}_\infty\} = \overline{w}_\infty^* \quad \text{a.s.} \tag{32}$$

Since $\{w^*(\overline{Q}_t), \overline{\mathscr{F}}_t\}_{0 \le t \le \infty}$ is a submartingale also, one similarly obtains

$$\begin{aligned} &\limsup_t w^*(\overline{Q}_t) \le w^*(\overline{Q}_\infty) \quad \text{a.s.} \\ \text{and} \quad &\lim_t \uparrow \mathbf{E}\{w^*(\overline{Q}_t)\} \le \mathbf{E}\{w^*(\overline{Q}_\infty)\}. \end{aligned} \tag{33}$$

But $\overline{Q}_t \to \overline{Q}_\infty$ weakly a.s. and $w^*(Q)$ is lower semicontinuous in Q, so that

$$\liminf_t w^*(\overline{Q}_t) \ge w^*(\overline{Q}_\infty) \quad \text{a.s.} \tag{34}$$

We conclude that $w^*(\overline{Q}_t) \to w^*(\overline{Q}_\infty)$ a.s. Since $w^*(Q)$ is also nonnegative Fatou's lemma implies that $\mathbf{E}\{w^*(\overline{Q}_t)\} \nearrow \mathbf{E}\{w^*(\overline{Q}_\infty)\}$. It follows that $\overline{w}_t^* \to \overline{w}_\infty^*$ a.s. and $\overline{W}_t^* \nearrow \overline{W}_\infty^*$, at least if $\mathbf{E}\{\log(\beta, X_0)\} > -\infty$ or $\sup_k \mathbf{E}\{w^*(\overline{Q}_k)\} < \infty$.

5. The asymptotic optimality principle. We now prove the asymptotic optimality principle for sequential log-optimum investment. The market is described by a sequence of return vectors $\{X_t\}_{0 \le t < \infty}$ defined on a perfect probability space $(\Omega, \mathscr{F}, \mathbf{P})$, and capital invested according to a portfolio b_t at the beginning of period t will grow by a factor (b_t, X_t) when the random outcome X_t is revealed at the end of that period. If the initial fortune is normalized to $S_0 = 1$, then the compounded capital S_n after n periods is given by

$$S_n = \prod_{0 \le t < n} (b_t, X_t). \tag{35}$$

The objective is to select portfolios b_t so as to maximize the capital growth rate $\liminf_n n^{-1} \log S_n$. Portfolio b_t must be selected on the basis of an information field $\mathscr{F}_t$ that embodies what is known at the beginning of period t. In other words, b_t must be $\mathscr{F}_t$-measurable ($b_t \in \mathscr{F}_t$, for short).

Let P_t denote a regular conditional probability distribution of X_t given $\mathscr{F}_t$, and let $b^*(\cdot)$ be a measurable selector of log-optimum portfolios. Then $b_t^* =$

$b^*(P_t)$ is an $\mathscr{F}_t$-measurable portfolio attaining the maximum conditional expected log return

$$(36)\qquad w_t^* = w^*(P_t) = \mathbf{E}\{\log(b_t^*, X_t)|\mathscr{F}_t\} = \sup_{b\in\mathscr{F}_t} \mathbf{E}\{\log(b, X_t)|\mathscr{F}_t\}.$$

The expectation of the log return $\log(b_t^*, X_t)$ and of its conditional expectation w_t^* are both equal to the maximum expected log return for period t given $\mathscr{F}_t$,

$$(37)\qquad W_t^* = \mathbf{E}\{w_t^*\} = \mathbf{E}\{\log(b_t^*, X_t)\} = \sup_{b\in\mathscr{F}_t} \mathbf{E}\{\log(b, X_t)\}.$$

We argue that $\{b_t^*\}_{0\le t<\infty}$ is optimum in the long run.

Theorem 5 (Asymptotic optimality principle). *Suppose the random outcomes $\{X_t\}_{0\le t<\infty}$ are defined on a perfect probability space $(\Omega, \mathscr{F}, \mathbf{P})$, and $\{\mathscr{F}_t\}_{0\le t<\infty}$ is an increasing sequence of sub-σ-fields of $\mathscr{F}$ such that $\sigma(X_0,\dots,X_{t-1})\subseteq\mathscr{F}_t$ for all $0\le t<\infty$. Let the compounded capital after n periods of investment according to the log-optimum strategy $\{b_t^*\}_{0\le t<\infty}$ and some competing nonanticipating strategy $\{b_t\}_{0\le t<\infty}$ be denoted by*

$$(38)\qquad S_n^* = \prod_{0\le t<n}(b_t^*, X_t) \quad \text{and} \quad S_n = \prod_{0\le t<n}(b_t, X_t).$$

Then $\{S_n/S_n^*, \mathscr{F}_n\}_{0\le n<\infty}$ is a nonnegative supermartingale converging almost surely to be a random variable Y with $\mathbf{E}\{Y\}\le 1$. Furthermore, $\mathbf{E}\{S_n/S_n^*\}\le 1$ for all n, and

$$(39)\qquad \limsup_n n^{-1}\log\left(\frac{S_n}{S_n^*}\right)\le 0 \quad a.s.$$

Proof. The log-optimum investor and his competitor start with equal fortunes, so that $S_0/S_0^*=1$. The ratio $S_n/S_n^* = \prod_{0\le t<n}(b_t, X_t)/(b_t^*, X_t)$ is $\mathscr{F}_n$-measurable, and the conditional log-optimality of b_n^* given $\mathscr{F}_n$ is equivalent to the Kuhn–Tucker condition

$$\mathbf{E}\left\{\frac{(b_n, X_n)}{(b_n^*, X_n)}\middle|\mathscr{F}_n\right\}\le 1.$$

It follows that

$$\mathbf{E}\left\{\frac{S_{n+1}}{S_{n+1}^*}\middle|\mathscr{F}_n\right\} = \mathbf{E}\left\{\frac{S_n}{S_n^*}\frac{(b_n, X_n)}{(b_n^*, X_n)}\middle|\mathscr{F}_n\right\} = \frac{S_n}{S_n^*}\mathbf{E}\left\{\frac{(b_n, X_n)}{(b_n^*, X_n)}\middle|\mathscr{F}_n\right\}\le\frac{S_n}{S_n^*}.$$

So $\{S_n/S_n^*, \mathscr{F}_n\}_{0\le n<\infty}$ is a nonnegative supermartingale. Any nonnegative supermartingale converges almost surely, and the expectations decrease monotonically to a limit no smaller than the expectation of the limit, by Fatou's lemma. Thus S_n/S_n^* converges almost surely to a nonnegative random variable Y and

$$1 = \mathbf{E}\{S_0/S_0^*\}\ge\mathbf{E}\{S_n/S_n^*\}\ge\lim_n\downarrow\mathbf{E}\{S_n/S_n^*\}\ge\mathbf{E}\{Y\}.$$

Since $\mathbf{E}\{S_n/S_n^*\} \le 1$, it follows from the Markov inequality that, for $r_n > 0$,

$$\mathbf{P}\{S_n/S_n^* \ge r_n\} \le r_n^{-1}\mathbf{E}\{S_n/S_n^*\} \le r_n^{-1}.$$

If r_n increases sufficiently fast so that $\sum_n r_n^{-1} < \infty$, then

$$\sum_n \mathbf{P}\{S_n/S_n^* \ge r_n\} \le \sum_n r_n^{-1} < \infty,$$

and hence $S_n/S_n^* < r_n$ eventually for large n, by the Borel–Cantelli lemma. In particular, choosing $r_n = \exp(n\varepsilon)$ with $\varepsilon > 0$ proves that

$$\mathbf{P}\left\{n^{-1}\log(S_n/S_n^*) \ge \varepsilon \text{ infinitely often}\right\} = 0.$$

Since $\varepsilon > 0$ was arbitrary we may conclude that $\limsup_n n^{-1}\log(S_n/S_n^*) \le 0$ a.s. [This fact can be proved also by observing that S_n/S_n^* converges to a random variable Y with $\mathbf{E}\{Y\} \le 1$ and hence $0 \le Y < \infty$ a.s. Indeed, $S_n/S_n^* \le (1 + Y)$ for large n and hence $\limsup_n n^{-1}\log(S_n/S_n^*) \le \lim_n n^{-1}\log(1 + Y) = 0$ a.s.] $\square$

Theorem 5 asserts that any alternative is dominated in the long run by the log-optimum strategy. Indeed, $\mathbf{E}\{S_n/S_n^*\} \le 1$ for all n, and the Borel–Cantelli lemma implies that $S_n/S_n^* < r_n$ eventually for any sequence $\{r_n\}$ such that $\sum_n r_n^{-1} < \infty$ (e.g., $r_n = n^{1+\varepsilon}$ or $r_n = e^{n\varepsilon}$). The maximal inequality for nonnegative supermartingales [cf. Neveu (1972), Proposition II-2-7, page 23] asserts that

$$\mathbf{P}\left\{\sup_n S_n/S_n^* \ge \lambda\right\} \le 1/\lambda. \tag{40}$$

Thus with probability at least $1 - 1/\lambda$, a competing investor will never outperform S_n^* by a factor greater than λ. The random variable $\sup_n S_n/S_n^*$ is finite almost surely, although its expectation is generally infinite. A game-theoretic sense in which S_n^* dominates S_n for games with payoff $\mathbf{E}\{\phi(S_n^{(1)}/S_n^{(2)})\}$ with ϕ increasing is given in Bell and Cover (1980, 1986).

The conclusions of Theorem 5 hold if $\{\mathscr{F}_t\}$ is an increasing sequence of information fields with $\sigma(U_0,\dots,U_{t-1}) \subseteq \mathscr{F}_t$ for all t. Indeed, $S_n/S_n^* = \prod_{0\le t<n}(b_t, U_t)/(b_t^*, U_t)$ is completely determined by the history of the scaled outcomes $U_t = u(X_t)$.

6. The asymptotic equiparitition property. Breiman (1960, 1961) considered a market with outcomes $\{X_t\}$ that are independent and identically distributed according to an atomic measure, and he argued that repeated choice of the log-optimum portfolio b^* is optimum according to various criteria. In particular, the capital $S_n^* = \prod_{0\le t<n}(b^*, X_t)$ will grow exponentially fast almost surely with limiting rate equal to the maximum expected log return $w^* = \sup_b E\{\log(b, X)\}$, by the strong law of large numbers,

$$n^{-1}\log S_n^* = n^{-1}\sum_{0\le t<n}\log(b^*, X_t) \to w^* = E\{\log(b^*, X)\} \quad \text{a.s.} \tag{41}$$

We prove an asymptotic equipartition property for log-optimum investment in a market that is stationary ergodic. The successive outcomes $X_t(\omega) = X(T^t\omega)$ are defined in terms of a random vector $X(\omega) \in \mathscr{R}_+^m$ and an invertible measurepreserving and metrically transitive transformation T defined on a perfect probability space $(\Omega, \mathscr{F}, \mathbf{P})$. Since T is invertible, the returns can be embedded in a two-sided sequence $\{X_t\}_{-\infty<t<\infty}$.

Let b_t^* be a log-optimum portfolio for period t based on the t-past $\mathscr{F}_t = \sigma(X_0, \dots, X_{t-1})$, and let $\bar{b}_t^*$ be log-optimum for period 0 based on the shifted information field $\bar{\mathscr{F}}_t = T^t\mathscr{F}_t = \sigma(X_{-1}, \dots, X_{-t})$. Portfolios b_t^* and $\bar{b}_t^*$ attain the maximum conditional expected log returns $w_t^* = \sup_{b \in \mathscr{F}_t} \mathbf{E}\{\log(b, X_t)|\mathscr{F}_t\}$ and $\bar{w}_t^* = \sup_{b \in \bar{\mathscr{F}}_t} \mathbf{E}\{\log(b, X_0)|\bar{\mathscr{F}}_t\}$. We denote by $W_t^* = \mathbf{E}\{\log(b_t^*, X_t)\}$ and $\bar{W}_t^* = \mathbf{E}\{\log(\bar{b}_t^*, X_0)\}$ the maximum expected log returns. Then $\bar{W}_t^* = W^*(X_0|X_{-1}, \dots, X_{-t})$ equals $W_t^* = W^*(X_t|X_{t-1}, \dots, X_0)$ by stationarity. If $\bar{b}_\infty^*$ is a log-optimum portfolio for period 0 based on the limiting σ-field $\bar{\mathscr{F}}_\infty = \sigma(X_{-1}, X_{-2}, \dots)$, then $W_t^* = \bar{W}_t^*$ increases monotonically to $\bar{W}_\infty^* = \mathbf{E}\{\log(\bar{b}_\infty^*, X_0)\}$. This limiting expectation is equal to the maximum expected log return given the infinite past, and is denoted by $\bar{W}_\infty^* = W^*(X_0|X_{-1}, X_{-2}, \dots)$. It may be noted that $W^*(X_0|X_{-1}, \dots, X_{-k})$ is the maximum expected log return given the infinite past under the stationary kth-order Markov process having the same $(k+1)$st-order marginal distribution as $\{X_t\}$.

Let $S_n^* = \Pi_{0 \le t < n}(b_t^*, X_t)$ denote the capital growth over n periods of log-optimum investment. The AEP asserts that the time-averaged growth rate $n^{-1}\log S_n^*$ and its expectation $n^{-1}\mathbf{E}\{\log S_n^*\} = n^{-1}W^*(X_0, \dots, X_{n-1})$ converge to the same limit.

THEOREM 6 (Asymptotic equipartition property). *If the sequence of stock return vectors $\{X_t\}$ is stationary ergodic, then capital will grow exponentially fast under the log-optimum investment strategy, almost surely with constant asymptotic rate equal to the maximum expected log return given the infinite past*

$$n^{-1}\log S_n^* \to \bar{W}_\infty^* = W^*(X_0|X_{-1}, X_{-2}, \dots) \quad a.s., \tag{42}$$

where

$$\begin{aligned} W^*(X_0|X_{-1}, X_{-2}, \dots) &= \lim_t \uparrow W^*(X_0|X_{-1}, \dots, X_{-t}) \\ &= \lim_t \uparrow W^*(X_t|X_{t-1}, \dots, X_0) \\ &= \lim_n \uparrow n^{-1}W^*(X_0, \dots, X_{n-1}). \end{aligned} \tag{43}$$

PROOF. One potential approach to establish the AEP for log-optimum investment is to invoke the extended ergodic theorem that was used by Breiman (1957/1960) to prove the AEP of information theory. This extension of the ergodic theorem asserts that

$$n^{-1}\log S_n^* = n^{-1} \sum_{0 \le t < n} \bar{\mathbf{w}}_t^*(T^t\omega) \to \bar{W}_\infty^* = \mathbf{E}\{\bar{\mathbf{w}}_\infty^*\} \quad \text{a.s. and in } L^1, \tag{44}$$

if $\bar{\mathbf{w}}_t^* = \log(\bar{b}_t^*, X_0)$ converges to $\bar{\mathbf{w}}_\infty^* = \log(\bar{b}_\infty^*, X_0)$ and $\{\bar{\mathbf{w}}_t^*\}_{0 \le t < \infty}$ is L^1-dominated. Theorem 4 asserts that $\bar{\mathbf{w}}_t^* \to \bar{\mathbf{w}}_\infty^*$ a.s., but it seems hard to check the integrability condition $\mathbf{E}\{\sup_t|\bar{w}_t^*|\} < \infty$. We shall instead reduce the AEP to direct applications of the ergodic theorem, using a sandwich argument.

The information field $\mathscr{F}_t = \sigma(X_0, \ldots, X_{t-1})$ is approximated by a more refined σ-field $\mathscr{F}_t^{(\infty)}$ and by less refined σ-fields $\mathscr{F}_t^{(k)}$, defined for $0 \le k < \infty$ as follows:

$$\mathscr{F}_t^{(k)} = T^{-t}\overline{\mathscr{F}}_{t \wedge k} = \begin{cases} \sigma(X_0, \ldots, X_{t-1}), & \text{if } 0 \le t < k, \\ \sigma(X_{t-k}, \ldots, X_{t-1}), & \text{if } k \le t < \infty, \end{cases} \tag{45}$$

$$\mathscr{F}_t^{(\infty)} = T^{-t}\overline{\mathscr{F}}_\infty = \sigma(\ldots, X_{-1}, X_0, \ldots, X_{t-1}). \tag{46}$$

Let $b_t^{(k)}$ and $b_t^{(\infty)}$ denote log-optimum portfolios for period t based on the approximating σ-fields $\mathscr{F}_t^{(k)}$ and $\mathscr{F}_t^{(\infty)}$, and let the corresponding capital growths over n periods be denoted by

$$S_n^{(k)} = \sum_{0 \le t < n} \left(b_t^{(k)}, X_t\right) \quad \text{and} \quad S_n^{(\infty)} = \prod_{0 \le t < n} \left(b_t^{(\infty)}, X_t\right). \tag{47}$$

Thus $S_n^{(k)}$, S_n^* and $S_n^{(\infty)}$ denote the capital growth over n periods of log-optimum investment when the investor is allowed to look back at each stage, respectively, at the k-past (but not beyond period 0), up to time 0 and into the infinitely distant past.

Observe that $b_t^{(k)}(\omega) = b_k^*(T^{t-k}\omega)$ if $t \ge k$. Given the expansion

$$n^{-1}\log S_n^{(k)} = n^{-1}\log S_k^* + n^{-1} \sum_{k \le t < n} \log\left(b_t^{(k)}, X_t\right), \tag{48}$$

it follows from the ergodic theorem that

$$n^{-1}\log S_n^{(k)} \to W_k^* = \mathbf{E}\{\log(b_k^*, X_k)\} \quad \text{a.s.} \tag{49}$$

The sequence $\{\log(b_t^{(\infty)}, X_t)\}$ is stationary ergodic and $b_0^{(\infty)} = \bar{b}_\infty^*$, so that again by the ergodic theorem,

$$n^{-1}\log S_n^{(\infty)} = n^{-1} \sum_{0 \le t < n} \log\left(b_t^{(\infty)}, X_t\right) \to \overline{W}_\infty^* = \mathbf{E}\left\{\log\left(\bar{b}_\infty^*, X_0\right)\right\} \quad \text{a.s.} \tag{50}$$

The log-optimum $\mathscr{F}_t^{(k)}$-measurable portfolio $b_t^{(k)}$ is $\mathscr{F}_t$-measurable since $\mathscr{F}_t^{(k)} \subseteq \mathscr{F}_t$, and the log-optimum $\mathscr{F}_t$-measurable portfolio b_t^* is $\mathscr{F}_t^{(\infty)}$-measurable since $\mathscr{F}_t \subseteq \mathscr{F}_t^{(\infty)}$. It follows from the asymptotic optimality principle that

$$\limsup_n n^{-1}\log\left(\frac{S_n^{(k)}}{S_n^*}\right) \le 0 \quad \text{and} \quad \limsup_n n^{-1}\log\left(\frac{S_n^*}{S_n^{(\infty)}}\right) \le 0 \quad \text{a.s.} \tag{51}$$

Thus we obtain the chain of asymptotic inequalities

$$\begin{aligned} W_k^* = \lim_n n^{-1}\log S_n^{(k)} &\le \liminf_n n^{-1}\log S_n^* \\ &\le \limsup_n n^{-1}\log S_n^* \le \lim_n n^{-1}\log S_n^{(\infty)} = \overline{W}_\infty^* \quad \text{a.s.} \end{aligned} \tag{52}$$

The AEP follows since $W_k^* = \overline{W}_k^* \nearrow \overline{W}_\infty^*$ with no gap as $k \to \infty$. □

The sandwich proof of the AEP remains valid if the log-optimum portfolios b_t^* are based on information fields $\mathscr{F}_t$ other than the history of past outcomes $\sigma(X_0, \ldots, X_{t-1})$. However, $\{\mathscr{F}_t\}_{0 \le t < \infty}$ must be monotonically increasing and the history of the scaled return vectors $\sigma(U_0, \dots, U_{t-1})$ must be contained in $\mathscr{F}_t$, so that the asymptotic optimality principle can be invoked. Monotonicity of $\{\mathscr{F}_t\}$ means that information available about the past should never be erased from memory. In addition, one must assume that the shifted fields $\bar{\mathscr{F}}_t = T^t\mathscr{F}_t$ are monotonically increasing to a limiting σ-field $\bar{\mathscr{F}}_\infty$, so that $\overline{W}_t^* \nearrow \overline{W}_\infty^*$ by Theorem 4. Monotonicity of $\{\bar{\mathscr{F}}_t\}$ means that later investors have an advantage in information when compared on common grounds, after shifting back to the reference period 0, where all face the same decision problem of selecting b_0.

Suppose in particular that side information $Y_t(\omega) = Y(T^t\omega)$ is revealed together with the return vector X_t at the end of period t. Then $\mathscr{F}_t = \sigma(X_0, Y_0, \dots, X_{t-1}, Y_{t-1})$ and $\bar{\mathscr{F}}_t = \sigma(X_{-t}, Y_{-t}, \dots, X_{-1}, Y_{-1})$ are monotonically increasing, and $n^{-1}\log S_n^* \to \overline{W}_\infty^*$ almost surely where $\overline{W}_\infty^* = W^*(X_0|X_{-1}, Y_{-1}, X_{-2}, Y_{-2}, \dots)$ is the maximum expected log return given the infinite past. The proof is identical to that of Theorem 6, except that $b_t^* = b^*(Q_t)$ and $\bar{b}_t^* = b^*(\overline{Q}_t)$ are now defined by applying a measurable selector of log-optimum portfolios $b^*(\cdot)$ to regular conditional probability distributions Q_t and $\overline{Q}_t$ of $U_t = u(X_t)$ given $\mathscr{F}_t$ and of $U_0 = u(X_0)$ given $\bar{\mathscr{F}}_t$.

The true log return $\log(b_t^*, X_t)$ will generally differ from the conditional expected log return $w_t^* = \mathbf{E}\{\log(b_t^*, X_t)|\mathscr{F}_t\}$. If conditional expected log returns were always exactly realized then the capital growth over n periods would be not S_n^* but rather

$$\tilde{S}_n^* = \prod_{0 \le t < n} \exp[\mathbf{E}\{\log(b_t^*, X_t)|\mathscr{F}_t\}]. \tag{53}$$

If $\tilde{S}_n = \prod_{0 \le t < n} \exp[\mathbf{E}\{\log(b_t, X_t)|\mathscr{F}_t\}]$ denotes the corresponding quantity under the competing strategy $\{b_t\}$, then $\tilde{S}_n \le \tilde{S}_n^*$ for all n, and hence

$$\limsup_n n^{-1}\log\left(\tilde{S}_n/\tilde{S}_n^*\right) \le 0 \quad \text{a.s.} \tag{54}$$

This may be called an asymptotic optimality principle for the hypothetical growth rate $\tilde{S}_n^*$. If the market is stationary ergodic, then an asymptotic equipartition property for $\tilde{S}_n^*$ can be proved as well, under certain integrability conditions. Let $L \log L$ designate the class of random variables $g(\omega)$ such that $\mathbf{E}\{|g|\log|g|\} < \infty$.

Theorem 7. *If the market is stationary ergodic and* $\mathbf{E}\{\log(\beta, X_0)|\bar{\mathscr{F}}_\infty\}$ *belongs to* $L \log L$, *then*

$$n^{-1}\log \tilde{S}_n^* \to \tilde{W}_\infty^* \quad a.s.\ and\ in\ L^1. \tag{55}$$

Proof. Breiman's (1957/1960) extension of the ergodic theorem asserts that $n^{-1}\sum_{0 \le t < n} g_t(T^t\omega) \to \mathbf{E}\{g\}$ *a.s.* and in L^1 if $g_t \to g$ a.s. and $\mathbf{E}\{\sup_t|g_t|\} < \infty$. In particular, if $\{g_t, \bar{\mathscr{F}}_t\}_{0 \le t < \infty}$ is a martingale or a nonnegative submartingale with limit g in $L \log L$, then the integrability condition $\mathbf{E}\{\sup_t|g_t|\} < \infty$ is satisfied. Indeed, Wiener's dominated ergodic theorem [cf. Chung (1974), example 7, page 355] asserts that

$$\mathbf{E}\left\{\sup_t|g_t|\right\} \le \frac{e}{e-1}\left[1 + \sup_t \mathbf{E}\{|g_t|\log^+|g_t|\}\right] \le \frac{e}{e-1}[1 + \mathbf{E}\{|g|\log^+|g|\}].$$

Consider the decomposition $\overline{w}_t^* = \bar{r}_t + w^*(\overline{Q}_t)$, where $\bar{r}_t = \mathbf{E}\{\log(\beta, X_0)|\overline{\mathscr{F}}_t\}$. Since $\{\bar{r}_t, \overline{\mathscr{F}}_t\}_{0 \le t < \infty}$ is a martingale with limit $\bar{r}_\infty = \mathbf{E}\{\log(\beta, X_0)|\overline{\mathscr{F}}_\infty\}$ in $L \log L$, Breiman's extended ergodic theorem implies that

$$n^{-1} \sum_{0 \le t < n} \mathbf{E}\{\log(\beta, X_t)|\mathscr{F}_t\} \to \mathbf{E}\{\log(\beta, X_0)\} \quad \text{a.s. and in } L^1.$$

Since $\{w^*(\overline{Q}_t)\}$ is bounded and $w^*(\overline{Q}_t) \to w^*(\overline{Q}_\infty)$ a.s., we also have

$$n^{-1} \sum_{0 \le t < n} \mathbf{E}\{\log(b_t^*, U_t)|\mathscr{F}_t\} \to \mathbf{E}\{w^*(\overline{Q}_\infty)\} \quad \text{a.s. and in } L^1.$$

By summation we may conclude that

$$n^{-1} \log \tilde{S}_n^* = n^{-1} \sum_{0 \le t < n} w_t^* \to \mathbf{E}\{\overline{w}_\infty^*\} = \overline{W}_\infty^* \quad \text{a.s. and in } L^1. \qquad \square$$

7. Stationary markets. We shall prove the AEP for markets that are stationary but not necessarily ergodic. A stationary market is a mixture of stationary ergodic modes [cf. Maitra (1977)], but no finite number of observations may suffice to exactly identify the (random) ergodic mode of $\{X_t\}$. However, log-optimum portfolios based on the t-past are better and better suited to the ergodic mode as t increases, and the log-optimum portfolio given the infinite past will be perfectly tailored because the ergodic mode is uniquely determined by the infinite past. It is therefore not surprising that S_n^* will grow with the same asymptotic rate as if the ergodic mode were known to begin with.

The AEP may hold even if the market is stationary in an asymptotic sense only. A dynamical system $(\Omega, \mathscr{F}, \mathbf{P}, T)$ asymptotically mean stationary (a.m.s.) if the Cesàro averages $n^{-1}\Sigma_{0 \le t < n}\mathbf{P}(T^{-t}F)$ converge for any event $F \in \mathscr{F}$. Setting the limit equal to $\overline{\mathbf{P}}(F)$ then defines a stationary (T-invariant) probability distribution $\overline{\mathbf{P}}$ on $(\Omega, \mathscr{F})$, and $\overline{\mathbf{P}}$ is perfect whenever $\mathbf{P}$ is. $\overline{\mathbf{P}}$ the stationary mean of $\mathbf{P}$, and expectations with respect to $\overline{\mathbf{P}}$ are denoted by $\overline{\mathbf{E}}\{\cdot\}$. The measures $\mathbf{P}$ and $\overline{\mathbf{P}}$ have the same restriction to the invariant σ-field $\mathscr{I} = \{F \in \mathscr{F}: T^{-1}F = F\}$, so that $\mathbf{E}\{\cdot|\mathscr{I}\} = \overline{\mathbf{E}}\{\cdot|\mathscr{I}\}$. See Gray and Kieffer (1980) for further discussion of asymptotically mean stationary measures, and Section 34.2 in Loève (1978) for a proof that the following strong law of large numbers holds for nonnegative measurable $g(\omega)$:

$$n^{-1} \sum_{0 \le t < n} g(T^t\omega) \to \mathbf{E}\{g|\mathscr{I}\} = \overline{\mathbf{E}}\{g|\mathscr{I}\} \quad \text{a.s. } (\mathbf{P}) \text{ and a.s. } (\overline{\mathbf{P}}). \tag{56}$$

A market asymptotically mean stationary if the underlying dynamical system $(\Omega, \mathscr{F}, \mathbf{P}, T)$ is a.m.s. As before we assume that T is invertible, $\mathbf{P}$ is perfect, and $X_t(\omega) = X(T^t\omega)$ for some random vector $X(\omega) \in \mathscr{R}_+^m$. The AEP holds for an asymptotically mean stationary market, unless the investor goes broke after a few rounds and remains trapped in a state that is infinitely worse than any other. The investor should not be completely ruined by the time he reaches the asymptotic regime, so that he can recover from transient losses.

THEOREM 8. *Suppose the market is stationary, and $\bar{b}_\infty^*$ is a log-optimum portfolio for period* 0 *given the infinite past $\overline{\mathscr{F}}_\infty$. Then*

$$n^{-1} \log S_n^* \to \mathbf{E}\{\log(\bar{b}_\infty^*, X_0)|\mathscr{I}\} \quad a.s. \tag{57}$$

The same conclusion holds if the market is asymptotically mean stationary and $\bar{b}_\infty^*$ is log-optimum under the stationary mean, at least if $n^{-1}\log S_{k_n}^* \to 0$ a.s. for some sequence $\{k_n\}$ such that $k_n \nearrow \infty$ and $k_n/n \to 0$.

PROOF. We consider the asymptotically mean stationary case. Recall that $S_n^* = \prod_{0\le t<n}(b_t^*, X_t)$, where b_t^* is a log-optimum portfolio for period t based on the t-past $\mathscr{F}_t = \sigma(X_0, \dots, X_{t-1})$. Portfolio b_t^* is log-optimum with respect to the true distribution $\mathbf{P}$. Let $\bar{b}_k^*$ and $\bar{b}_\infty^*$ designate portfolios for period 0 that are log-optimum with respect to the stationary mean $\overline{\mathbf{P}}$, based on the shifted information field $\overline{\mathscr{F}}_k = \sigma(X_{-k}, \dots, X_{-1})$ and the limiting σ-field $\overline{\mathscr{F}}_\infty = \sigma(\dots, X_{-2}, X_{-1})$. If an investor selects log-optimum portfolios b_t^* during the first k periods $0 \le t < k$, and in later periods $t \ge k$ switches to suboptimum portfolios $\bar{b}_k^*(T^t\omega)$ (i.e., portfolios based on the k-past that are log-optimum with respect to the stationary mean $\overline{\mathbf{P}}$), then capital growth over n periods will be given by

$$S_n^{(k)} = \begin{cases} S_n^*, & \text{if } 0 \le n < k, \\ S_k^* \prod_{k\le t<n} \left(\bar{b}_k^*(T^t\omega), X_t\right), & \text{if } k \le n < \infty. \end{cases}$$

If the investor always selects the portfolio $\bar{b}_\infty^*(T^t\omega)$ that is log-optimum based on the infinite past with respect to the stationary mean $\overline{\mathbf{P}}$, then capital growth is given by

$$S_n^{(\infty)} = \prod_{0\le t<n} \left(\bar{b}_\infty^*(T^t\omega), X_t\right).$$

It is clear that $\mathbf{E}\{S_n^{(k)}/S_n^*\} \le 1$ and $\overline{\mathbf{E}}\{S_n^*/S_n^{(\infty)}\} \le 1$, so that by Markov's inequality and the Borel–Cantelli lemma (cf. the proof of Theorem 5),

$$\limsup_n n^{-1}\log\left(\frac{S_n^{(k)}}{S_n^*}\right) \le 0 \quad \text{a.s. } (\mathbf{P}) \quad \text{and} \quad \limsup_n n^{-1}\log\left(\frac{S_n^*}{S_n^{(\infty)}}\right) \le 0 \quad \text{a.s. } (\overline{\mathbf{P}}).$$

The ergodic theorem for a.m.s. measures implies that

$$n^{-1}\log S_n^{(k)} \to \mathbf{E}\left\{\log\left(\bar{b}_k^*(T^k\omega), X_k\right)|\mathscr{I}\right\} = \overline{\mathbf{E}}\left\{\log\left(\bar{b}_k^*, X_0\right)|\mathscr{I}\right\} \quad \text{a.s. } (\mathbf{P})$$

and

$$n^{-1}\log S_n^{(\infty)} \to \overline{\mathbf{E}}\left\{\log\left(\bar{b}_\infty^*, X_0\right)|\mathscr{I}\right\} \quad \text{a.s. } (\overline{\mathbf{P}}).$$

Combining the previous results yields

$$\mathbf{E}\left\{\log\left(\bar{b}_k^*(T^k\omega), X_k\right)|\mathscr{I}\right\} = \overline{\mathbf{E}}\left\{\log\left(\bar{b}_k^*, X_0\right)|\mathscr{I}\right\} \le \liminf_n n^{-1}\log S_n^* \quad \text{a.s. } (\mathbf{P})$$

and

$$\limsup_n n^{-1}\log S_n^* \le \overline{\mathbf{E}}\left\{\log\left(\bar{b}_\infty^*, X_0\right)|\mathscr{I}\right\} \quad \text{a.s. } (\overline{\mathbf{P}}).$$

The last inequality also holds a.s. $(\mathbf{P})$ since both sides are invariant (the left-hand side by assumption). We obtain the chain of asymptotic inequalities

$$\overline{\mathbf{E}}\{\log(\bar{b}_k^*, X_0)|\mathscr{I}\} \le \liminf_n n^{-1}\log S_n^*$$
$$\le \limsup_n n^{-1}\log S_n^* \le \overline{\mathbf{E}}\{\log(\bar{b}_\infty^*, X_0)|\mathscr{I}\} \quad \text{a.s. } (\mathbf{P}).$$

We claim that $\overline{\mathbf{E}}\{\log(\bar{b}_k^*, X_0)|\mathscr{I}\}$ is increasing in k. Indeed, if $k \le l$, then the event where $\overline{\mathbf{E}}\{\log(\bar{b}_k^*, X_0)|\mathscr{I}\}$ exceeds $\overline{\mathbf{E}}\{\log(\bar{b}_l^*, X_0)|\mathscr{I}\}$ must have zero probability, since conditioning on this invariant event and taking expectations would otherwise contradict the inequality $\overline{W}_k^* \le \overline{W}_l^*$. The expectations $\overline{W}_k^* = \mathbf{E}\{\log(\bar{b}_k^*, X_0)\}$ increase to $\overline{W}_\infty^* = \mathbf{E}\{\log(\bar{b}_\infty^*, X_0)\}$ as $k \to \infty$, so that by the monotone convergence theorem,

$$\overline{\mathbf{E}}\{\log(\bar{b}_k^*, X_0)|\mathscr{I}\} \nearrow \overline{\mathbf{E}}\{\log(\bar{b}_\infty^*, X_0)|\mathscr{I}\} \quad \text{a.s. } (\overline{\mathbf{P}}).$$

Convergence also holds a.s. (**P**) since we are dealing with invariant random variables, and Theorem 8 follows since $\mathbf{E}\{\log(\bar{b}_\infty^*, X_0)|\mathscr{I}\} = \overline{\mathbf{E}}\{\log(\bar{b}_\infty^*, X_0)|\mathscr{I}\}$. □

8. Gambling as investment. We consider a market in which exactly one stock will yield a nonzero return, the jth stock with probability q^j. The random outcome X is then oriented along one of the coordinate axes of $\mathscr{R}^m$, and the scaled return U is an extreme point of the simplex $\mathscr{U} = \{u \in \mathscr{R}_+^m : (\beta, u) = 1\}$. As observed by Kelly (1956), investing in such a market is like gambling on the outcome of a horse race in which horse j has win probability q^j. Since one unit bet on horse j yields $U^j = 1/\beta^j$ if horse j wins, we have

$$w(b, Q) = E_Q\{\log(b, U)\} = \sum_{1\le j\le m} q^j\log(b^j/\beta^j) = D(q\|\beta) - D(q\|b).$$

The information divergence $D(q\|b) = \sum_{1\le j\le m} q^j\log(q^j/b^j)$ is nonnegative, and equal to zero iff $b = q$. It follows that the bet vector $b = q = (q^j)_{1\le j\le m}$ is the unique log-optimum portfolio. Thus the gambler should ignore the odds $1/\beta^j$ and place an amount on each horse j proportional to its win probability q^j. The maximum expected log of the scaled return $w^*(Q)$ is precisely the Kullback–Leibler divergence between the probability vector q and the reference portfolio β that defines the odds, i.e.,

$$w^*(Q) = D(q\|\beta) = \sum_{i\le j\le m} q^j\log(q^j/\beta^j). \tag{58}$$

Gambling on a set of m stocks out of which exactly one will yield a nonzero return is a most risky type of investment game. Least risky is a market whose return vector has a fixed direction, so that the stock(s) with highest return can be predicted with certainty. In general, we say that a distribution Q on $\mathscr{U}$ is less risky than another distribution Q', and write $Q \le Q'$, if there exists a dilation $\Gamma(du|\mu)$ of $\mathscr{U}$ such that $Q' = \Gamma Q$, i.e., $Q'(\cdot) = \int_{\mathscr{U}}\Gamma(\cdot|u)Q(du)$. [A dilation of $\mathscr{U}$ is a transition probability $\Gamma(du|\mu)$ from $\mathscr{U}$ to $\mathscr{U}$ such that μ is equal to the barycenter of $\Gamma(\cdot|\mu)$ for all $\mu \in \mathscr{U}$.] See Alfsen (1971) for more discussion of this so-called dilation or Choquet order on the space $\mathscr{Q}$ of probability measures on $\mathscr{U}$.

If $Q \le Q'$, then Q is less risky and more attractive than Q', in terms of expected log return. Indeed, $Q \le Q'$ iff $\int_{\mathscr{U}} \varphi(u) Q(du) \le \int_{\mathscr{U}} \varphi(u) Q'(du)$ for all lower semicontinuous convex $\varphi: \mathscr{U} \to (-\infty, \infty]$. Choosing $\varphi(u) = -\log(b, u)$ proves that $w(b, Q) = E_Q\{\log(b, U)\}$ is increasing in Choquet order, and taking suprema proves

THEOREM 9. *The maximum expected log return* $w^*(Q) = \sup_b E\{\log(b, U)\}$ *is monotonically decreasing in Choquet order on* $\mathscr{Q}$, *i.e.*,

$$\text{if } Q \le Q' \text{ in } \mathscr{Q}, \text{ then } w^*(Q) \ge w^*(Q'). \tag{59}$$

$\mathscr{U}$ is a Choquet simplex, so every distribution Q on $\mathscr{U}$ admits a barycenter $\mu(Q) \in \mathscr{U}$ and for every point $\mu \in \mathscr{U}$ there exists a unique probability measure π_μ on the set of extreme points of $\mathscr{U}$ that admits μ as barycenter. Two measures that are comparable in Choquet order have the same barycenter. The point mass δ_μ that is concentrated at μ is minimal and the measure π_μ on the extreme points of $\mathscr{U}$ is maximal with respect to Choquet order on $\mathscr{Q}$, among all distributions that admit the point $\mu \in \mathscr{U}$ as barycenter. Notice that $\pi_{\mu(Q)} = \Pi Q$, where $\Pi(du|\mu) = \pi_\mu(du)$ is the maximal dilation that sweeps all mass to the extreme points of $\mathscr{U}$; the minimal dilation is the identity kernel $\Delta(du|\mu) = \delta_\mu(du)$ that leaves all mass put.

Among all distributions Q on $\mathscr{U}$ with a given barycenter $\mu(Q) = \mu$, the most concentrated measure δ_μ is best and the most dilated measure π_μ is worst in terms of expected log return. Indeed, let $\mu = \mu(Q)$ and let q denote the probability vector proportional to μ [with components $q^j = \mu^j / (\sum_j \mu^j)$]. Since $\delta_\mu \le Q \le \pi_\mu$, Theorem 9 implies that

$$\max_j \log \mu^j = w^*(\delta_\mu) \ge w^*(Q), \tag{60}$$

$$w^*(Q) \ge w^*(\pi_\mu) = D(q\|\beta) = \sum_j q^j \log q^j. \tag{61}$$

The most natural choice for β is the uniform portfolio $(1/m)_{1 \le j \le m}$, which allocates an equal amount to each of the m stocks and whose yield (β, X) is the arithmetical average return $m^{-1}(X^1 + \cdots + X^m)$. Then $D(q\|\beta) = \log m - \mathscr{H}(q)$, where $\mathscr{H}(q) = -\sum_j q^j \log q^j$ is the Shannon entropy of the probability vector q. In general, one may interpret $h^*(Q) = \log m - w^*(Q)$ as the minimum loss of expected log return relative to the ideal reference level $r(P) + \log m$. When the chain of inequalities $0 \le w^*(\delta_\mu) \le w^*(Q) \le w^*(\pi_\mu) \le \log m$ is rewritten in terms of $h^*(Q)$, one obtains

$$0 \le \min_j (-\log q^j) = h^*(\delta_\mu) \le h^*(Q) = \log m - w^*(Q), \tag{62}$$

$$h^*(Q) \le h^*(\pi_\mu) = \mathscr{H}(q) = \sum_j q^j (-\log q^j). \tag{63}$$

If one starts with a point mass δ_μ located at $\mu \in \mathscr{U}$ and repeatedly dilates mass, then Q traces out a linearly $\le$-ordered chain of distributions all having barycenter $\mu(Q) = \mu$ in $\mathscr{U}$. Ultimately, one ends up with a measure π_μ, when all mass is swept to the corners (extreme points) of the simplex $\mathscr{U}$. Initially (when $Q = \delta_\mu$), one can place all bets on the stock(s) j for which the minimum information loss $(-\log q^j)$ is minimum, but in the end (when $Q = \pi_\mu$) one has to place proportional bets and concede an average loss equal to the Shannon entropy $\mathscr{H}(q) = \sum_{1 \le j \le m} q^j(-\log q^j)$.

Gambling on the next outcome of a horse race is a special type of investment game. Proportional betting is log-optimum, and the asymptotic optimality principle and asymptotic equipartition property can be formulated in a way that does not seem to involve a maximization, since the log-optimum strategy is explicitly known. The same is true for proportional betting on the next outcome of a random process with values in a Polish space. Indeed, let $p(x_0, \dots, x_{n-1})$ denote the marginal density with respect to some dominating measure of the first n outcomes of a random process $\{X_t\}$, and let $q(x_0, \dots, x_{n-1})$ denote the density under some alternative distribution. The likelihood ratio $q(X_0, \dots, X_{n-1})/p(X_0, \dots, X_{n-1})$ is then a nonnegative supermartingale converging almost surely to a random variable Y with $\mathbf{E}\{Y\} \le 1$, and

$$\limsup_n n^{-1} \log\left(\frac{q(X_0, \dots, X_{n-1})}{p(X_0, \dots, X_{n-1})} \right) \le 0 \quad \text{a.s.} \tag{64}$$

If, moreover, $\{X_t\}$ is stationary ergodic and densities are taken with respect to a Markovian reference measure, then $p(X_0, \dots, X_{n-1})$ will grow exponentially fast with constant limiting rate almost surely equal to the relative entropy rate of the true distribution with respect to the reference measure, i.e.,

$$n^{-1} \log p(X_0, \dots, X_{n-1}) \to \sup_n \mathbf{E}\{\log p(X_0, \dots, X_{n-1})\} \quad \text{a.s.} \tag{65}$$

See Barron (1985) and Orey (1985) for a proof of this generalized Shannon–McMillan–Breiman theorem using Breiman's extension of the ergodic theorem and Algoet and Cover (1988) for a sandwich proof.

Acknowledgments. We wish to acknowledge supporting discussions with John Gill and David Larson.

REFERENCES

ALFSEN, E. M. (1971). *Compact Convex Sets and Boundary Integrals*. Springer, New York.

ALGOET, P. H. and COVER, T. M. (1988). A sandwich proof of the Shannon–McMillan–Breiman theorem. *Ann. Probab.* **16** 899–909.

BARRON, A. R. (1985). The strong ergodic theorem for densities: Generalized Shannon–McMillan–Breiman theorem. *Ann. Probab.* **13** 1292–1303.

BELL, R. and COVER, T. M. (1980). Competitive optimality of logarithmic investment. *Math. Oper. Res.* **5** 161–166.

BELL, R. and COVER, T. M. (1986). Game-theoretic optimal portfolios. Preprint.

BERTSEKAS, D. P. and SHREVE, S. E. (1978). *Stochastic Optimal Control, the Discrete Time Case.* Academic, New York.

BREIMAN, L. (1957/1960). The individual ergodic theorem of information theory. *Ann. Math. Statist.* **28** 809–811. Correction **31** 809–810.

BREIMAN, L. (1960). Investment policies for expanding businesses optimal in a long run sense. *Naval Res. Logist. Quart.* **7** 647–651.

BREIMAN, L. (1961). Optimal gambling systems for favorable games. *Proc. Fourth Berkeley Symp. Math. Statist. Probab.* **1** 65–78. Univ. California Press.

CHUNG, K. L. (1974). *A Course in Probability Theory*, 2nd ed. Academic, New York.

COVER, T. M. (1984). An algorithm for maximizing expected log investment return. *IEEE Trans. Inform. Theory* **IT-30** 369–373.

FINKELSTEIN, M. and WHITLEY, R. (1981). Optimal strategies for repeated games. *Adv. in Appl. Probab.* **13** 415–428.

GRAY, R. M. and KIEFFER, J. C. (1980). Asymptotically mean stationary measures. *Ann. Probab.* **8** 962–973.

JIŘINA, M. (1954). Conditional probabilities on σ-algebras with countable basis. *Czechoslavak Math. J.* **4** 372–380. (English translation in *Amer. Math. Soc. Transl. Ser. 2* **2** (1962) 79–86.)

KELLY, J. L., JR. (1956). A new interpretation of information rate. *Bell System Tech. J.* **35** 917–926.

KURATOWSKI, K. and RYLL-NARDZEWSKI, C. (1961). A general theorem on selectors. *Bull. Acad. Polon. Sci.* **13** 397–403.

LOÈVE, M. (1978). *Probability Theory* **2**, 4th ed. Springer, New York.

MAITRA, A. (1977). Integral representations of integral measures. *Trans. Amer. Math. Soc.* **229** 209–225.

MARKOWITZ, H. M. (1952). Portfolio selection. *J. Finance* **7** 77–91.

MARKOWITZ, H. M. (1959). *Portfolio Selection*. Wiley, New York.

NEVEU, J. (1972). *Martingales à Temps Discret.* Masson, Paris.

OREY, S. (1985). On the Shannon-Perez-Moy theorem. *Contemp. Math.* **41** 319–327.

RAMACHANDRAN, D. (1979). *Perfect Measures I—Basic Theory* and *II—Special Topics. ISI Lecture Notes* **5**, **7**. Macmillan, New York.

SAMUELSON, P. (1967). General proof that diversification pays. *J. Financial and Quantitative Anal.* **2** 1–13.

SAMUELSON, P. (1970). The fundamental approximation theorem of portfolio analysis in terms of means, variances and higher moments. *Rev. Econom. Stud.* **37** 537–542.

SAMUELSON, P. (1971). The "fallacy" of maximizing the geometric mean in long sequences of investing or gambling. *Proc. Nat. Acad. Sci. U.S.A.* **68** 2493–2496.

SHARPE, W. F. (1985). *Investments*, 3rd ed. McGraw-Hill, New York.

THORP, E. O. (1971). Portfolio choice and the Kelly criterion. In *Stochastic Optimization Models in Finance* (W. T. Ziemba and R. G. Vickson, eds.) 599–619. Academic, New York.

COLLEGE OF ENGINEERING
BOSTON UNIVERSITY
110 CUMMINGTON STREET
BOSTON, MASSACHUSETTS 02215

DEPARTMENTS OF STATISTICS AND
ELECTRICAL ENGINEERING
STANFORD UNIVERSITY
STANFORD, CALIFORNIA 94305