Portfolio Performance Measurement: Theory and Applications

Zhiwu Chen
Ohio State University

Peter J. Knez
University of Wisconsin-Madison
and Northwestern University

Any admissible portfolio performance measure should satisfy four minimal conditions: it assigns zero performance to each reference portfolio and it is linear, continuous, and nontrivial. Such an admissible measure exists if and only if the securities market obeys the law of one price. A positive admissible measure exists if and only if there is no arbitrage. This article characterizes the (infinite) set of admissible performance measures. It is shown that performance evaluation is generally quite arbitrary. A mutual fund data set is also used to demonstrate how the measurement method developed here can be applied.

The literature on mutual fund performance measurement goes back to the beginning of asset pricing

This article supercedes two previous papers respectively titled, "A Pricing Operator Based Approach to Performance Evaluation" and "Mutual Fund Performance: A Nonparametric Empirical Investigation." The authors would like to thank John Cochrane, Wayne Ferson, William Goetzmann, Allstair Hall, Lars Hansen, Robert Korajczyk (the editor), Blake LeBaron, Mike Spitzer, Ken West, an anonymous referee, and seminar participants at the University of Chicago, the University of North Carolina, the University of Minnesota and the University of Wisconsin-Madison for comments. We are grateful to Dean Johnson and Walter Prahl for research assistance. Morning Star Inc. graciously provided much of the mutual fund data used in this study. Thanks are also due to Wayne Ferson who provided several data sets that greatly reduced the data acquisition process. Any remaining errors are ours alone. Address correspondence to Peter J. Knez, J. L. Kellogg School of Management, Northwestern University, Leverone Hall, 2001 Sheridan Road, Evanston, IL 60208.

The Review of Financial Studies Summer 1996 Vol. 9, No. 2, pp. 511–555

theory, if not further. Since the early formal measures of Jensen (1968), Sharpe (1966), and Treynor (1965), numerous new performance measures have been proposed. While we do not intend to conduct a literature review here, there are the APT-based measures of Connor and Korajczyk (1986) and Lehmann and Modest (1987), the period weighting measures of Grinblatt and Titman (1989), and the intertemporal marginal rates of substitution-based measures of Glosten and Jagannathan (1994), to list just a few. From a performance evaluator's perspective, this array of measures offers a rich choice set but at the same time makes the selection of a method difficult (if possible at all). To make matters worse, there is no general theoretical framework that allows the evaluator to examine these and many other proposed measures.[1] What constitutes a "performance measure"? What is the minimal market condition in order for there to exist an admissible performance measure? Are there other performance measures beyond those that have been proposed? How differently will these and other "admissible performance measures" rank a given managed fund? The goal of this article is to address these questions and study the entire class of admissible performance measures. In the analysis, we also propose performance measures that can be identified from market data and are independent of any asset pricing model.

The essence of performance evaluation is to measure the value of the services (if any) provided by the portfolio management industry. It is to investigate whether a fund manager helps enlarge the investment opportunity set faced by the investing public and, if so, to what extent the manager enlarges it. Put differently, if the manager provides a portfolio that is also achievable by the investing public, he offers no service; it is when the managed fund lies outside of the existing investment opportunity set faced by the public that the manager offers a genuine service. With this in mind, we say a function is an "admissible performance measure" if it satisfies the following four minimal conditions. First, it assigns zero performance to every portfolio in some reference set. For instance, if the uninformed investors constitute the investing public, the reference set will then contain all portfolio returns that are achievable by any uninformed investor. More generally, the reference portfolio set can be enlarged to include all dynamic portfolios that are obtainable using public information. Second, the function is linear so that a manager cannot create better or worse performance by simply re-

[1] In the empirical literature, there have been studies to examine how different measures may rank funds differently. See, for instance, Chen, Copeland and Mayers (1987) and Lehmann and Modest (1987).

bundling other funds. This ensures that superior performance is only a result of superior information. Third, the function is continuous, which guarantees that any two funds whose returns are indistinguishable from one another will always be assigned performance values that are arbitrarily close. This imposes some sense of fairness to all fund managers. Finally, the function is nontrivial in the sense that if a fund's excess return over a reference portfolio is proportional to some traded security's payoff, the fund will not be assigned a zero performance.

What is the minimal market condition under which an admissible performance measure exists? It turns out that there is an admissible performance measure if and only if the securities market satisfies the law of one price. This means that for performance evaluation purposes, requiring market conditions (e.g., general equilibrium) stronger than the law of one price may not be necessary. Now, suppose that the law of one price holds on the market from which a performance evaluator selects reference portfolios and that the evaluator is given some arbitrary function $\lambda(\cdot)$. Then, can this arbitrary $\lambda(\cdot)$ be admissible as a performance measure? The answer is yes if and only if it can be represented by some random variable d (satisfying certain conditions) as in $\lambda(x) = E(d\,x)$, for every random variable x in some proper space. This result, stated in Theorem 2, provides the strongest possible characterization of admissible performance measures, since it asserts that every admissible measure must be representable in this inner-product form *and* only functions representable in this form can be admissible as performance measures. This characterization answers the basic question of what can be a performance measure and what cannot. It not only allows us to reexamine existing proposed performance measures but also provides a criterion for the introduction of new performance measures. For instance, we demonstrate that the Jensen measure, the APT-based measures, and the measures in Glosten and Jagannathan (1994) are all representable in this form and are therefore admissible in our sense, but the period weighting measures may not be admissible unless one additional condition on the period weights is also imposed. Nonetheless, these are only a few members in the infinite set of admissible performance measures.

Given the infinity of admissible performance measures, will they evaluate the same fund drastically differently? If a managed return lies in the benchmark return set, the answer is no and all admissible measures will give the fund a zero performance. However, if a fund's return is not achievable by an uninformed investor, then by suitably choosing an admissible performance measure any performance value can be assigned to the fund. Furthermore, for two funds whose returns are not achievable by any uninformed investor, there will always be admissible measures that rank the two funds totally differently: if one measure ranks one fund higher, there exists another measure that will simply reverse the ranking (so long as the two funds are not perfectly correlated). These two results are similar in spirit to, but not the same as, Roll's (1978) conclusions.[2] In the CAPM context, Roll argues that *(i)* by choosing the right inefficient market index any Jensen α value can be assigned to a managed fund and *(ii)* when inefficient market indices are used for the Jensen measure, the ranking of two managed funds can be easily reversed by the choice of two inefficient market indices. Note that in his case the arbitrariness of performance evaluation is caused by the use of an inefficient market index and hence by the use of inadmissible performance measures (in the sense defined here), whereas in our context the arbitrariness is due to the multiplicity of admissible performance measures.

The above negative results about performance measurement obtain in part because of the fact that each measure is only required to satisfy the four minimal conditions stated earlier. To make performance evaluation less dependent on the choice of a particular performance measure, we add one more condition: each measure must be positive, in the sense that if a managed excess return (over any reference return) is positive with probability one, then the fund will be assigned a positive performance. Such a performance measure exists if and only if the securities market offers no-arbitrage (NA) opportunities. For this reason, we refer to each admissible positive performance measure as an NA-based measure. When only NA-based measures are used for performance measurement, performance values assigned to a given fund will lie in some interval rather than the entire real line. It is no longer the case that for any desired value, one can find a performance measure that assigns this value. In particular, if a managed fund dominates some reference portfolio in every state of nature, then each NA-based measure will give it a positive performance ranking.

NA-based measures possess other nice properties as well. For example, if a fund is given a positive ranking by some NA-based measure, there must exist some investor in a certain preference class who values (at the margin) this fund more than any reference portfolio, which means that positive performance is associated with genuine service provided by the manager and appreciated by at least one investor. Conversely, if a fund is valued by some investor more

---

[2] Roll's points are further analyzed by Dybvig and Ross (1985) and Green (1987).

than any reference portfolio, there must exist an NA-based measure that gives the fund a positive performance. This result is encouraging in the sense that if an evaluator uses an NA-based measure and finds a fund with positive performance, the evaluator can immediately conclude that the manager has achieved an outperforming portfolio with superior information, at least from some investor's perspective.

Mayers and Rice (1979) were the first to formally argue that within the CAPM framework it is internally consistent for some information-driven fund to have a nonzero Jensen α value. Then, Dybvig and Ross (1985) further elaborated on this point.[3] Using a functional analytic framework, our analysis further generalizes this argument. It shows that having superior information per se does not automatically lead to superior performance, depending on how the manager uses the information. If the manager takes advantage of the information in the right direction, he will typically produce a portfolio return that lies outside of the linear span of reference portfolios. It then follows that many admissible performance measures will assign a nonzero value to the fund.

Finally, we argue that for application purposes, one does not need to rely on asset pricing models to define an admissible performance measure. Instead, using our characterizations, an evaluator can estimate many admissible performance measures and NA-based measures from the available market data and by solving certain systems of pricing equations. For this purpose, we offer implementation guidelines and use a set of mutual fund data from Morning Star Inc. to illustrate the performance evaluation methods that result from our general analysis. As pointed out later, our "look into the data for performance measures" approach has quite a few advantages over existing ones. Most notably, it renders the evaluation measures independent of any asset pricing model.

This article is organized as follows. In Section 1, we outline the basic framework and define and characterize admissible performance measures. Section 2 studies properties of admissible measures. Section 3 discusses the extension of our analysis to allowing public information-conditioned portfolios as performance reference. With this extension, one can use the resulting measures to distinguish managers who possess only public information from those with both access to better information sources and more portfolio management skills. Section 4 offers estimation guidelines. Section 5 provides performance eval-

[3] See, among others, Admati et al. (1986), Admati and Ross (1985), Connor and Korajczyk (1986), and Grinblatt and Titman (1989), for further discussion.

uation results for a group of mutual funds. Section 6 concludes the article. The proof of each result is given in Appendix A.

1. Admissible Performance Measures

Consider a single-period market, with the understanding that this single period is taken as a snapshot of a multiperiod securities market. The basic setup consists of a set of traded securities, the investing public and, for simplicity, one privately informed portfolio manager.

Take as given N securities that constitute the investment universe for all investors including the portfolio manager, whose time-0 prices are normalized to one dollar and whose time-1 gross returns are denoted by x_n, for $n \in N$, where N is also the index set of the securities. Assume that the securities market is incomplete and that each x_n is a mean-square integrable random variable defined on some probability space $\{\Omega, F, Pr\}$, where Ω is the set of states of nature, F a sigma field on Ω, and Pr the probability measure. For convenience, let L^2 be the linear space of mean-square integrable random variables on $\{\Omega, F, Pr\}$. Following convention, for any random variable x, we write $x \geq 0$ if $Pr(x \geq 0) = 1$; $x > 0$ if $x \geq 0$ and $Pr(x \neq 0) > 0$; and $x \gg 0$ if $x > 0$ and $Pr(x \neq 0) = 1$.

The investing public is comprised of a large set of investors that possess strictly increasing, continuous and convex preferences over future portfolio payoffs. For now, each individual investor (other than the portfolio manager) is taken to be uninformed, which is relaxed later when we discuss conditional performance measures. Then, for the investing public, the set of achievable gross returns per dollar invested is simply

$$R_0 \equiv \left\{ x \in L^2\colon\ \exists \alpha \in \Re^N \ \text{ s.t. } \sum_{n \in N} \alpha_n = 1 \ \text{ and } \sum_{n \in N} \alpha_n x_n = x \right\},$$

that is, the uninformed investors can only form nothing more than constant-composition portfolios. In this sense, their investment strategies are quite passive.

The informed portfolio manager, however, faces much better investment opportunities. Assume that the manager observes some private signal s, which contains information either about particular firms or about the entire securities market or both. Using this information and starting with one dollar, the manager can choose any portfolio, $\alpha(s) \equiv (\alpha_1(s), \ldots, \alpha_N(s))'$, such that $\sum_{n \in N} \alpha_n(s) = 1$ and $\sum_{n \in N} \alpha_n(s)\, x_n \in L^2$, where the amount invested in security n, $\alpha_n(s)$, is a function of the signal s. The restriction that $\sum_{n \in N} \alpha_n(s)\, x_n \in L^2$ means at least two things. First, the manager cannot fully observe at time 0 what everyone else will be able to observe at time 1. Second, the manager is not allowed to adopt strategies that generate future returns with unbounded variation, that is, every feasible strategy should lead to a square-integrable portfolio return.[4] The portfolio weight function, $\alpha_n(s)$, can be quite nonlinear, and many types of dynamic portfolio strategy are allowed here. For instance, Merton (1981) examines an option-like trading strategy in which the investor puts all the money in the market portfolio if he sees the market return to be high and otherwise all the money in the riskfree asset. Clearly, if the market signal is what the manager observes, he can adopt this strategy. Given the private signal, thus the manager's opportunity set is

$$R_s \equiv \left\{ x \in L^2\colon\ \exists \alpha(s) \ \text{ s.t. } \sum_{n \in N} \alpha_n(s) = 1 \ \text{ and } \sum_{n \in N} \alpha_n(s)\ x_n = x \right\}.$$

Since the informed manager can always choose to ignore the signal and adopt a constant-composition portfolio strategy, it is true that the set R_s contains the set R_0, which means the informed manager should in general do better than an uninformed individual investor. But there is no guarantee that the manager will choose the best dynamic strategy in R_s and make the most out of the signal. Therefore, possessing better information does not necessarily imply better performance.

To formally define the performance measurement problem, suppose that the manager has adopted some strategy that leads to a portfolio return $x_s \in R_s$. As in the existing performance literature, use the uninformed set R_0 as the performance benchmark. That is, any managed portfolio that does no better and no worse than any portfolio in R_0 deserves a zero performance ranking, while any portfolio doing better than some portfolio in R_0 deserves a positive performance ranking. With this in mind, take an arbitrary gross return $x' \in R_0$ as a reference and subtract it from the managed return x_s to get the excess return: $(x_s - x')$. Then, the performance measurement problem is to find some function $\lambda(\cdot)\colon\ L^2 \to \Re$ such that $\lambda(x_s - x')$ determines the performance for the managed portfolio and λ satisfies the following four conditions:

[4] This is an important technical requirement in order for our discussion to stay within the L^2 framework and for the standard statistical techniques to be used in empirical performance evaluation (since most statistical tools are developed assuming the L^2 framework). It nonetheless limits the set of admissible dynamic strategies. For instance, for certain square-integrable portfolio strategies [i.e., $\alpha_n(s) \in L^2$], the resulting portfolio returns will be absolute-integrable but not square-integrable (i.e., in L^1 but not in L^2).

Condition I. *If the managed return, x_s, lies in the uninformed set R_0, then it is given a zero performance: $\lambda(x_s - x') = 0$. That is, for every $x \in R_0$, $\lambda(x - x') = 0$.*

Condition II. *The function λ is linear. In other words, suppose there are two managed portfolios with returns $x_{s'}$ and $x_{s''}$; then, if a third portfolio is a constant combination of the two given by $\theta x_{s'} + (1 - \theta) x_{s''}$, for any constant $\theta \in \Re$, the performance of the third portfolio is determined by the same constant combination of the two portfolios' performance values:*

$$\lambda(\theta\, x_{s'} + [1 - \theta]\, x_{s''} - x') = \theta\, \lambda(x_{s'} - x') + [1 - \theta]\, \lambda(x_{s''} - x').$$

Condition III. *The function λ is continuous. That is, if the gross returns produced by two managed portfolios are arbitrarily close (in the mean-square metric), the performance values assigned to them will also be arbitrarily close: for any managed fund $x_{s'} \in L^2$ and any real number $\epsilon > 0$, there is a $\delta > 0$ such that $|\, \lambda(x_{s'} - x') - \lambda(x_{s''} - x')\, | < \epsilon$ for all (managed) portfolio returns $x_{s''}$ satisfying $\|x_{s'} - x_{s''}\| < \delta$, where $\| \cdot \|$ is the mean-square norm defined by $\|x\|^2 \equiv E(x^2)$.*

Condition IV. *The function λ is nontrivial, in the sense that if λ would be used to value any traded security, it would not assign a zero value, that is, for any $n \in N$, $\lambda(x_n) \neq 0$.*

These four conditions constitute a minimum set of requirements for any performance measure. Condition I states that any managed portfolio return that is achievable by an uninformed investor is automatically assigned zero performance. This condition is implicit in most existing performance measures. For instance, according to the Jensen measure, any portfolio on the security market line is assigned zero performance. According to Sharpe's (1992) style analysis, any manager whose portfolio is a linear combination of the reference funds is classified as having no ability. Condition II makes sense because a simple rescaling of a fund or a simple mix of two funds should not by itself improve one's performance ranking. Any change in performance ranking has to result from additional information. Condition III ensures that two managers producing indistinguishable returns are ranked the same. Condition IV guarantees that if a fund's excess return was proportional to the gross return on some security, this fund would not be assigned zero performance.

Conditions I and II imply that the choice of the reference return x' is inconsequential and any gross return in R_0 will serve the purpose. To briefly see this, take any $x'' \in R_0$. By Condition I, $\lambda(x'' - x') = 0$, which means, by Condition II, $\lambda(x'') = \lambda(x')$. Thus, $\lambda(x_s - x') = \lambda(x_s) - \lambda(x') = \lambda(x_s - x'')$. For this reason, we do not restrict the choice of the reference return to any particular payoff in R_0, even though in practice a common choice is the riskfree asset.

A function λ is said to be an admissible performance measure if it satisfies Conditions I through IV. Next, we seek to identify market conditions under which such measures exist.

Definition 1. *Let M be the linear span of R_0: $M \equiv \left\{\sum_{n \in N} \alpha_n x_n \colon \alpha_n \in \Re \right\}$. Then, the law of one price (LOP) is said to hold on the securities market if, for every $x \in M$, the set $\left\{\sum_{n \in N} \alpha_n \colon \alpha_n \in \Re \text{ and } \sum_{n \in N} \alpha_n x_n = x \right\}$ is a singleton.*

The LOP requires each payoff to have only one price, regardless of the portfolio composition that generates it. For more characterizations of the LOP, see Chen and Knez (1994, 1995) and Hansen and Jagannathan (1991, 1994).

Theorem 1. *There is an admissible performance measure if and only if the securities market supports the LOP.*

The LOP is thus a necessary and sufficient market condition for the existence of an admissible performance measure. For this reason, we also refer to admissible performance measures as LOP-based measures, whenever stating so adds convenience.

Theorem 2. *A function λ is an admissible performance measure if and only if there is some $d \in L^2$ such that d represents the function λ in the sense that*

$$\lambda(x - x') = E[d\,(x - x')] \qquad \forall x \in L^2, \tag{1}$$

and d satisfies

$$E(d\, x_n) = k \qquad \forall n \in N, \tag{2}$$

for some $k \in \Re$ and $k \neq 0$. Furthermore, each admissible performance measure λ is uniquely represented by such a d satisfying Equation (2).

Every admissible measure must therefore be representable by some d as shown in Equation (1) and only functions that are representable in this way can be admissible performance measures.[5] This is a rather

[5] In representing every admissible performance measure in the inner product format, some sort of risk adjustment is implicitly done. To see this, expand Equation (1) to get

$$\lambda(x_s - x') = E(d)\,E(x_s - x') + \mathrm{Cov}(d,\ x_s - x'),$$

where $\mathrm{Cov}(\cdot, \cdot)$ is the covariance operator. The covariance term in this expression reflects how risk is rewarded or adjusted.

strong characterization. Many known performance measures fall into this admissible class. In particular, since the LOP is a weaker market condition than general equilibrium, most equilibrium-based performance measures are admissible.

Example 1. The Jensen measure. *Based on the CAPM, Jensen's α measure is*

$$J(x_s) \equiv [E(x_s) - x_0] - \frac{\text{Cov}(x_s, x_m)}{\sigma_m^2}[E(x_m) - x_0],$$

where x_0 and x_m are, respectively, the riskfree gross return and the gross return on the market portfolio, and σ_m is the standard deviation of x_m. From Dybvig and Ingersoll (1982), the stochastic discount factor implied by the CAPM is given by

$$d_j = \frac{1}{x_0}\left(1 - \frac{E(x_m) - x_0}{\sigma_m^2}[x_m - E(x_m)]\right). \tag{3}$$

It is then straightforward to verify that

$$J(x_s) = x_0\, E[d_j\,(x_s - x_0)] = E[d_j'\,(x_s - x_0)], \tag{4}$$

where the riskfree return is chosen as the reference return and $d_j' \equiv x_0\, d_j$. The Jensen performance measure is representable by the d_j' given above and thus admissible.

Example 2. The APT-based performance measure *[Connor and Korajczyk (1986) and Lehmann and Modest (1987)]. Suppose there are K orthogonal factors f_k, $k = 1, \ldots, K$. Following Connor and Korajczyk (1986), the APT-based measure is defined as*

$$J_{APT}(x_s) \equiv [E(x_s) - x_0] - \sum_{k=1}^{K} b_k \gamma_k,$$

where b_k is the k-th factor beta of x_s and γ_k is the k-th factor premium. Chen and Knez (1994, Theorem 2.1) show that the APT is equivalent to having a stochastic discount factor exactly representable as a linear combination of the K factors:

$$d' = \sum_{k=1}^{K} \beta_k f_k, \tag{5}$$

for some $\beta_k \in \Re$. Using the proof of Theorem 2.1 in Chen and Knez (1994), one can verify that $J_{APT}(x_s) = x_0\, E[d'\,(x_s - x_0)]$. Thus, the APT-based measure is also admissible in our sense.

Example 3. The Grinblatt and Titman (1989) period weighting measures. *Grinblatt and Titman propose the following performance measures:*

$$J_{GT} \equiv \sum_{t=1}^{T} w_t\,(x_{s,t} - x_{0,t}),$$

subject to the constraint that $\sum_{t=1}^{T} w_t = 1$ *and* $\text{plim}[\sum_{t=1}^{T} w_t\, r_{E,t}] = 0$, *where* T *is the number of sample periods,* $(x_{s,t} - x_{0,t})$ *is the period-*t *excess return on the managed fund over the riskfree rate,* w_t *is the weight assigned to period* t, *and* $r_{E,t}$ *is the excess return on some mean-variance efficient portfolio. It is clear that depending on how the period weights are chosen, the resulting performance measure may not be admissible in our sense, because it may not necessarily assign a zero performance to every uninformed portfolio in* R_0 *(i.e., Condition I may be violated). However, if an additional requirement is met, that is, the period weights* w_t *satisfy*

$$\sum_{t=1}^{T} w_t\,(x_{n,t} - x_{0,t}) = 0, \qquad \textit{for each traded security } n \in N, \tag{6}$$

then the resulting function J_{GT} *will define an admissible performance measure.*

Example 4. The Glosten and Jagannathan (1994) performance measures. *Glosten and Jagannathan use the intertemporal marginal rates of substitution (IMRS) from each investor's consumption portfolio choice problem as the basis to define a class of performance measures:*

$$J_{GJ}(x_s) \equiv E[m^i\,(x_s - x_0)], \tag{7}$$

where m^i *is investor* i*'s IMRS.*[6] *By the first-order condition of each investor's problem, the function* J_{GJ} *must satisfy Condition I and hence each* J_{GJ} *is an admissible performance measure that is representable by* m^i.

Another popular measure, the Sharpe measure, captures the manager's ability to diversify away idiosyncratic risks, and it does not assign the same performance ranking to every uninformed portfolio in R_0, which implies that the Sharpe measure is not an admissible performance measure in our sense (because it violates Condition I). The

[6] Cumby and Glen (1990) use the IMRS of some investor with a power utility function to determine the period weights in their implementation of the Grinblatt and Titman (1989) period weighting measures. In this sense, their measure also falls into this class of IMRS-based performance measures.

purpose of this measure is thus different and it is not supposed to reflect whether the manager has superior information. The Sharpe measure is ideal, for instance, for ranking uninformed portfolios within R_0.

The performance measures given in the above examples are only a few out of infinitely many admissible ones. The set of all admissible performance measures is completely identified by

$$\overline{D} \equiv \{d \in L^2\colon\ d \text{ satisfies the equation system in Equation (2) for some } k \in \Re\}.$$

Each $d \in \overline{D}$ defines a performance measure λ as in Equation (1). For instance, the d_j in Equation (3), the d' in Equation (5), and the m^i in Equation (7) are all in $\overline{D}$. The set $\overline{D}$ is infinite. To see this, suppose that d is in $\overline{D}$. Then, as can be checked, $\theta\, d \in \overline{D}$ for every nonzero scaler $\theta \in \Re$.

Let D be the set of random variables $d \in L^2$ satisfying

$$E(d\,x_n) = 1 \qquad \forall n \in N. \tag{8}$$

Following Hansen and Jagannathan (1991), we refer to each member $d \in D$ as a *stochastic discount factor*. Since Equation (8) is a special case of Equation (2), D is clearly a subset of $\overline{D}$ and each $d \in D$ defines an admissible performance measure. This subset has two distinctive features. First, the larger set $\overline{D}$ can be obtained by rescaling every stochastic discount factor in D: $\overline{D} = \{\theta\ d\colon\ d \in D$ and for any nonzero $\theta \in \Re\}$. Thus, the set D provides sufficient information about $\overline{D}$. Second, since the time-0 price of each gross return x_n is one dollar, Equation (8) implies that each stochastic discount factor $d \in D$ actually prices every traded security consistently with the market. It then becomes natural to focus attention on the basic set D. In the set D there is one unique stochastic discount factor, denoted by d^*, that is in both D and the linear span M. Properties of this d^* have been extensively studied by Chamberlain and Rothschild (1983), Chen and Knez (1994, 1995), Hansen and Jagannathan (1991, 1994), and Hansen and Richard (1987). For example, every stochastic discount factor $d \in D$ can be decomposed as $d = d^* + \varepsilon$, for some $\varepsilon \in L^2$ that is orthogonal to every x_n. As seen shortly, this d^* is of particular interest for performance evaluation.

As far as performance measurement is concerned, each admissible measure in D can be used in empirical implementations, except that some of the measures will require stronger economic assumptions. For instance, the Jensen measure requires the CAPM assumptions to hold; and the APT-based measures depend on whether the assumed factor structure for asset returns is correct. Certain assumptions underlying these and other related performance measures may not be empirically supported. An alternative to these "parametric" performance measurement methods is to look into the security price data and identify admissible performance measures from there. Literally speaking, every stochastic discount factor in D, and hence the resulting performance measure, can be obtained by solving the equation system in Equation (2) or (8). In doing so, the only market condition required is the LOP holding. This alternative approach allows performance measurement to be independent of the assumptions that come with an asset pricing model. One can, for instance, use the performance measure defined by d^* since d^* is easily estimable. We will return to this topic in later sections.

Conditions I through IV do not rule out the possibility that well-managed portfolios may actually be assigned a negative performance. To see this, suppose the managed return is such that $(x_s - x') > 0$, for some reference $x' \in R_0$, that is, x_s dominates x' even though the initial costs for both gross returns are one dollar. Clearly, the manager should be given a positive ranking. But, Conditions I through IV together cannot guarantee it. For example, the Jensen measure is admissible, but this measure is known to have the potential to assign negative performance to superior managers.[7] To avoid this undesirable potential, we impose an additional condition:

Condition V. *The candidate performance measure* λ *is positive, that is,* $\lambda(x_s - x') \geq 0$ *whenever* $x_s - x' \geq 0$ *and* $\lambda(x_s - x') > 0$ *whenever* $x_s - x' > 0$.

A function λ is then said to be an admissible positive performance measure if it satisfies Conditions I through V.

Definition 2. *The securities market is said to offer no arbitrage opportunity if, for every positive payoff* $x \in M$ *such that* $x > 0$*, there exists no portfolio vector* $\alpha \in \Re^N$ *such that (i)* $\sum_{n \in N} \alpha_n x_n = x$ *and (ii)* $\sum_{n \in N} \alpha_n \leq 0$*, that is, every positive payoff has a positive price.*

Theorem 3. *Suppose that there is some traded security* x_n *whose gross return is positive almost surely (e.g., the riskfree asset):* $x_n \gg 0$*. Then there exists an admissible positive performance measure if and only if the securities market is free of arbitrage opportunities. Furthermore,*

[7] Dybvig and Ingersoll (1982) found that the stochastic discount factor d_j embedded in the CAPM takes negative values in some states of nature. Then, given the relationship between the Jensen measure and d_j as determined in Equation (4), it is not surprising that the Jensen measure may assign negative performance to superior managers. See Admati and Ross (1985), Dybvig and Ross (1985), and Jensen (1968, 1969) for more discussion on this point.

any function λ is an admissible positive performance measure if and only if there are a constant $\eta > 0$ and a positive stochastic discount factor $d^+ \in D$, such that $d^+ \gg 0$ and

$$\lambda(x_s - x') = \eta\, E[d^+(x_s - x')] \qquad \text{for every possible return } x_s \in L^2. \quad (9)$$

Admissible positive performance measures can be found when and only when there is no arbitrage opportunity, and every such measure has to be representable by some positive stochastic discount factor d^+. For this reason, we also refer to such performance measures as NA-based measures (to distinguish them from the LOP-based measures). To give a couple of known examples, the IMRS-based performance measures of Cumby and Glen (1990) and Glosten and Jagannathan (1994) are positive and admissible—so long as the equilibrium assumptions for the IMRS hold in the economy; the period weighting measures of Grinblatt and Titman (1989) present another such example—so long as the period weights satisfy Equation (6).

As in the case of the LOP-based measures, the set of all NA-based measures can be identified independent of any asset pricing model. According to Theorem 3, this set is completely represented by the set D^+, defined as

$$D^+ \equiv \{d^+ \in L^2:\ d^+ \gg 0 \text{ and } d^+ \text{ satisfies Equation (8)}\}. \quad (10)$$

D^+ contains all positive stochastic discount factors, and it is a subset of D. Each d^+ together with a positive constant η defines an admissible positive performance measure as in Equation (9). Therefore, for empirical performance studies, one only needs to estimate a d^+ by using Equation (8) and substituting it into Equation (9), from which one can next go on computing the performance value for any managed fund. We provide a detailed estimation procedure in Section 3.

2. Properties of Admissible Performance Measures

In the preceding section, it was established that the set of admissible performance measures is infinite. Among other things, we address in this section two main questions. First, to what extent will the performance ranking of managed funds depend on the choice of performance measure? Second, suppose that a managed portfolio is assigned a positive performance by some admissible positive measure; what does this positive performance value mean?

2.1 The LOP-based measures

Consider a manager who observes a signal s and whose portfolio return per dollar invested, x_s, is not in the uninformed set R_0. In other words, x_s is not achievable by any uninformed investor. Then, what performance value will this manager be given? The answer depends on which admissible performance measure is used in the evaluation.

First, suppose that there is some admissible measure λ that assigns a nonzero performance to it: $\lambda(x_s - x') = v$, for some constant $v \neq 0$ and any reference return $x' \in R_0$. According to Theorem 2, a rescaling of λ gives another admissible performance measure, that is, $h\lambda(\cdot)$ is also admissible, for any nonzero constant $h \in \Re$. Then, for any number $v' \neq 0$, we can choose $h = \frac{v'}{v}$ and arrive at a performance value given by $h\lambda(x_s - x') = \frac{v'}{v} v = v'$. This is to say that any performance value can be assigned to this manager by suitably choosing the performance measure.

One may argue that the above pessimistic conclusion is due to the fact that any rescaling of an admissible performance measure generates another admissible one. As observed earlier, the set $\overline{D}$ contains all admissible performance measures and it is a rescaled version of the set of stochastic discount factors, D. The latter set captures all the information relevant for performance evaluation and no stochastic discount factor in D is a rescaling of another member in D. Then, will the above conclusion change if the evaluator only chooses a performance measure from those defined by the discount factors in D? Unfortunately, as the proof of the next theorem demonstrates, the answer will stay intact and the multiplicity of admissible performance measures thus makes performance evaluation difficult.

Theorem 4. *If the managed gross return x_s is achievable by an uninformed investor (i.e., $x_s \in R_0$), all admissible performance measures will assign zero performance to it. Conversely, if every admissible performance measure assigns zero performance to the fund, x_s must be achievable by every uninformed investor. However, when x_s is not achievable by any uninformed investor (i.e., $x_s \notin R_0$), then, for any desired performance value $v \in \Re$, there is an admissible performance measure λ that assigns this performance v to the informed manager: $\lambda(x_s - x') = v$, for any reference return $x' \in R_0$.*

The first part of the above result states that a managed fund x_s is in the uninformed set R_0 if and only if every admissible measure gives it a zero performance ranking. This characterization is quite useful for empirical work. To see this, fix a managed fund. Then, if there is any admissible measure that assigns a nonzero performance to it, the fund's return must lie outside of the set R_0 and the fund must therefore be providing a service by enlarging the uninformed investors' opportunity set. On the other hand, if one picks some measure and finds that the measure gives zero performance to the fund, this by itself does not represent enough evidence to conclude that the fund does not help enlarge the opportunity set faced by investors, because there may be other admissible measures that will assign nonzero performance values to it. This suggests that for empirical work one may need to use as many performance measures as possible to evaluate a mutual fund. This is sort of the positive side of the above theorem.

The second part of Theorem 4 is rather negative about performance evaluation. It goes beyond the negative results of Roll (1978). Roll shows that if one uses an inefficient market index to stand for the market portfolio in the Jensen measure, any performance value is possible even for an arbitrary uninformed portfolio. His point concerns mostly the use of inefficient market indices. Using our terminology, the Jensen measure will no longer be admissible when the betas are obtained using an inefficient market index, since in that case Condition I will be violated by the resulting Jensen measure. For this reason, the choice of an inefficient market index will only make the Jensen measure-based performance evaluation arbitrary. Theorem 4, however, says more than this. It says that even if an evaluator chooses among admissible performance measures, any performance value is possible for a managed portfolio return lying outside of R_0.

Theorem 5. *Suppose that a performance evaluator is to rank two managed funds respectively with gross returns x_s and $x_{s''}$, both of which are not achievable by uninformed investors: $x_s \notin R_0$ and $x_{s''} \notin R_0$. Further, suppose the noises in the two private signals, s and s'', are not perfectly correlated. Then, when one admissible performance measure, say λ, assigns a higher performance to fund x_s, that is, $\lambda(x_s - x') > \lambda(x_{s''} - x')$ for any reference return x', there must exist another admissible measure λ' that reverses the ranking: $\lambda'(x_s - x') < \lambda'(x_{s''} - x')$.*

The relative ranking of two funds according to one admissible measure can thus be reversed by another admissible measure. This result is also similar in spirit to, but different from, a related result in Roll (1978). The comments preceding Theorem 5 apply here as well.

The above two results together reinforce our earlier argument that Conditions I through IV may not impose a strong enough set of requirements for performance measures. For instance, not every admissible performance measure under these conditions will be increasing in the managed excess return $(x_s - x')$. That is, suppose there are two funds with gross returns x_s and $x_{s''}$, such that $(x_s - x') > (x_{s''} - x')$, for any reference return x'; Then, some admissible performance measures may even rank fund x_s lower than fund $x_{s''}$. Nonetheless, there are widely used performance measures that only satisfy Conditions I through IV, but not Condition V. The Jensen measure discussed in Example 1 is one such example. Therefore, if one is willing to accept and use the Jensen measure, one should not have problems using any measure satisfying Conditions I through IV. But, there are too many functions that satisfy these conditions. In order to get around these negative results, we have to impose more conditions such that the set of the then admissible performance measures becomes smaller.

2.2 The NA-based measures

The NA-based performance measures will still assign different performance values to the same managed portfolio, but they nonetheless possess several more desirable properties. First, observe the following properties which are not shared by the LOP-based performance measures:

1. If a managed fund x_s is such that $x_s - x' > 0$ for some reference return $x' \in R_0$, then every NA-based performance measure will assign a positive performance to this fund. This uniformity in sign among all NA-based measures is a result of Condition V.

2. If $x_s - x' < 0$ for some reference return x', then every NA-based measure will assign to it a negative performance. This follows from Theorem 3.

3. Since every NA-based measure λ^+ can be represented as $\lambda(x_s - x') = \eta\, E[d^+ (x_s - x')]$, for some constant $\eta > 0$ and some $d^+ \in D^+$ (Theorem 3), every NA-based performance measure is strictly increasing in the managed excess return $(x_s - x')$. This property is useful for ranking funds relative to each other. If one fund (x_s) does better than another $(x_{s''})$ in generating excess returns [i.e., $(x_s - x') > (x_{s''} - x')$], then each NA-based measure will rank the former higher than the latter.

More generally, performance values that can be assigned to a managed fund by an NA-based measure will lie in a range. Depending on the nature of the fund, this range can be a small segment of the real line or as wide as the entire real line (as in the case of the LOP-based measures). Thus, we have four different cases. (i) Suppose that there are two NA-based measures, λ^+ and ψ^+, one assigning a positive and the other assigning a negative performance to the managed fund x_s: $\lambda^+(x_s - x') > 0$ and $\psi^+(x_s - x') < 0$. Since both $\eta\,\lambda^+$ and $\eta\,\psi^+$ are still admissible positive performance measures for any positive number $\eta > 0$, the range of possible performance values assigned by the NA-based measures is the entire real line $\Re$. (ii) If there is some NA-based measure that assigns a positive performance to x_s but no NA-based measure gives it a negative performance, then the range of possible performance values assigned to x_s by any NA-based measure will be the positive segment of the real line. The justification is the same as in case (i). (iii) If there is some NA-based measure that assigns a negative performance to x_s but no NA-based measure gives it a positive performance, then the range of possible performance values will be the negative segment of the real line. (iv) Suppose that the performance evaluator only chooses a measure from those defined by the positive discount factors in D^+. An acceptable justification for this restriction is that every such d^+ can price the traded securities consistently with the market, while every other NA-based measure is only a rescaled version of the NA-based measure defined by some d^+ in D^+. Then, performance evaluation is much less arbitrary and the range of possible performance values with these NA-based measures is a possibly open interval (ℓ, u), with $\ell \equiv \inf\{E[d^+(x_s - x')]: \ d^+ \in D^+\}$ and $u \equiv \sup\{E[d^+(x_s - x')]: \ d^+ \in D^+\}$. The convexity of the set D^+ ensures that for any value between ℓ and u, there is a positive discount factor that defines an NA-based measure assigning this performance value to x_s. See Kreps (1981) for a related discussion in a different context.

Theorem 6. *Let x_s be the managed gross return and x' be any reference return.*

(i) There exists an NA-based measure λ that assigns a positive performance to the fund (i.e., $\lambda(x_s - x') > 0$) if and only if there is an investor in the monotonically increasing, convex and continuous preference class who values x_s more than any uninformed return in R_0;

(ii) There exists an NA-based measure λ that assigns a negative performance to the fund (i.e., $\lambda(x_s - x') < 0$) if and only if there is an investor in the monotonically increasing, convex and continuous preference class who values x_s less than any portfolio in R_0.

Note that each d^+ can be thought of as some investor's IMRS and hence each NA-based measure can be interpreted as some investor's marginal valuation function. Then, it may not be any surprise that if a managed fund is ranked positively by some NA-based measure, there must exist at least one investor who would like to hold more of this fund rather than of any portfolio in R_0. Conversely, if a fund is valued by some investor over every uninformed portfolio, at least the NA-based measure defined by this investor's IMRS will rank the fund positively. This result is quite encouraging since it ensures that positive performance according to any NA-based measure certifies superior portfolio service to the investor community.

Aided by Theorem 6, we can also interpret the performance value, $\lambda^+(x_s - x')$, as the "equilibrium management fee" that an uninformed investor is willing to pay per dollar invested. To see this, suppose λ^+ can be represented by some positive d^+ such that $\lambda^+(x_s - x') = E[d^+(x_s - x')]$. From the proof of Theorem 6, we can interpret this d^+ as some investor's intertemporal marginal rate of substitution. Then, if $E[d^+(x_s - x')] > 0$, this positive quantity indicates how much the investor's total expected utility may increase at the margin when he sells short one dollar of the reference portfolio x' and invests the proceeds in the managed fund x_s. In other words, $E[d^+ (x_s - x')]$ reflects the marginal value of the fund to the investor. This quantity is in some sense a price for both the information s and the manager's skills in using it. For a discussion on equilibrium management fees, see Merton (1981).

According to the second part of Theorem 6, negative performance will be assigned to a managed fund by some NA-based measure when and only when some investor in the stated class would rather sell short the fund and invest more in any uninformed portfolio in R_0. Then, from this investor's perspective, the manager has not used the private information to this investor's advantage and has produced an inferior portfolio.

There can be managed funds that are assigned a positive performance by some NA-based measures but a negative performance by others. In that case, it means that such a fund is valued and appreciated by some investors but disliked and not wanted by others. Given that mutual funds are set up to satisfy different investor clienteles, such an evaluation outcome may not be unrealistic.

Suppose in an empirical performance study one finds that an LOP-based measure assigns a negative performance value to a given fund. Then, is it possible for there to also exist an NA-based measure that ranks the same fund positively? The answer is a definitive yes—as long as there is an investor who values this fund more than any reference portfolio in R_0. To see this, let x_s be such a fund. By Theorem 4, any negative performance can be assigned to x_s by some LOP-based measure. By Theorem 6, there is also an NA-based measure that ranks this fund positively. Therefore, cases can arise in which an LOP-based measure and an NA-based measure assign to the same fund performance values of opposite signs. For an example, see the findings in Section 5.

3. Extension to Public Information-Conditioned Benchmarks

In the discussion so far, the reference set of portfolio returns with respect to which zero performance is determined has been the uninformed set R_0. This has also been the general practice in the existing literature. Recent empirical studies have, however, documented that it is possible to use such public information as interest rates to forecast stock returns. See, for example, Fama and Schwert (1977), Ferson (1989), Ferson and Harvey (1991), Ferson and Korajczyk (1995), and Keim and Stambaugh (1986). Breen, Glosten, and Jagannathan (1989) show that managers who time using publicly available information will be classified by traditional evaluation methods as possessing superior ability. In light of this, it is necessary to expand the performance reference set from R_0 to include all those returns that are achievable using public information, so that managers who solely rely on public information will be classified as zero performers and only those who use private information efficiently will be classified as superior managers.

To do this, suppose the public information signal p is coarser than the manager's private signal s. Accordingly, the set of gross returns achievable with public information is

$$R_p \equiv \left\{ x \in L^2 \colon \exists \alpha(p) \text{ such that } \sum_{n \in N} \alpha_n(p) = 1 \right.$$
$$\left. \text{and } \sum_{n \in N} \alpha_n(p)\, x_n = x \right\}. \tag{11}$$

Since the signal p is coarser than the manager's s, we have $R_0 \subset R_p \subset R_s$.

Aided with the tools from Hansen and Richard (1987), we can replace the reference set R_0 in the previous sections by R_p and all the results that we established will still hold, provided that appropriate changes are made in the interpretation. In this case, Condition I is replaced by requiring λ to satisfy

$$\lambda(x_p - x') = 0, \qquad \forall x_p \in R_p,$$

for any given reference return $x' \in R_p$. This ensures that managers who rely on no more than public information will be given zero performance. A function λ is said to be an *admissible conditional performance measure* if it satisfies the extended Condition I and the linearity, continuity, and nontriviality conditions.

The conditional counterpart to Theorem 2 is that a function λ is an admissible conditional performance measure if and only if there is some $d_p \in L^2$ such that

$$\lambda(x - x') = E[d_p\,(x - x')] \qquad \forall x \in L^2,$$

and

$$E(d_p\, x_p) = k \qquad \forall x_p \in R_p,$$

for some $k \in \Re$ and $k \neq 0$. For this reason, we sometimes refer to such a measure as a *conditional LOP-based measure*. The conditional counterpart to every other result holds as well. To save space, we omit the detailed discussion which would be a virtual repetition of the preceding two sections.

In general, replacing the reference set R_0 by the publicly informed set R_p should mean a "tougher" benchmark for performance evaluation. Only managers who use more than public information have the potential to be assigned a positive performance by a conditional performance measure. Does this imply that for a given fund, any conditional measure will assign a performance value lower than what an unconditional LOP-based measure does? Given the infinity of admissible conditional and unconditional measures, the answer is again negative. To see this, let x_s be a managed fund such that no publicly informed investor can achieve it with one initial dollar: $x_s \notin R_p$. Naturally, $x_s \notin R_0$. First, for an arbitrary negative number, there is by Theorem 4 an unconditional LOP-based measure that assigns this performance value to the fund (negative unconditional performance). Next, for any positive number, there is, by the conditional counterpart to Theorem 4, an admissible conditional measure that assigns this positive performance to the same fund (positive conditional performance). This explains the empirical finding by Ferson and Schadt (1992) that for some funds performance switches from negative to positive when they change from an unconditional measure to a conditional one (also see our empirical results in Section 5). Therefore, due to the nonuniqueness of admissible measures, switching from unconditional to conditional performance measures does not necessarily mean lowering the performance ranking of every fund.

4. Implementing Performance Measurement Tests

The main message from the discussion so far seems quite negative about performance evaluation. In the case of LOP-based measures, basically nothing can be said about a manager's ability, unless the manager's portfolio happens to lie in the span of the reference assets. Since every performance value is possible for a managed fund lying outside of the reference set, the relative ranking between two funds also becomes quite arbitrary. With Condition V imposed, performance evaluation is made less arbitrary, but still any performance value within a certain range can be assigned to the same fund by the choice of the right NA-based measure. In spite of the negative conclusions, portfolio performance evaluation as a practical matter is, and will always be, conducted so long as there is investment management. For this reason, in what follows we present and illustrate the use of several nonparametric performance evaluation procedures. We refer to them as being nonparametric because the validity of these performance measures does not depend on any parametric or equilibrium asset pricing model such as the CAPM and the consumption CAPM. Rather, they rely only on either the LOP or the NA condition. In addition, as will become clear shortly, they also offer implementational ease. We first discuss the issues involved in selecting the set of reference assets, and then describe detailed procedures for respectively estimating a particular LOP-based and a particular NA-based measure.

4.1 Selecting the reference assets

There are two choices that must be made in specifying the set of reference assets. The first is how many assets to include in the reference set. The second is which assets to include. While in theory this set should include all assets available to the investing public, the econometrician can only use at best a finite sample in the estimation. This imposes limitations on the number of assets that the investigator can include. The significance of this constraint depends on the estimation procedure used. In the next subsection we describe the relationship between the estimation method and the number of assets that may be included.

The decision of which assets to include is guided first by the type of assets in which the fund invests. That is, it should at least include the same assets that comprise the fund under evaluation. This is because the mission of performance assessment is to investigate whether the investment opportunity set has been significantly enlarged by the dynamic trading strategy of the fund. If we find the investment opportunity set has been enlarged, we want this to be attributed to the efforts of the fund manager as embodied in the trading strategy. In other words, we do not want the existence of performance to be caused by the inclusion of certain reference assets and the exclusion of others. For example, a small-firm fund invests mainly in firms with market capitalizations less than $1 billion. This implies that the set of reference assets should include small cap stocks when such a mutual fund is evaluated. In addition, many funds have a significant portion of their capital in cash such as short-term Treasury bonds. Thus, for the evaluation of those funds, Treasury securities may be included in the reference set.

In theory, the reference assets should include more than just the subset from which the informed manager has decided to choose. Typically, evaluation results depend on the set of reference assets used. For a good example, Elton et al. (1993) explicitly investigate the sensitivity of performance assessment to non-S&P stocks and bonds. They find that both security types are significant factors for the correct assessment of performance. As in other empirical studies, a trade-off has to be made here between economic and statistical power as well.

4.2 Test methods

In this subsection we discuss the econometric methods used to conduct performance evaluation.[8] Recall from Section 1 that each stochastic discount factor in D and D^+ defines an admissible measure. Hansen and Jagannathan (1991, 1994) have put forward two particular stochastic discount factors, denoted by d^* and d^+, where d^+ is as estimated in Hansen and Jagannathan (1991). Since these two discount factors are probably among the most familiar ones, we choose to focus on the two measures, denoted λ and λ^+, that they respectively define.

4.2.1 Estimating λ and λ^+. Assume that asset price processes are sufficiently regular that a version of the time-series law of large numbers applies so that the sample moments converge to the corresponding population moments. For this reason, unless it is necessary, we omit the time subscripts. Depending on the context, we use X to stand for either the $N \times 1$ vector of gross returns on N reference assets or the $N \times T$ matrix of time-series observations on the gross returns of the assets. In the latter case, X_t stands for the $N \times 1$ vector of gross returns on the N assets during the t-th period. Assume that $E[XX']$ is nonsingular.

Method 1. By construction the stochastic discount factor d^* is the payoff on some constant composition portfolio whose weights are given by some α^* in $\Re^N$, that is, $d^* = X'\alpha^*$, satisfying

$$\mathbf{1}_N = E[X\, d^*] = E[X\, X'\alpha^*], \tag{12}$$

where $\mathbf{1}_N$ is an $N \times 1$ vector of 1s. Thus, $\alpha^* = E[XX']^{-1}\,\mathbf{1}_N$. For a managed fund with gross return x_s, its LOP-based performance value is

$$\lambda = E[x_s\, d^*] - 1 = E[x_s\, X'\alpha^*] - 1. \tag{13}$$

The performance test is on whether λ is significantly different from zero, which is conducted in two steps:

Step 1. Analytically compute $d^* = X'\alpha^*$, where $\alpha^* = E[XX']^{-1}\,\mathbf{1}_N$. For the t-th period, the performance value for x_s is given by

$$\hat{\lambda}_t(\alpha^*) = x_{s,t}\,(X_t'\,\alpha^*) - 1. \tag{14}$$

[8] The methods described here employ a standard GMM procedure so as to accommodate heteroskedasticity and serial correlation issues. For discussions on using GMM to estimate stochastic discount factors, see Chen and Knez (1994, 1995), Cochrane and Hansen (1992), Knez (1994), and Snow (1991). While there are good reasons to use GMM for this estimation problem, theory does not require the use of this estimation procedure. See He, Ng, and Zhang (1994) for an alternative method of estimating stochastic discount factors.

Step 2. To test if $\hat{\lambda}_t$ is significantly different from zero, form a h_T statistic:

$$h_T = T\left[\frac{1}{T}\sum_{t=1}^{T}\hat{\lambda}_t(\alpha^*)\right] W_T \left[\frac{1}{T}\sum_{t=1}^{T}\hat{\lambda}_t(\alpha^*)\right], \tag{15}$$

where $W_T = \frac{1}{\hat{\sigma}_\lambda^2}$, with $\hat{\sigma}_\lambda^2$ being a consistent estimate of the variance of $\hat{\lambda}$ [see Newey and West (1987)]. Under the null hypothesis that $E[\hat{\lambda}_t(\alpha^*)]$ is zero, h_T is asymptotically χ^2-distributed with one degree of freedom.

Tests on whether the NA-based measure, λ^+, is significantly different from zero can be conducted by replacing d^* in Equations (12) through (14) with d^+. The counterpart to Equation (12) is

$$E[X\,(X'\alpha)^+] = \mathbf{1}_N, \tag{16}$$

where $(X'\alpha^*)^+ \equiv max(X'\alpha^*, 0)$ and $(\cdot)^+$ stands for the "truncation at zero" nonnegative operator. The estimation of $d^+ = (X'\alpha^*)^+$ is significantly more complicated than that of d^*, because of the nonlinear representation in α. Hansen and Jagannathan (1991) develop a procedure in which the objective is to search for a portfolio payoff whose nonnegative truncation satisfies Equation (16) and that has the minimum second moment. Realizing that one does not have to use this minimum second-moment d^+ and that any d^+ satisfying Equation (16) is equally qualified for performance evaluation, we nonetheless choose to adopt the Hansen and Jagannathan procedure to generate a d^+ for performance evaluation. See Hansen and Jagannathan (1991) for details. The h_T statistic for the measure λ^+ implied by this minimum second-moment d^+ can be similarly defined as for the LOP-based measure.

Since d^* can be computed analytically, a relatively large number of reference assets can be included in X—so long as the second-moment matrix, $E(X\,X')$, can be conveniently inverted. In contrast, since the estimation of d^+ involves a system of equations nonlinear in α, only a relatively small number of reference assets can be included in conducting performance evaluations using the NA-based measure.

Method 2. The procedure in Method 1 involves first computing the stochastic discount factor and then using it to evaluate the performance of a mutual fund. One drawback of the test statistic h_T is that it ignores the potential sampling error in computing d^* from a finite number of observations. An alternative procedure is to simultaneously estimate α^* and test if λ is significantly different from zero.

Step 1. Form the system of moment equations. Let $\widetilde{X}$ be an $(N+1)\times 1$

vector, the first N elements of which are comprised of the N elements in X and the last element of which is the return on the candidate mutual fund, that is, $\widetilde{X} = (X', x_s)'$. Then, let $\alpha \equiv (\alpha^x, \alpha^s)'$, where $\alpha^x \in \Re^N$ and $\alpha^s \in \Re$, and form the following system of $(N + 1)$ equations:

$$E\left[\widetilde{X}\,(\widetilde{X}'\,\alpha)\right] = \mathbf{1}_{N+1}. \tag{17}$$

The test on whether λ is significantly different from zero is conducted by testing if there is a solution, α^*, to Equation (17) such that the corresponding component $\alpha^{*,s}$ is zero.

Step 2. Note that the system of equations in Equation (17) has $N + 1$ moment conditions and $N + 1$ parameters to be estimated. However, since α^s is hypothesized to be zero, this system is overidentified, which permits the use of a GMM procedure [see Hansen (1982) and Hansen and Singleton (1982)]. Define

$$\lambda_t(\alpha) \equiv \widetilde{X}_t\,\widetilde{X}_t'\alpha - \mathbf{1}_{N+1}. \tag{18}$$

Equation (18) implies that, when evaluated at $\alpha = \alpha^*$, $E[\lambda_t(\alpha^*)] = 0$.

Step 3. Form the sample moments and estimate the parameters. This is done by replacing the population mean in Equation (17) with the sample mean:

$$\bar{\lambda}(\alpha) \equiv \left(\sum_{t=1}^{T} \widetilde{X}_t\,\widetilde{X}_t'\,\alpha - \mathbf{1}_{N+1}\right) / \,T, \tag{19}$$

and choose α to minimize the scaled J-statistic given by

$$J_T \equiv T\,\bar{\lambda}(\alpha)'\,W\,\bar{\lambda}(\alpha),$$

where W is a symmetric positive, definite weighting matrix. When the set of overidentifying restrictions holds, J_T is approximately χ^2-distributed with one degree of freedom.

To test if the performance value for x_s is significantly different from zero according to the NA-based measure, we need to replace $(\widetilde{X}'\,\alpha)$ in Equation (17) with $(\widetilde{X}'\,\alpha)^+$. The test is then on whether there is a solution, α^*, to

$$E\left[\widetilde{X}\,(\widetilde{X}'\,\alpha)^+\right] = \mathbf{1}_{N+1},$$

such that the $(N+1)$ component, $\alpha^{*,s}$, equals 0. The J-statistic for this case can be formed as before with some minor adjustments.

Note that this method also has a single degree of freedom. However, in this method the number of reference assets that can be included should in general be smaller than in Method 1, out of the consideration that there is a trade-off between the number of moment conditions used in estimation and the precision with which the optimal weighting matrix may be estimated.[9]

4.2.2 Estimating conditional λ. This subsection describes how to estimate and test the significance of λ using conditional performance measures. Let d_p^* represent a stochastic discount factor that can consistently price every payoff generated by a public information-based dynamic trading strategy.[10] In order to compute d_p^*, in principle we should include all publicly observable information and consider all feasible payoffs conditioned on such information. However, only a subset of such information is available to the econometrician. Denote by Z_t the column vector of K information variables observable by the econometrician as of period t. To construct d_p^* we first need to identify the reference set of returns, R_p. For this purpose we follow Hansen and Singleton (1982) in assuming that expectations conditional on public information are linear in the information variables in Z_t. In effect, this means that all public information-conditioned reference returns are generated with linear trading strategies. That is,

$$R_{p,t+1} = \left\{ \sum_{n \in N} (\alpha_n' Z_t)\, x_{n,t+1}\colon\ \alpha_n \in \Re^K,\ \forall n, \right. \\ \left. \text{such that } \sum_{n \in N} \alpha_n' Z_t = 1 \right\},$$

where $(\alpha_n' Z_t)$ is the Z_t-conditioned portfolio weight in security n and, for convenience of discussion, a time dimension is explicitly incorporated, with $R_{p,t+1}$ being the set of time $(t+1)$ gross returns achievable using public information. Let α be an $N \times K$ matrix, with its n-th row being the transpose of the column vector α_n. Since we choose $d_{p,t+1}^*$ to be the stochastic discount factor that lies in the linear span of $R_{p,t+1}$, there must be some matrix $\alpha^* \in \Re^{NK}$ such that

$$d_{p,t+1}^* = X_{t+1}' \alpha^* Z_t, \tag{20}$$

and that

$$E\left[(X_{t+1}'\alpha' Z_t)\, X_{t+1}' \alpha^* Z_t\right] = 1, \tag{21}$$

[9] In the context of estimating a stochastic volatility model, Andersen and Sorensen (1994) recently used Monte Carlo simulations to argue that (i) the number of moment conditions should be small (e.g., 14) for typical sample sizes and (ii) when too many moment conditions are included the p-values associated with overidentifying restriction tests become inflated, causing the tests to underreject.

[10] See Farnsworth et al. (1995) and Ferson and Schadt (1992) for an alternative method for constructing conditional performance benchmarks.

for any matrix $\alpha \in \Re^{NK}$ such that $\sum_{n \in N} \alpha_n' Z_t = 1$ (i.e., $X_{t+1}' \alpha' Z_t \in R_{p,t+1}$). Alternatively, the requirement for $d_{p,t+1}^*$ as given in Equation (21) can be replaced by $E(X_{t+1} d_{p,t+1}^* \mid Z_t) = \mathbf{1}_N$, which means that the pricing errors for the N assets are orthogonal to Z_t, that is,

$$E\left[(X_{t+1} d_{p,t+1}^* - \mathbf{1}_N) Z_t'\right] = 0_{N \times K}, \tag{22}$$

where $0_{N \times K}$ is an $N \times K$ matrix of 0s.[11]

Substituting Equation (20) into Equation (22) and replacing each moment by its time-series sample counterpart produces

$$\frac{1}{T} \sum_{t=1}^{T} \left\{(X_{t+1} X_{t+1}' \alpha^* Z_t - \mathbf{1}_N) Z_t'\right\} = 0_{N \times K},$$

or

$$\sum_{t=1}^{T} (Z_t Z_t' \otimes X_{t+1} X_{t+1}') \, \mathbf{vec}(\alpha^*) = (Z \otimes \mathbf{1}_N) \, \mathbf{1}_T,$$

where Z is the $K \times T$ matrix with its t-th column being Z_t, which gives

$$\mathbf{vec}(\alpha^*) = \left[\sum_{t=1}^{T} (Z_t Z_t' \otimes X_{t+1} X_{t+1}')\right]^{-1} (Z \otimes \mathbf{1}_N) \, \mathbf{1}_T. \tag{23}$$

Together, Equations (20) and (23) provide the desired expression for the computation of d_p^*. With this d_p^*, we can conduct conditional performance evaluations on any managed fund by following the same steps as in Method 1 for the unconditional case.

One can also conduct conditional performance evaluations following a one-step, simultaneous estimation procedure. Recall that $\widetilde{X}$ is the stacked $(N + 1)$ vector with the first N components given by X and the last by the managed fund x_s. Replacing the X in Equation (22) by $\widetilde{X}$, we have

$$E\left[(\widetilde{X}_{t+1} \widetilde{X}_{t+1}' \tilde{\alpha}^* Z_t - \mathbf{1}_{N+1}) Z_t'\right] = 0_{(N+1) \times K}, \tag{24}$$

where $\tilde{\alpha}^*$ is now an $(N + 1) \times K$ matrix. As before, Equation (24) represents an overidentified system under the null hypothesis that $[\tilde{\alpha}_{N+1}^*]' Z_t = 0$. This system can again be tested using GMM. Relying on $d_{p,t+1}^* = \widetilde{X}_{t+1}' \tilde{\alpha}^* Z_t$, one can use the estimated d_p^* from the GMM procedure and proceed with conditional performance assessments for the managed fund under consideration.

[11] We thank Heber Farnsworth for suggesting this alternative derivation.

5. An Illustration

Following Method 2, we illustrate in this section the use of the performance measures respectively defined by d^*, d^+, and d_p^*. Besides the purpose of illustration, we also make the collection of reference assets *both* larger than typically used in the literature *and* interesting to a typical investor, so that the evaluation results to be discussed are of independent interest. Specifically, 68 equity mutual funds are included in the study and they are classifiable into five classes: growth, growth-income, income, stability-growth-income, and maximum capital gain. We choose these funds because they have been evaluated using the performance measures given in Examples 1 and 2 of Section 1. The five classifications, described in Weisenberger's Investment Companies annual compendium, reflect the funds' investment objectives. For each of the 68 funds, monthly returns net of transactions costs are used in our estimation, covering the period from January 1968 to December 1989. A more detailed description of the data used is contained in Appendix B.

We address four questions regarding the performance of equity mutual funds: (1) Do funds, on average, exhibit significant abnormal performance? (2) Do there exist mutual funds (as a group or on an individual basis) that exhibit strong evidence of abnormal performance? If so, is this abnormal performance positive or negative? (3) Does the performance of mutual funds vary substantially across fund types or across funds in general? (4) How are the answers to these questions affected when conditioning information is incorporated into the construction of the benchmark?

Table 1 summarizes the unconditional performance results for the five fund groups.[12] In the table the p-value is reported for the hypothesis that the equally weighted portfolio of funds of a given type performs no better than a constant-composition portfolio. Consider first the LOP-based measure in panel A of Table 1. For four of the five groups there is insufficient sample evidence to reject the hypothesis that λ is not different from zero at all the standard levels of significance. For only one class of funds, stability-growth-income, is there sufficient sample information to reject the hypothesis at the 10 percent confidence level. However, at the 5 percent confidence level this hypothesis is not rejected even for this group. Alternatively, we can

[12] In the tables we only report results obtained for the entire sample period 1968:01 to 1989:12. In order to check the stability of our results, we also investigated the performance of the 68 funds over two subperiods: 1968:01 to 1978:12 and 1979:01 to 1989:12. We found that for most funds the performance results over the subperiods do not differ significantly from what is reported in the tables. For space considerations, these results are not included here. They are, however, available from the authors upon request.

Table 1
Unconditional performance evaluation by fund groups

Fund type	λ	χ^2	p-value
Panel A: LOP-based measure			
Growth	−0.06	1.92	0.166
Growth-income	−0.91	1.11	0.292
Income	−0.37	2.28	0.105
Stability-growth-income	0.13	2.95	0.086
Maximum capital gain	−2.4	2.46	0.117
Panel B: NA-based measure			
Growth	0.19	2.08	0.149
Growth-income	−0.65	1.22	0.268
Income	−0.11	2.35	0.125
Stability-growth-income	0.69	3.15	0.076
Maximum capital gain	−1.61	2.68	0.102

The annualized monthly returns for all funds of a given type are equally weighted to form a return time series for the fund type. The sample time period is from 1968:01 to 1989:12. Unconditional performance evaluation is then done on each of the five equally weighted return series, one for each fund type. The tests use the generalized method of moments and the Newey and West (1987) procedure to allow for heteroskedasticity and serial correlation. The lag length is set at 17 for all fund types. The number of observations is 264. The tests are χ^2 tests of the overidentifying restriction with one degree of freedom. Expressed in basis points, λ is the performance value, based on either the LOP- or the NA-based measure.

draw inference based on the χ^2 statistic. For a χ^2 variate with one degree of freedom, a significance level of 5 percent (for the p-value) corresponds to a χ^2 statistic of 3.84. In Table 1 (and in the other tables as well), the reported χ^2-statistic is always below 3.84. Thus, according to this statistic, there is also insufficient evidence to reject the null hypothesis of no abnormal performance.

Panel B of Table 1 contains the results for the NA-based performance measure. The p-values for the NA-based measure are slightly smaller than those for the LOP-based measure. As with the LOP-based measure, for four out of the five groups the value of λ is statistically insignificant at standard confidence levels, and the hypothesis of no abnormal performance is not rejected at the 5 percent level for all fund groups. For the stability-growth-income group, its λ under the NA-based measure is positive and statistically significant at confidence levels greater than 7.6 percent. This means, for example, that at the 10 percent confidence level we would reject the null hypothesis that λ is zero and assign positive performance to this group.

Table 2 presents the simple averages of the individual estimates for each fund within a group. For all five groups, the average p-values suggest that at all standard significance levels there is no abnormal

Table 2
Unconditional performance evaluation by individual funds

Fund type	λ	χ^2	p-value
Panel A: LOP-based measure			
Growth	−0.868	1.29	0.277
Growth-income	−1.71	3.21	0.150
Income	−0.984	1.91	0.338
Stability-growth-income	−0.597	1.97	0.242
Maximum capital gain	−0.974	1.63	0.338
All	−1.026	2.00	0.269
Panel B: NA-based measure			
Growth	−0.760	1.34	0.267
Growth-income	−1.79	3.27	0.141
Income	−1.03	2.02	0.332
Stability-growth-income	−0.461	2.13	0.228
Maximum capital gain	−0.926	1.73	0.333
All	−1.03	2.09	0.246

The sample time period is from 1968:01 to 1989:12. Unconditional performance evaluation is conducted for each individual mutual fund. The tests use the generalized method of moments and the Newey and West (1987) procedure to allow for heteroskedasticity and serial correlation. The lag length is set at 17 for all fund types. The number of monthly observations is 264. The tests are χ^2 tests of the overidentifying restriction with one degree of freedom. Expressed in basis points, λ is the simple average of individual performance values for all funds in a given group, based on either the LOP- or the NA-based measure. The reported χ^2-statistic and p-value are, respectively, the simple average of individual χ^2-statistic values and individual p-values for all funds in a given group. Among fund types, the rows under type "All" correspond to the simple averages of all 68 funds included.

performance. In addition, λ is negative on average for all funds. Similar conclusions follow for the NA-based measure (see panel B).

The results in Tables 1 and 2 suggest that on average, or on a group-by-group basis, these mutual funds display very little evidence of abnormal performance. Table 2 also reports the average of the unconditional LOP- or NA-based λ values for all 68 funds (under group type "All" in Table 2). Based on the average p-value for all funds included, there is insufficient sample information to reject the hypothesis of zero performance for all funds as a whole, at all standard significance levels. For all the funds included, the NA-based measure on average assigns a negative performance, while the zero performance hypothesis would be rejected at confidence levels greater than 24.1 percent.

It is important to understand if the lack of performance can be attributed to the extremely poor performance on the part of a few funds or all funds. To address this issue, we show in Figures 1 and 2 the p-values for the individual mutual funds within each fund group.

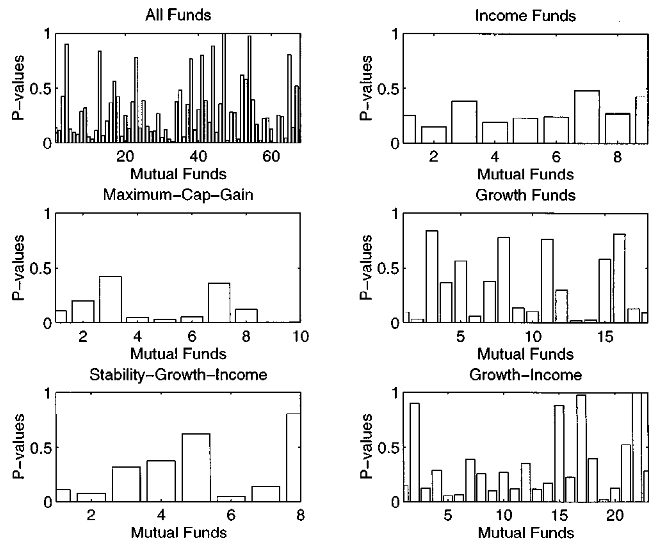

Figure 1
Unconditional LOP-based performance measurement
The *p*-values are from the χ^2-test of the null hypothesis of no abnormal performance using the LOP-based measure. Each *p*-value represents the probability that the corresponding variate is greater than the sample value of the test statistic. The test statistic is asymptotically χ^2-distributed with one degree of freedom.

These figures demonstrate that the *p*-values exhibit a large degree of variation across funds. For example, the minimum and the maximum *p*-values are, respectively, 0.3 percent and 97 percent for the LOP-based measure and 0.3 percent and 96 percent for the NA-based measure. Eight percent of the funds have a *p*-value less than 5 percent using the LOP-based measure, and 13.2 percent of the funds have a *p*-value less than 5 percent using the NA-based measure (see Table 3). Of the 68 funds, the number of funds with positive performance values is 11 by the LOP-based measure and 7 by the NA-based measure. Recall that zero performance for the equally weighted portfolio of stability-growth-income funds is rejected at the 10 percent confidence level, while the average *p*-value for the funds in this group is 24.2 percent. In Figure 1, three of the seven funds in this group show *p*-values less than 10 percent while three others have *p*-values between 20 and 65 percent. Thus, examining each fund separately may give one more information about mutual fund performance.

In the maximum capital gain group, there are two funds for which zero performance is strongly rejected by the data, with their *p*-values at 0.3 and 0.4 percent. In addition, both of these funds have positive NA-based performance values. Since the mutual funds are classified

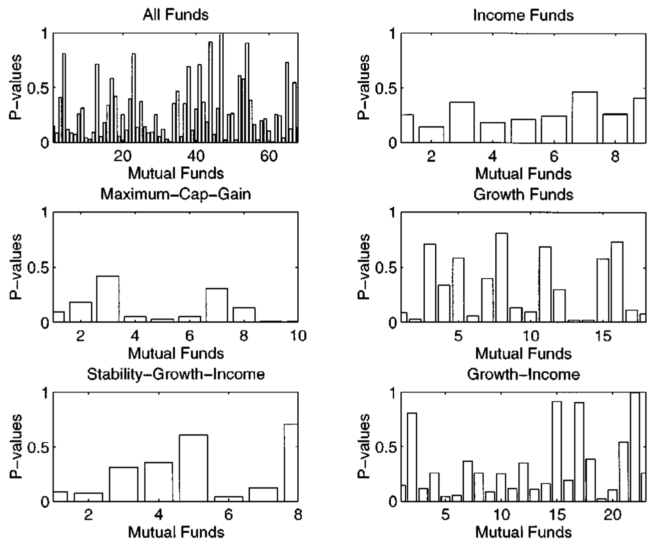

Figure 2
Unconditional NA-based performance measurement
The p-values are from the χ^2-test of the null hypothesis of no abnormal performance using the NA-based measure. Each p-value represents the probability that the corresponding variate is greater than the sample value of the test statistic. The test statistic is asymptotically χ^2-distributed with one degree of freedom.

Table 3
Summary performance statistics for individual funds

Performance measure	Max p	Min p	Percent of funds with $p < 5\%$	Number of funds with $\lambda > 0$ and $p < 5\%$
Unconditional LOP	0.97	0.003	8.8	3
Unconditional NA	0.96	0.003	13.2	2
Conditional LOP	0.98	0.004	2.9	1

Reported above are the statistics based on individual p-values and λ-values for all 68 funds, using the respective performance measures: the unconditional LOP-based, the unconditional NA-based, and the conditional LOP-based measures. That is, the individual estimates are obtained by running the estimation separately for each fund and for each measure. The sample period is from 1968:1 to 1989:12. See the descriptions in Tables 2 and 4. Here, p stands for p-value.

into five groups based on a reading of each fund's prospectus prior to the beginning of the sample period, the classification can serve as an indicator of the fund's risk level. The ordering of the fund types from the lowest to the highest risk level is income, stability-growth-income, growth-income, growth, and maximum capital gain. This means that the funds that lead to the strongest rejection of zero performance are members of the highest risk fund class: maximum capital gain funds.

In addition, zero performance for 6 of the 10 funds in this group is rejected at the 10 percent confidence level. Figure 2 also depicts the individual *p*-values for funds that comprise the lowest risk fund group: income. Note that at the 10 percent confidence level, zero performance is not rejected for any fund in this group. Therefore, it appears that abnormal performance is associated with the risk level of the fund. The riskier the fund, the more likely that the fund possesses abnormal performance.

Note that the statistics in Table 3 may not be independent across funds. To see the robustness of the results, we can conduct the Bonferroni test. According to the Bonferroni inequality, at a critical *p*-value of say 5 percent, the joint hypothesis of no abnormal performance among all 68 funds can be rejected if there is any fund with a *p*-value less than 0.073 percent (= $\frac{5}{68}$ percent). In Table 3, however, the minimum *p*-value for any fund and under any performance measure is 3 percent. Therefore, the joint hypothesis of no abnormal performance among all funds cannot be rejected according to the Bonferoni test.

The overall lack of evidence for the existence of positive performance is consistent with the findings of previous researchers.[13] Two recent studies that employ admissible performance measures are by Connor and Korajczyk (1991) and Lehmann and Modest (1987) (see Examples 1 and 2 in Section 1). The set of funds we examine is a subset of the funds examined by these researchers. However, their sample period is a subperiod of ours. The first two authors find, in contrast to our results, evidence for the existence of "widespread" abnormal performance using the APT-based measures. Consistent with our findings, they find that alphas from the APT-based performance measures are predominantly negative. However, Lehmann and Modest also carefully document the sensitivity of their results to, for example, the choice of the factor estimation procedure, the number of factors, and the number of securities used in estimating the factors. They find that the performance results based on the CAPM differ dramatically from those obtained from the APT-based measures. Connor and Korajczyk (1991) also use the CAPM and a five-factor version of the APT to investigate the performance of the same five classes of mutual funds, either by averaging the results of the individual funds within a group or by looking at an equally weighted portfolio of the funds within each group. Using the CAPM, they find significant negative performance for only the growth-income and the maximum capital gain funds. They point out that even though these performance values are statistically significant, the size of the APT-based performance values

[13] See, for example, Henriksson (1984), Jensen (1968, 1969), and Sharpe (1966).

Table 4
Conditional performance evaluation by fund groups

Fund type	λ	χ^2	*p*-value
Growth	−0.169	1.88	0.169
Growth-income	−0.032	0.811	0.368
Income	−0.117	1.11	0.292
Stability-growth-income	−0.129	3.04	0.081
Maximum capital gain	−0.187	1.57	0.210

The three information variables used in constructing the conditional stochastic discount factor are nominal 1-month Treasury bill rate, dividend yield on the CRSP value-weighted stock index, and the difference in yield to maturity between bonds with greater than 15 years to maturity and bonds with 5 to 15 years to maturity. The annualized monthly returns for all funds of a given type are equally weighted to form a return time series for the fund type. The sample time period is from 1968:01 to 1989:12. Conditional performance evaluation is then done on each of the five equally weighted return series, one for each fund type. The tests use the generalized method of moments and the Newey and West (1987) procedure to allow for heteroskedasticity and serial correlation. The lag length is set at 17 for all fund types. The tests are χ^2 tests of the overidentifying restriction with one degree of freedom. Expressed in basis points, λ is the performance value using the conditional LOP-based measure.

is within the range that would be explained by reasonable mutual fund transactions costs. In another study, Grinblatt and Titman (1988) also report no evidence of positive abnormal performance, based on returns for any of the seven classes of funds that they consider.[14]

To see how a conditional performance measure may rank funds differently, we use three information variables: the nominal 1-month Treasury bill, the dividend yield on the CRSP value-weighted NYSE stock index, and the term premium, which is the difference between the yield on bonds with more than 15 years to maturity and bonds with 5 to 15 years to maturity. For further discussion of these variables, see Appendix B.

Table 4 gives the conditional evaluation results for the five fund groups. As in the unconditional case, for four out of the five groups, there is insufficient sample information to reject the hypothesis of zero abnormal performance at all standard significance levels. In general,

[14] Grinblatt and Titman (1988) estimate survivorship bias by computing the difference between the traditional Jensen measure estimated from a sample of hypothetical returns which is *not* subject to survivorship bias and from a sample of hypothetical returns which *is* subject to survivorship bias [also see Brown et al. (1992)]. They conclude that the survivorship bias for a collection of 157 equity mutual funds is relatively small, 0.5 percent per year or less. Since the funds in our sample are a subset of the funds that survived the entire period, there is the possibility of survivorship bias in our test results. However, this is a positive bias in the sense that it may exaggerate positive performance values since only those managers that have survived are being evaluated. Given that we find little or no evidence of positive abnormal performance, and given the findings of Grinblatt and Titman (1988) on the magnitude of the bias, our conclusion is most likely robust to survivorship bias considerations.

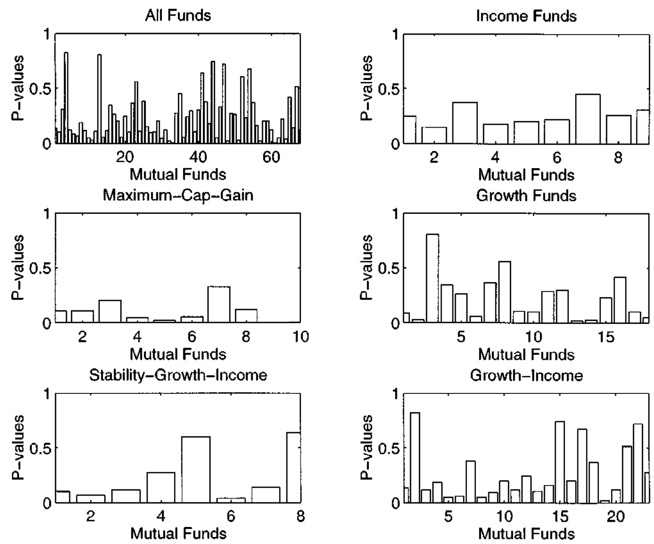

Figure 3
Conditional LOP-based performance measurement
The p-values are from the χ^2-test of the null hypothesis of no abnormal performance using the LOP-based measure. Each p-value represents the probability that the corresponding variate is greater than the sample value of the test statistic. The test statistic is asymptotically χ^2-distributed with one degree of freedom.

the p-values in the conditional case are larger than in the unconditional case. In addition, the λ's are negative for each group. Figure 3 displays the p-values for the conditional LOP-based measure in terms of individual funds by group. The p-values for most individual funds are respectively higher than in the unconditional case. The largest impact of conditional information is with regard to the growth funds, where these funds which previously had p-values below the 5 percent level now have p-values greater than 5 percent. About 2.9 percent of the funds have p-values less than 5 percent for the conditional LOP measure, compared to 13.2 percent of the funds for the unconditional case. There are two possible reasons for this increase in p-values for the individual funds. First, there are more parameters to estimate in the conditional case. Given the same number of observations, the statistical power of the estimates generally goes down, resulting in higher p-values. Alternatively, as discussed before, using public information-based dynamic portfolios as performance references leads to harder performance yardsticks. This makes it more likely for funds to show no abnormal performance. This point is also reflected in the λ values reported in Table 4. For instance, the performance value for the fund group stability-growth-income changes from 0.13 in the unconditional

LOP case to −0.129 when the conditional measure is used, while that for the growth funds changes from −0.06 to −0.169.

For the fund groups growth-income, income, and maximum capital gain, their performance values go up somewhat (while remaining negative as before) in the conditional case. This is consistent with the finding in Ferson and Schadt (1992) that performance switches from negative to positive when a conditional performance measure is used. See the discussion in Section 3 as to why this can occur.

There is only one fund that has a positive conditional performance value and a p-value less than 5 percent. For some funds the specified conditioning information has little or no effect on the test statistic. For example, funds five and nine in the maximum capital gain group have, respectively, p-values of 1 percent and 0.3 percent in the unconditional case and 1.1 percent and 0.4 percent in the conditional case. One interpretation of this is that the portfolio managers of these funds have generated abnormal returns by exploiting information that is not included in the three information variables. That is, the managers have shown an ability to use conditional information to guide investment decisions. An alternative interpretation is that the three selected variables do not reflect all the relevant information for the estimation of the stochastic discount factor d_p^*.

6. Concluding Remarks

In this article we have developed a general framework for evaluating the performance of a managed portfolio. In particular, we specify a minimum set of conditions that any performance measure must satisfy. Using this framework to assess existing performance measures and performance measurement in general leads to the following conclusions.

1. The first conclusion is negative in spirit. It says that if a managed fund has truly enlarged the investment opportunity set, then any performance value can be assigned by choosing the right performance measure that satisfies these minimal conditions. In addition, for any two such funds the ranking is performance measure-specific.

2. The second conclusion is positive in spirit. It says that if we also require each performance measure to be positive, performance measurement then becomes less dependent on the particular admissible performance measure used in the evaluation. When the performance measure used is positive, any managed portfolio whose excess return is positive with probability one will be classified as outperforming. Furthermore, if a fund is assigned a positive ranking, it must be the case that at least some investor would like to hold more of the managed fund. The converse of this is also true. Another point made in

this article of a positive nature is that given the negative theoretical results, it becomes even more important to evaluate performance using a battery of methods before drawing strong inferences.

3. The framework developed in this article provides a means of conducting performance evaluations independent of asset pricing models. Since performance evaluation by its very nature involves testing the economic and statistical significance of something that is relatively small (i.e., an excess return), a small amount of model misspecification may corrupt the entire inference. In this sense, identifying an admissible performance measure by estimating the equation system in Equation (8) is particularly useful because, unlike existing performance measures that build on equilibrium asset pricing models, such a measure does not rely on any asset pricing model and hence is not subject to any model misspecification. Furthermore, for performance evaluation purposes, such a measure is as admissible and qualified as any equilibrium-based measures.

4. Finally, our framework enables one to conduct conditional performance evaluation. When public information-generated returns are used as performance references, managers will need to truly utilize private information efficiently in order to receive a positive performance ranking.

Appendix A: Proof of Results

Proof of Theorems 1 and 2. From Chamberlain and Rothschild (1983), the LOP holds if and only if there is some $d \in L^2$ such that

$$E(d\,x_n) = 1 \qquad \forall n \in N. \tag{25}$$

See, alternatively, Chen and Knez (1994) and Hansen and Jagannathan (1991, 1994).

Suppose the LOP holds. Then, there is a d satisfying Equation (25) and, for any $x \in R_0$,

$$E(d\,x) = 1. \tag{26}$$

Now, define $\lambda(\cdot)\colon\ L^2 \to \Re$ by $\lambda(x) = E(d\,x)$ for every $x \in L^2$. Next, for any given $x' \in R_0$, $\lambda(x - x') = E[d\,(x - x')] = E(d\,x) - E(d\,x') = 1 - 1 = 0$, for each $x \in R_0$, which means λ satisfies Condition I. By the Riesz representation theorem,[15] the function λ so defined must

[15] Let H be a Hilbert space defined on the probability space (Ω, F, Pr) and equipped with the mean-square inner product, and $f(\cdot)\colon\ H \to \Re$. The Riesz representation theorem states that f is a continuous linear functional if and only if f can be uniquely represented by some $d \in H$ such that $f(x) = E(d\,x)$ for every $x \in H$.

also be linear and continuous, satisfying Conditions II and III. Equation (25) ensures that λ meets Condition IV. Thus, λ is an admissible performance measure.

Suppose that there is a function λ that is an admissible performance measure. By Condition II and III, λ must be linear and continuous on L^2, which implies, by the Riesz representation theorem, there is a unique $d \in L^2$ such that $\lambda(x) = E(d\,x)$ for each $x \in L^2$. Furthermore, for any given $x' \in R_0$, it holds by Condition I that $\lambda(x - x') = E[d\,(x - x')] = 0$, for every $x \in R_0$. Thus,

$$E(d\,x) = E(d\,x') = k, \tag{27}$$

for each $x \in R_0$ and some constant $k \in \Re$. By Condition IV, $k \neq 0$ (because $x_n \in R_0$). In summary, each admissible measure can be represented by a unique d satisfying Equations (1) and (2).

To close the loop of implications, suppose there is a $d \in L^2$ such that it both defines a function λ as in Equation (1) and satisfies Equation (2). By the Riesz representation theorem, this λ is continuous and linear. Since $k \neq 0$, dividing both sides of Equation (2) by k gives Equation (25), which, by the above stated result from Chamberlain and Rothschild (1983), means that the LOP holds. ■

Proof of Theorem 3. From Harrison and Kreps (1979) and Ross (1978), there is no arbitrage if and only if there is a $d^+ \in D$ such that $d^+ \gg 0$ [see also Hansen and Jagannathan (1991, 1994) and Hansen and Richard (1987)]. Therefore, suppose there is no arbitrage. Then, there is a $d^+ \in D$ such that $d^+ \gg 0$ and d^+ defines an admissible performance measure λ as in Equation (1). Clearly, this admissible measure λ satisfies Condition V.

Next, suppose there is an admissible positive performance measure λ. First, by Theorem 2, there must be a unique $d \in L^2$ such that d represents λ as in Equation (1) and d satisfies Equation (2). Second, since λ is positive on L^2, this d must also be so: $d \gg 0$. Third, rewrite Equation (2): $E(d\,x_n) = k$, $\forall n \in N$ and for some constant k. Since there is one x_n such that $x_n \gg 0$ and since $d \gg 0$, the above k must also be positive: $k > 0$. Then, dividing both sides of the above equation by the constant k and letting $d^+ \equiv \frac{d}{k}$, we have $E(d^+\,x_n) = 1$ for each $n \in N$, which means that $d^+ \gg 0$ and $d^+ \in D$. Finally, we have from Equation (1) that, for every $x_s \in L^2$ and some $x' \in R_0$, $\lambda(x_s - x') = E[d\,(x_s - x')] = k\,E[d^+\,(x_s - x')]$, which, letting $\eta = k$, gives Equation (9).

Suppose there is a $d^+ \in D$ such that $d^+ \gg 0$ and it satisfies Equation (9) for some constant $\eta > 0$. Then, clearly, there is not arbitrage. This completes the loop of implications. ■

Proof of Theorem 4. Since the set of all admissible measures is completely represented by the set $\overline{D}$, and since $\overline{D}$ is only a rescaled version of the set D (i.e., for each $\overline{d} \in \overline{D}$, there are a nonzero constant η and a $d \in D$ such that $\overline{d} = \eta\, d$), we only need to focus on those admissible measures represented by D.

Following the usual practice, assume the linear span M is a closed subspace in L^2. Let $M^{\perp}$ be the orthogonal complement of M in L^2. Recall from Chen and Knez (1995) that each $d \in D$ can be expressed as $d = d^* + \varepsilon$, for some $\varepsilon \in M^{\perp}$, where d^* is the unique member in D that is also in M. Conversely, $d^* + \varepsilon \in D$ for any $\varepsilon \in M^{\perp}$.

First, by Condition I, every admissible performance measure has to assign a zero performance to x_s if $x_s \in R_0$. This part is trivial. Next, suppose that every admissible measure λ assigns a zero performance to x_s: $\lambda(x_s - x') = 0$, for any given $x' \in R_0$. By Theorem 2, this means that

$$\eta\, E[d\,(x_s - x')] = \eta\, E[(d^* + \varepsilon)\,(x_s - x')] = 0, \qquad (28)$$

where $d \in D$ and $\eta\, d$ represents λ as in Equation (1). Note that the function λ^* defined by $\lambda^*(x) = E(d^*\, x)$ for each $x \in L^2$ gives an admissible measures, which implies by assumption that $E[d^*\,(x_s - x')] = 0$. This and Equation (28) together yield

$$E[\varepsilon\,(x_s - x')] = 0, \qquad \forall \varepsilon \in M^{\perp}, \qquad (29)$$

which can only hold when the excess return $(x_s - x')$ is orthogonal to every random variable in $M^{\perp}$. Thus, $(x_s - x')$ must be in M, which is possible only when x_s is in the uninformed set R_0 (because M is a closed subspace). In turn, x_s must be achievable by an uninformed investor.

To prove the last part, suppose that $x_s \notin R_0$. Then, $x_s \notin M$. Project x_s onto M (using the mean-square inner product) to yield, by the projection theorem, $x_s = x^* + \varepsilon^*$, where $x^* \in M$ is the projected point and $\varepsilon^* \in M^{\perp}$ is the component orthogonal to M. Note that $d^* + \eta\, \varepsilon^*$ is in D, for any constant η. Let λ' be the admissible performance measure defined by $d^* + \eta\, \varepsilon^*$. We have, for a given $x' \in R_0$ and for any value $v \in \Re$,

$$\begin{aligned}\lambda'(x_s - x') &= E[(d^* + \eta\, \varepsilon^*)\,(x^* + \varepsilon^* - x')] \\ &= E[d^*\,(x^* - x')] + \eta\, \|\varepsilon^*\|^2 = v\end{aligned}$$

by choosing $\eta = \frac{v - E[d^*\,(x^* - x')]}{\|\varepsilon^*\|^2}$, where $\|\varepsilon^*\|^2 \neq 0$ since $x_s \notin M$ by assumption. Thus, for any desired performance value for the fund, there is an admissible measure that gives it. ■

Proof of Theorem 5. Suppose that two managed gross returns, x_s and

$x_{s''}$, are such that $x_s \notin R_0$ and $x_{s''} \notin R_0$. Since the market is frictionless, this means that $x_s \notin M$ and $x_{s''} \notin M$. Assume that there is an admissible measure, λ, uniquely represented by some $d \in \overline{D}$, that ranks x_s higher than $x_{s''}$:

$$\lambda(x_s - x') - \lambda(x_{s''} - x') = \lambda(x_s - x_{s''}) = E[d\,(x_s - x_{s''})] > 0,$$

for a given reference return x'. Now, project $x_{s''}$ onto the closed linear span of x_s and M, denoted by $span(x_s, M)$, and obtain, by the projection theorem, $x_{s''} = x'' + z$, where $x'' \in span(x_s, M)$ and z is orthogonal to $span(x_s, M)$ (which also means $z \in M^{\perp}$), and $\|z\| \neq 0$ because the noises in the signals s and s'' are not perfectly correlated. Since $d + \mu\, z$ is in D for any constant μ, $d + \mu\, z$ defines another admissible measure λ_z:

$$\begin{aligned}\lambda_z(x_s - x_{s''}) &= E[(d + \mu\, z)\,(x_s - x_{s''})] \\ &= E[(d + \mu\, z)\,(x_s - x'' - z)] \\ &= E[d\,(x_s - x'' - z)] - \mu\,\|z\|^2,\end{aligned}$$

which can be made negative by the choice of a very large constant μ. Then, the admissible performance measure λ_z reverses the ranking of the two funds. ■

Proof of Theorem 6. A general proof can be constructed following the proof of Harrison and Kreps (1979, Theorem 1). For the purpose here, let's assume that each investor's preferences can be represented by von Neumann-Morgenstern utility functions $U(\cdot)$ that are strictly increasing and continuously differentiable. We only need to prove the first part of the theorem because the other part follows similarly. Without loss of generality, each uninformed investor starts with \$1.00. Take some investor with preferences $U(\cdot)$ and assume his optimal portfolio gross return is $x^* \in R_0$. By the first-order condition for the investor's portfolio problem, we have

$$E[x\, U'(x^*)] = 1 \qquad \forall x \in R_0, \tag{30}$$

where $U'(\cdot)$ is the first-order derivative of U.

First, suppose there is some NA-based measure λ^+ that assigns a positive performance to x_s. By Theorem 3, there are some $\eta > 0$ and $d^+ \in D^+$ such that

$$\lambda^+(x_s - x') = \eta\, E[d^+\,(x_s - x')] > 0, \tag{31}$$

for any $x' \in R_0$, which implies, for each $x \in R_0$,

$$E[d^+\,(x_s - x)] = E(d^+\, x_s) - E(d^+\, x) > 0. \tag{32}$$

Now, let $U'(x^*) = d^+ \gg 0$ for some investor with utility function $U(\cdot)$ and optimal portfolio return x^*. Then, the above equation yields: $E[U'(x^*)\, x_s] - E[U'(x^*)\, x] > 0$, which means that the investor's marginal valuation of x_s is higher than that of any portfolio in R_0. That is, this investor prefers x_s over every reference portfolio at the margin.

Next, suppose some investor with utility function $U(\cdot)$ and optimal portfolio x^* values (at the margin) x_s more than any reference portfolio in R_0. That is, for each $x \in R_0$, $E[U'(x^*)\, x_s] - E[U'(x^*)\, x] > 0$. Letting $d^+ = U'(x^*) \gg 0$, we then have $\lambda^+(x_s - x') \equiv E(d^+\, x_s) - E(d^+\, x) = E[d^+\,(x_s - x)] > 0$. This function λ^+ so defined clearly satisfies Condition I (because of the first-order condition for the investor's portfolio problem), Conditions II and III (because of the Riesz representation theorem), and Conditions IV and V (because of the fact that $d^+ \gg 0$). Thus, λ^+ is an NA-based measure that assigns a positive performance to x_s. ■

Appendix B: Data Description

Three data sets are used: the set of stock returns used to generate the reference payoffs, the set of information variables used to generate conditional benchmark payoffs, and the set of mutual funds.

Stock returns

Monthly returns for all individual stocks listed on the NYSE and the AMEX are used, beginning in January 1968 and ending in December 1990, as provided by CRSP. These return series are assigned to their respective industrial groups according to the two-digit SIC codes. For each industry, an equally weighted portfolio is constructed by taking the simple average of the existing returns for a given month, resulting in a total of 12 equally weighted industrial portfolios [see Ferson and Harvey (1991) and Ferson and Korajczyk (1995) for a similar construction].

Information variables

Three information variables are included and chosen based on evidence of their predictive power from existing research: (i) nominal 1-month Treasury bill rate, which is obtained from the CRSP riskfree files and shown to be a predictor of future stock returns by, among others, Fama and Schwert (1977), Ferson (1989), Ferson and Korajczyk (1995), and Keim and Stambaugh (1986); (ii) dividend yield of the CRSP value-weighted NYSE stock index, which is documented to possess predictive power by, for example, Campbell and Shiller (1988) and Fama and French (1988, 1989); and (iii) term premium as measured by the difference between yields on bonds with more than

15 years to maturity and bonds with 5 to 15 years to maturity [e.g., Ferson and Harvey (1991)], constructed from the Ibbotson corporate bond module (module 14).

Mutual fund data

The monthly returns net of transaction costs for the 68 funds are obtained from Standard and Poor's "Over-the-Counter Daily Stock Price Record," Weisenberger's "Investment Companies," and Moody's Annual Dividend Report for the period from January 1968 to December 1982, and from Morningstar Inc. for the remaining months until December 1991. For the first period, the data was graciously provided to us by Wayne Ferson. As both data sets have the names of each fund and there are 7 years of overlap, merging the data was straightforward. The Morningstar tape contains information on each fund's portfolio composition and investment objectives, which allows us to select only funds whose stated objective was to invest primarily in equities. This restriction on the funds selected is motivated by the fact that we use only equity securities to estimate the stochastic discount factors.

References

Admati, A., S. Bhattacharya, P. Pfleiderer, and S. Ross, 1986, "On Timing and Selectivity," *Journal of Finance*, 15, 715–730.

Admati, A., and S. Ross, 1985, "Measuring Investment Performance in a Rational Expectations Equilibrium Model," *Journal of Business*, 58, 1–26.

Andersen, T., and B. Sorensen, 1994, "GMM Estimation of a Stochastic Volatility Model: A Monte Carlo Study," working paper, Northwestern University.

Breen, W., L. Glosten, and R. Jagannathan, 1989, "Economic Significance of Predictable Variations in Stock Index Returns," *Journal of Finance*, 44, 1177–1190.

Brown, S., W. Goetzmann, R. Ibbotson, and S. Ross, 1992, "Survivorship Bias in Performance Studies," *Review of Financial Studies*, 5, 553–580.

Campbell, J., and R. Shiller, 1988, "The Dividend-Price Ratio and Expectations of Future Dividends and Discount Factors," *Review of Financial Studies*, 1, 195–228.

Chamberlain, G., and M. Rothschild, 1983, "Arbitrage, Factor Structure, and Mean-Variance Analysis on Large Asset Markets," *Econometrica*, 51, 1281–1304.

Chen, N., T. Copeland, and D. Mayers, 1987, "A Comparison of Single and Multifactor Portfolio Performance Methodologies," *Journal of Financial and Quantitative Analysis*, 22, 401–418.

Chen, Z., and P. Knez, 1994, "A Pricing Operator-Based Testing Foundation for a Class of Factor Pricing Models," *Mathematical Finance*, 4, 121–141.

Chen, Z., and P. Knez, 1995, "Measurement of Market Integration and Arbitrage," *Review of Financial Studies*, 8, 287–325.

Cochrane, J., and L. Hansen, 1992 "Asset Pricing Explorations for Macroeconomics," *National Bureau of Economic Research Macroeconomics Annual*, 1992, 115–165.

Connor, G., and R. Korajczyk, 1986, "Performance Measurement with the Arbitrage Pricing Theory: A New Framework for Analysis," *Journal of Financial Economics*, 15, 373–394.

Connor, G., and R. Korajczyk, 1988, "An Intertemporal Equilibrium Beta Pricing Model," *Review of Financial Studies*, 2, 33–52.

Connor, G., and R. Korajczyk, 1991, "The Attributes, Behavior, and Performance of U.S. Mutual Funds," *Review of Quantitative Finance and Accounting*, 1, 5–26.

Cumby, R., and S. Glen, 1990, "Evaluating the Performance of International Mutual Funds," *Journal of Finance*, 45, 497–521.

Dybvig, P., and J. Ingersoll, 1982, "Mean-Variance Theory in Complete Markets," *Journal of Business*, 55, 233–251.

Dybvig, P., and S. Ross, 1985, "The Analytics of Performance Measurement Using a Security Market Line," *Journal of Finance*, 40, 401–416.

Elton, E., M. Gruber, S. Das, and M. Hlavka, 1993, "Efficiency with Costly Information: A Reinterpretation of Evidence from Managed Portfolios," *Review of Financial Studies*, 6, 1–22.

Fama, E., and K. French, 1988, "Dividend Yields and Expected Stock Returns," *Journal of Financial Economics*, 22, 3–25.

Fama, E., and K. French, 1989, "Business Conditions and Expected Stock Returns," *Journal of Financial Economics*, 22, 3–26.

Fama, E., and W. Schwert, 1977, "Asset Returns and Inflation," *Journal of Financial Economics*, 5, 115–146.

Farnsworth, H., W. Ferson, D. Jackson, S. Todd, and B. Yomtov, 1995, "Conditional Performance Evaluation," working paper, University of Washington.

Ferson, W., 1989, "Changes in Expected Security Returns, Risk and the Level of Interest Rates," *Journal of Finance*, 44, 1191–1218.

Ferson, W., and C. Harvey, 1991, "The Variation of Economic Risk Premiums," *Journal of Political Economy*, 99, 385–415.

Ferson, W., and R. Korajczyk, 1995, "Do the Arbitrage Pricing Models Explain the Predictability of Stock Returns," *Journal of Business*, 68, 309–349.

Ferson, W., and R. Schadt, 1992, "Measuring Fund Strategy and Performance in Changing Economic Conditions," working paper, University of Washington.

Glosten, L., and R. Jagannathan, 1994, "A Contingent Claim Approach to Performance Evaluation," *Journal of Empirical Finance*, 1, 133–166.

Green, R., 1987, "Benchmark Portfolio Inefficiency and Deviations from the Security Market Line," *Journal of Finance*, 41, 295–312.

Grinblatt, M., and S. Titman, 1988, "The Evaluation of Mutual Fund Performance: An Analysis of Monthly Returns," working paper, University of California, Los Angeles.

Grinblatt, M., and S. Titman, 1989, "Portfolio Performance Evaluation: Old Issues and New Insights," *Review of Financial Studies*, 2, 393–421.

Hansen, L., 1982, "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, 50, 1029–1079.

Hansen, L., and R. Jagannathan, 1991, "Implications of Security Market Data for Models of Dynamic Economies," *Journal of Political Economy*, 99, 225–262.

Hansen, L., and R. Jagannathan, 1994, "Assessing Specification Errors in Stochastic Discount Factor Models," manuscript, University of Chicago.

Hansen, L., and S. Richard, 1987, "The Role of Conditioning Information in Deducing Testable Restrictions Implied by Dynamic Asset Pricing Models," *Econometrica*, 55, 587–613.

Hansen, L., and K. Singleton, 1982, "Generalized Instrumental Variables Estimation of Nonlinear Rational Expectations Models," *Econometrica*, 50, 1269–1286.

Harrison, J., and D. Kreps, 1979, "Martingales and Arbitrage in Multiperiod Securities Markets," *Journal of Economic Theory*, 20, 381–408.

He, J., L. Ng, and C. Zhang, 1994, "Asset Pricing Specification Errors and Mutual Fund Performance Evaluation," working paper, University of Houston.

Henriksson, R., 1984, "Market Timing and Mutual Fund Performance: An Empirical Investigation," *Journal of Business*, 57, 73–96.

Jagannathan, R., and R. Korajczyk, 1986, "Assessing the Market Timing Performance of Managed Portfolios," *Journal of Business*, 59, 217–235.

Jensen, M., 1968, "The Performance of Mutual Funds in the Period 1945–1964," *Journal of Finance*, 23, 389–416.

Jensen, M., 1969, "The Pricing of Capital Assets and the Evaluation of Investment Portfolios," *Journal of Business*, 42, 167–247.

Keim, D., and R. Stambaugh, 1986, "Predicting Returns in the Stock and Bond Markets," *Journal of Financial Economics*, 17, 357–390.

Knez, P., 1994 "Pricing Money Market Securities with Stochastic Discount Factors," working paper, University of Wisconsin, Madison.

Kreps, D., 1981, "Arbitrage and Equilibrium in Economies with Infinitely Many Commodities," *Journal of Mathematical Economics*, 8, 15–35.

Lehmann, B., and D. Modest, 1987, "Mutual Fund Performance Evaluation: A Comparison of Benchmarks and Benchmark Comparisons," *Journal of Finance*, 42, 233–265.

Mayers, D., and E. Rice, 1979, "Measuring Portfolio Performance and the Empirical Content of Asset Pricing Models," *Journal of Financial Economics*, 7, 3–29.

Merton, R., 1981, "On Market Timing and Investment Performance I: An Equilibrium Theory of Value for Market Forecasts," *Journal of Business*, 54, 363–406.

Newey, W., and K. West, 1987, "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55, 703–708.

Roll, R., 1978, "Ambiguity when Performance is Measured by the Security Market Line," *Journal of Finance*, 33, 1051–1069.

Ross, S. A., 1978, "A Simple Approach to the Valuation of Risky Streams," *Journal of Business*, 51, 453–475.

Sharpe, W., 1966, "Mutual Fund Performance," *Journal of Finance*, 39, 119–138.

Sharpe, W., 1992, "Asset Allocation: Management Style and Performance Measurement," *Journal of Portfolio Management*, Winter, 7–19.

Snow, K., 1991, "Diagnosing Asset Pricing Models Using the Distribution of Asset Returns," *Journal of Finance*, 46, 955–983.

Treynor, J., 1965, "How to Rate Management of Investment Funds," *Harvard Business Review*, 41, 63–75.