Portfolio Performance Measurement: A No Arbitrage Bounds Approach*

Dong-Hyun Ahn[a], H. Henry Cao[b], and Stéphane Chrétien[c,†]

[a] *Department of Economics, Seoul National University, San 56-1, Sillim-Dong, Kwanak-Gu, Seoul, 151-742, Korea*

[b] *Cheung Kong Graduate School of Business, Oriental Plaza 3/F, Tower E3, 1 East Chang An Avenue, Beijing, PR China, 100738*

[c] *Faculty of Business Administration, Pavillon Palasis-Prince, Laval University 2325, rue de la Terasse, Quebec City, QC, Canada, G1V 0A6*

November 2007

Abstract

This paper presents a new method to examine the performance evaluation of mutual funds in incomplete markets. Based on the *no arbitrage condition*, we develop bounds on admissible performance measures. We suggest new ways of ranking mutual funds and provide a diagnostic instrument for evaluating the admissibility of candidate performance measures. Using a monthly sample of 320 equity funds, we show that admissible performance values can vary widely, supporting the casual observation that investors disagree on the evaluation of mutual funds. In particular, we cannot rule out that more than 80% of the mutual funds are given positive values by some investors. Moreover, we empirically demonstrate that potential inference errors embedded in existing parametric performance measures can be of important magnitude.

JEL classification: G12; G23

Keywords: Portfolio Performance Measurement; Mutual Funds

---

*Special thanks to Geert Bekaert, Jacob Boudoukh, Jennifer Conrad, Bob Dittmar, Wayne Ferson, Don Fraser, Ron Gallant, Bob Korkie, Phil Sang Lee, Matthew Richardson, Anil Shivdansani, Steve Slezak, Mike Stutzer, Robert Whitelaw and seminar participants at the AFA and NFA Meetings, Cornell University, HEC-Montreal, Korea University, Laval University, Sherbrooke University, Simon Fraser University, the University of Alberta, the University of Calgary, the University of North Carolina, the University of Southern California, the University of Western Ontario, Yale University and York University for their helpful discussion and comments. Stéphane Chrétien acknowledges financial support from from the *Institut de Finance Mathématique de Montréal* and the Faculty of Business Administration at Laval University.

†Corresponding author. Phone: (418) 656-2131, ext. 3380. Fax: (418) 656-2624. Email: stephane.chretien@fsa.ulaval.ca.

Portfolio Performance Measurement: A No Arbitrage Bounds Approach

Abstract

This paper presents a new method to examine the performance evaluation of mutual funds in incomplete markets. Based solely on the *no arbitrage condition*, we develop bounds on admissible performance measures. We suggest new ways of ranking mutual funds and provide a diagnostic instrument for evaluating the admissibility of candidate performance measures. Using a monthly sample of 320 equity funds, we show that admissible performance values can vary widely, supporting the casual observation that investors disagree on the evaluation of mutual funds. In particular, we cannot rule out that more than 80% of the mutual funds are given positive values by some investors. Moreover, we empirically demonstrate that potential inference errors embedded in existing parametric performance measures can be of important magnitude.

JEL classification: G12; G23

Keywords: Portfolio Performance Measurement; Mutual Funds

1 Introduction

One of the most long-standing issues in financial economics is the measurement of mutual fund performance. Throughout the last few decades, as mutual funds have increasingly represented the dominant investment vehicle for individual investors, this issue has become of even higher profile. A difficult problem in the studies on performance measures is how to take into account the tradeoff between risk and returns which is based upon underlying asset pricing theories. Chen and Knez (1996) summarize the minimal conditions that an admissible performance measure should satisfy. The first, and perhaps the most substantial condition, is that the measure assigns zero performance to every *passive* portfolio that uninformed investors can construct. Consequently, the search for an admissible performance measure is consistent with a quest for an asset pricing model that can correctly value unmanaged portfolios. To put it differently, an admissible performance measure should be based on an admissible stochastic discount factor (SDF), the properties of which are extensively studied in the seminal work of Harrison and Kreps (1979).

Studies in asset pricing and, concomitantly, a search for performance measures have evolved primarily as two alternative approaches. The first approach derives a stochastic discount factor based on a full-fledged parametric asset pricing model. The early CAPM-based measures of Jensen (1968, 1969), Sharpe (1966), and Treynor (1965) belong to this approach. Most beta-pricing models and their corresponding performance measures fit this class as well.[1] Despite their contribution to the better understanding of the nature of performance evaluation, these studies are inevitably subject to one drawback, the *'bad model' problem*, as is mentioned in Fama (1998).[2] It is well-known that these measures fail to assign zero performance to passive, or reference, portfolios. As a result, performance measures developed in this line may not be admissible. This problem is accentuated by the empirical fact, emphasized by Lehmann and Modest (1987), that performance results may change significantly from one model to another.

The second approach does not rely upon a particular asset pricing model in defining an admissible performance measure. Instead, it estimates admissible performance measures from the available market data, i.e., passive portfolios. This "look into the data for performance measures" approach is pioneered in the period weighting measures of Grinblatt and Titman (1989), followed by numeraire portfolio measures

[1] See Connor and Korajczyk (1986), Lehmann and Modest (1987), Elton, Gruber, Das, and Hlavka (1993), Ferson and Schadt (1996) and Carhart (1997) among many others.

[2] For performance evaluation using the CAPM, see also Roll (1978), Dybvig and Ross (1985a, 1985b) and Green (1986).

advocated by Long (1990). This approach is developed into and culminates in the minimum-norm SDF-based measures of Chen and Knez (1996). The Chen-Knez measures are innovative since the performance measures are admissible, by construction, in the passive portfolios from which the minimum-norm SDF is derived. As such, these measurements do not suffer from the 'bad model' problem associated with the parametric approach. However, there is in general an infinite number of discount factors which assign different performance measures. This ambiguity in performance measures arises from the fact that markets may be incomplete, or at least are incomplete for econometricians with limited data. As shown by Harrison and Kreps (1979), in incomplete markets, there is an infinite set of admissible SDFs consistent with a subset of the economy, resulting in an infinite number of admissible measures. Therefore, one particular choice of SDFs, like the minimum-norm SDF, provides only one of the infinite performance measures and may not be admissible in a larger set of the market.

This paper extends the study of Chen and Knez (1996) by considering the infinite number of performance measures available in an incomplete market. We use only one minimal condition to restrict the set of admissible SDFs: the *no arbitrage* condition. Not only is this assumption economically appealing, but also it allows us to find the admissible range of, or bounds on, performance values of mutual funds. In doing so, we avoid any auxiliary assumptions inherent in existing studies (assumptions on preference systems or budget constraints in the parametric performance measures, or an implicit assumption in Chen and Knez (1996) to use the minimum-norm SDF). As such, our measures are free from the aforementioned potential 'bad model' problems innate in existing performance measures.[3]

Coupled with this analysis, we propose a new methodology to rank mutual funds. Depending on the relationship among the bounds on the performance measures, we develop three alternative ranking rules: Universal Dominance, Best Case Scenario Dominance and Worst Case Scenario Dominance. Consider two funds, A and B. Universal Dominance occurs when, using any identical admissible SDF, the performance measure of Fund A is always greater than the performance measure of Fund B. Best (Worst) Case Scenario Dominance occurs when the upper (lower) bound of Fund A is higher than the one of Fund B. One can interpret different admissible SDFs as the marginal utility of different classes of investors who invest a small amount in the managed fund. When Fund A universally dominates Fund B, then all investors will prefer

[3]Our approach is in the same spirit as Cochrane and Saá-Requejo (2000). They impose a maximum admissible Sharpe ratio condition to derive no "good deal" bounds on asset prices in incomplete markets.

Fund A to Fund B at the margin. For the Best (Worst) Case Scenario Dominance rule, the maximum (minimum) amount an investor is willing to pay at the margin, across the different investor classes, is higher for Fund A than for Fund B. We show that the Universal Dominance rule provides an incomplete ranking free from the 'bad model' problem, but the Best Case Scenario and the Worst Case Scenario ranking rules are subject to inference errors. We also present the benefits and shortcomings of each rule and illustrate their applications.

Another use of our performance bounds is to compare alternative parametric performance measures suggested in the existing literature. If candidate performance models are admissible, their performance values should be inside the bounds since they capture an *entire* set of admissible measures. Therefore, our bounds can be used as a diagnostic tool for evaluating candidate performance measures, in the same vein that the Hansen and Jagannathan (1991) bound is used to diagnose the validity of candidate SDFs of asset pricing models. This exercise can provide a justification for using one model over another in the context of performance evaluation, complementing studies by Kothari and Warner (2001), Farnsworth, Ferson, Jackson, and Todd (2002) and Coles, Daniel, and Nardari (2006) which look at simulations for such purpose.

As an empirical application of our approach, we use a monthly sample of 320 equity mutual funds during the period from 1984 to 1997. Out of 320 funds, the upper bounds of 55 funds are negative, while the lower bounds of only 8 funds are positive. Hence, the performance measures render sufficiently tight bounds to sign the performance of about 20% of the funds and the specific choice of SDF determines the performance sign of about 80% of the funds. We also investigate a rich menu of popular measures to determine their admissibility when using our bounds as a diagnostic tool. We document that consumption-based models often fail to generate admissible performance values, but that return-based models perform generally better. The best result is found in a conditional version of the Fama and French (1993) three-factor model that is inadmissible for only 8% of the funds. While the CAPM also generates mostly admissible performance values, the admissibility of the unconditional and conditional four-factor models proposed by Ferson and Schadt (1996) is somewhat poor since about 30% of their performance measures are outside the bounds.

We reach two main conclusions based on our analysis. First, it is often not possible to sign the performance of mutual funds. This finding not only suggest that inference errors can have a strong effect on the measurement of portfolio performance, but it can also potentially explain why the existing empirical literature documents an important sensitivity of performance to the benchmark chosen (see Lehmann and Modest (1987), Elton, Gruber, Das, and Hlavka (1993), Grinblatt and Titman (1994), Ferson and Schadt (1996), Carhart (1997) and Chan, Dimmock, and Lakonishok (2006)).

Second, we cannot rule out the possibility that a large number of mutual funds are evaluated positively by some investors. This result is comforting in light of the number of investors and the amount of money involved in the mutual fund industry. If markets are truly incomplete, then heterogeneous preferences implicit in the infinite admissible SDFs can potentially explain the disagreement in mutual fund valuation. In particular, given that mutual funds usually have a target investor class in mind, its mission is to serve its particular clientele. Presumably, the performance of the fund should be measured only with respect to that clientele's preferences. In this sense, the performance measure based on the upper bound and its associated Best Case Scenario Dominance rule can be thought as the performance assessment of the fund's most favorable investor class. Consequently, this measure can be regarded as a very relevant and practically meaningful way to evaluate the performance of mutual funds without knowing their particular clientele's specific preferences and without assuming a specific parametric asset pricing model.

The rest of the paper is organized as follows. Section 2 presents the basic framework in which we develop our performance measurement bounds and ranking rules. Section 3 offers guidelines on how to estimate the bounds and outlines the extension of our measures to allowing public information-conditioned portfolios as references. Section 4 discusses the use of our bounds as a diagnostic tool and presents candidate models that will be investigated. Section 5 describes the data used for our empirical applications. Section 6 presents the empirical results for a sample of mutual funds, and concluding remarks are offered in section 7.

2 Admissible Performance Measures

In this section, we first introduce a finite state economy in which we define and characterize the set of admissible SDFs. Then, we derive and interpret our main theoretical results, the performance evaluation bounds. We furthermore provide a simple example to illustrate our findings. Next, we present an extension of our analysis to a more general market economy. Finally, we discuss the performance ranking of mutual funds. Our setup is similar to Harrison and Kreps (1979), Hansen and Jagannathan (1991, 1997) and Chen and Knez (1995, 1996).

2.1 Finite State Economy

We consider a market economy represented by a probability space triplet $(\Omega,\ \mathcal{F},\ P)$ on which the space $\mathcal{L}^2$ of all random variables with finite second moment is defined. We endow $\mathcal{L}^2$ with its inner product $\langle x|y\rangle = E^P[xy]$ for x and $y \in \mathcal{L}^2$ to make the $\mathcal{L}^2$ space a Hilbert space. The corresponding second norm is $\|x\| = \langle x|x\rangle^{1/2}$.

The economy contains N basis assets, including $N-1$ risky assets and one riskless asset, with payoffs (in gross returns) denoted by a random $N \times 1$ random vector $\mathbf{x}$ and prices $\pi(\mathbf{x}) = 1_N$, where 1_N is an $N \times 1$ vector of ones. The set $\mathcal{A}$ of payoffs achievable by the investors of the economy includes all obtainable portfolios constructed with these payoffs. There are K states of the world with nonzero probability.

Assumption 1: *The number of states is strictly larger than the number of assets, $K > N$: i.e., the market is incomplete.*

Assumption 1 states that N basis assets and their portfolios are not sufficient for producing all possible state contingent payoffs . We suggest two justifications for this assumption. First, the whole market wherein the submarket $\mathcal{A}$ resides may in fact be incomplete, ruling out *perfect* risk sharing among investors. Put differently, the dimension of the assets in the whole economy itself may be small relative to the dimension of the states. Second, even though the market itself is truly complete, the N basis assets that econometricians rely on are only a subset of the whole market.

Before stating the second assumption, we require a formal definition of an arbitrage trading strategy.

Definition 1: *An arbitrage trading strategy is a trading strategy that gives an investor with zero endowment a nonnegative, nonzero payoff such that $(-\pi(\theta'\mathbf{x})\ \theta'\mathbf{x}) \geq^{+} (0\ 0)$, where the inequality $\geq^{+}$ is defined as: $x \geq^{+} y$ if $x(\omega) \geq y(\omega)\ \forall\ \omega \in \Omega$ and there exists at least one $\omega \in \Omega$ such that $x(\omega) > y(\omega)$.*

Definition 1 says that if payoff x is always as good as payoff y, and sometimes x is better, then the price of x must be greater that the price of y. Under this definition, a zero-investment trading strategy which earns a positive payoff in only one state and no payoffs otherwise will be counted as an arbitrage trading strategy.

The second assumption excludes the possibility of arbitrage trading strategies.

Assumption 2: *The price system in the submarket $\mathcal{A}$ is viable: i.e., the given price system is an equilibrium price system for some population of investors wherein arbitrage trading strategies are precluded.*

Assumption 2 means that the submarket of the N basis assets that econometricians choose should be *viable*, in the sense that a trading strategy based on portfolios of the basis assets should not lead to arbitrage profit opportunities. This assumption suggests an important criterion for the choice of basis assets: the basis assets should be *passive portfolios* which can be constructed by uninformed investors, and hence, unlikely to produce arbitrage trading opportunities. As a special case of this no arbitrage (NA) assumption, the so-called *law of one price* (LOP) must hold: two assets with the same payoffs must have the same price.

Under the above assumptions, we can define the SDF in the following proposition.

Proposition 1: *Under assumptions 1 and 2, there is a non-empty and non-singleton set of admissible stochastic discount factors, $\mathcal{M}$, which is closed and convex such that*

$$\mathcal{M} = \{M \mid \pi(y) = \theta' 1_N = E^P[M \cdot y] = \langle M|y \rangle \ \forall \ y = \theta' \mathbf{x} \in \mathcal{A} \text{ and } M \geq 0\},$$

where M is the SDF, a market-wide random variable.[4]

Proposition 1 states that there exists an infinite number of positive SDFs M which assign a unique price to a payoff $y \in \mathcal{A}$ such that $\pi(y) = \theta' 1_N$ where $y = \theta' \mathbf{x} \in \mathcal{A}$. Since the basis payoffs are gross returns, they have unit prices by construction. Hence, the price of the synthesized portfolio, y, equals the cost of the mimicking portfolio, $\theta' 1_N$. The infinite number of SDFs results from assumption 1 about market incompleteness. The linearity of the pricing functional and the positivity of the SDF result from assumption 2 and the definition of arbitrage trading strategies.

Proposition 1 is based on the *first valuation theorem* coupled with the *second valuation theorem* in the literature (see Duffie (1996)). It shows that, once the prices of the N basis assets chosen by the econometricians are viable, it is possible to find the SDFs defined on the physical probability measure P. Therefore,

[4] Proofs of propositions are made in Appendix A.

an econometrician should investigate the viability of price system in submarket $\mathcal{A}$ first before assessing the performance of mutual funds.[5]

2.2 Performance Evaluation Bounds

Since the active fund manager has the option to ignore his information and adopt a simple constant-composition portfolio, the rationale for investing in actively managed funds is to outperform passively managed portfolios, which are not built on superior information. These actively managed funds are, on average, more expensive to purchase; they not only charge various explicit costs such as front-end and/or back-end loads, 12b-1 expenses and brokerage fees, but also implicit costs such as higher transaction costs induced by higher turnover ratios ('smart money' related costs). Actively managed funds should provide superior returns to compensate investors for their superior costs. To do so consistently, the fund managers must possess superior information and must choose the best trading strategy to fully exploit that information.

We now investigate the performance measurement of mutual funds based on the admissible SDFs defined in proposition 1. As the performance measures rely on admissible SDFs, they assign zero performance to the passive portfolios achievable from the basis assets, and thus do not suffer from the 'bad model' problem. They are not subject to inference errors innate in existing performance measures that rely on auxiliary assumption on either specification for SDFs or market completeness.

Proposition 2: *Let x_{mf} be the gross return on a mutual fund. If $x_{mf} \in \mathcal{A}$, the performance value of the fund is a singleton:*

$$\alpha_{\$}(x_{mf}) = \psi(x_{mf}) - 1 = 0 \text{ where } \psi(x_{mf}) \stackrel{\triangle}{=} E^P[M \cdot x_{mf}] \text{ for any } M \in \mathcal{M}.$$

If $x_{mf} \notin \mathcal{A}$, there is a closed and compact interval of admissible performance values, $\mathcal{PM} = [\underline{\alpha}_{\$}(x_{mf}), \overline{\alpha}_{\$}(x_{mf})]$,

[5] We could proceed alternatively by finding the risk-neutral probability Q which corresponds to each SDF, and then using risk-neutral pricing under Q. Harrison and Kreps (1979) have shown that, under assumptions 1 and 2, there exists an equivalent martingale measure, Q, for each SDF in $\mathcal{M}$ such that

$$\mathcal{Q} = \left\{ Q | M = \frac{Q/P}{R_f} \;\; \forall \;\; M \in \mathcal{M} \right\}$$

where R_f is the gross riskless interest rate. Moreover, the fundamental valuation equation can be rewritten as:

$$\theta' 1_N = E^Q \left[\frac{y}{R_f} \right] \; \forall \, y = \theta' x \in \mathcal{A}.$$

Thus, we could state all our results on performance measurement and ranking in terms of risk-neutral pricing under Q. Given the one-to-one correspondence between the SDF and the equivalent martingale measure, we do not pursue this approach further.

such that the minimum admissible value and the maximum admissible value are respectively defined as

$$\begin{aligned}
\underline{\alpha}_{\$}(x_{mf}) &= \underline{\psi}(x_{mf}) - 1 \text{ where } \underline{\psi}(x_{mf}) \triangleq \inf_{M \in \mathcal{M}} E^P[M \cdot x_{mf}], \\
\overline{\alpha}_{\$}(x_{mf}) &= \overline{\psi}(x_{mf}) - 1 \text{ where } \overline{\psi}(x_{mf}) \triangleq \sup_{M \in \mathcal{M}} E^P[M \cdot x_{mf}].
\end{aligned}$$

Proposition 2 states that there are two cases when performance measurements are not subject to inference errors. The first case is when simple portfolios of the basis assets can precisely replicate the payoff of a mutual fund. In this case, the mutual fund's gross return is an element of the attainable set $\mathcal{A}$, and thus, regardless of the choice of SDF, it is given the same admissible price $\psi(x_{mf})$. Even though this perfect replication is an ideal case, it is also unlikely for two main reasons. First, if a portfolio manager truly possesses superior information, he should be able to generate payoffs that are not achievable by passive investors. Second, econometricians are able to use only a limited number of basis assets in evaluating mutual fund performance. This restriction reduces their ability to reproduce the payoffs of portfolio managers that generally invest in a large number of tradable assets.

The second case for performance measurements without inference errors is when simple portfolios of the basis assets do not span the mutual fund payoff. In this case, different SDFs, $M \in \mathcal{M}$, assign different admissible prices to the mutual fund's gross return. Thereby, an infinite number of performance measures are admissible. In the terminology of Harrison and Kreps (1979), there is an infinite number of valid pricing extensions ψ to the pricing functional π which assigns a unique price to a payoff $y = \theta'\mathbf{x} \in \mathcal{A}$ and a different price to other payoffs. However, since the set of admissible performance measures is *closed* and *convex*, proposition 2 states that there exist upper and lower bounds on the admissible performance values. Hence, it is possible to find the best and worst performance values for a mutual fund.

Apart from indicating the extreme admissible performance values, the upper and lower bounds have interesting interpretations as well when different SDFs are thought as intertemporal marginal rate of substitution of different investor classes. The upper bound represents the performance value of the investor class the most favorable to the mutual fund. Given that mutual funds generally serve a target investor class, its performance should ideally be measured with respect to that clientele's preferences. In this sense, the upper bound is relevant as it could represent the performance assessment the closest to the true value of a mutual fund without knowing its particular clientele's preferences. The lower bound gives the performance value of the investor class the least favorable to the mutual fund. While this value is less practically meaningful than the upper bound, the worst possible performance is in the same spirit as the Hansen and Jagannathan (1997) distance measure in which asset pricing models are assessed according to their worst pricing error.

The performance bounds of proposition 2 consider all admissible SDFs. Thus, any specific choice of admissible SDF will necessarily give a performance inside the bounds.

Corollary 2: *The NA performance measure of Chen and Knez (1996), $\alpha_\$^{MINNA}$, is inside the bound: $\alpha_\$^{MINNA} \in \mathcal{PM}$.*

The NA measure proposed in Chen and Knez (1996), $\alpha_\$^{MINNA}$, is based on a particular choice of M in $\mathcal{M}$, $M^{MINNA} = \inf_{M \in \mathcal{M}} ||M||$. Therefore, their measure is admissible, but it represents only one among the infinite number of admissible measures.[6] There is no reason why M^{MINNA} should be favored over other admissible SDFs. In fact, there is a strong likelihood that M^{MINNA} may not be admissible in an extended economy. When we enlarge the set of basis assets, the set of admissible SDFs, $\mathcal{M}$, tends to become smaller since there are more restrictions imposed on the *admissibility* of SDFs. One indication of this shrinking is that the Hansen-Jagannathan bound on the second norm of admissible SDFs shifts up when we increase the number of assets under analysis.[7] Therefore, given the limited number of basis assets included in any econometric analysis, the minimum second norm SDF M^{MINNA} estimated by econometricians may not be admissible in a larger economy.

Our performance bounds, $\mathcal{PM}$, are similar to "good deal" asset price bounds pioneered by Cochrane and Saá-Requejo (2000). They impose weak economic restrictions to derive bounds on asset prices in an incomplete market. Specifically, they obtain bounds which rule out high Sharpe ratios ("good deals"), as well as arbitrage trading opportunities, in pricing payoffs outside the attainable set. The "good deal" restriction reduces the set $\mathcal{M}$ of admissible SDFs by imposing an upper bound on the volatility of M such that

$$\sigma(M) \leq \frac{h}{R_f} \implies \overline{\sigma}(M) = \frac{h}{R_f},$$

where h is the maximum Sharpe ratio (defining "good deals") and $\overline{\sigma}(M)$ is the *exogenously* specified upper

[6] Notice that M^{MINNA} is not identical to M leading to $\underline{\alpha}_\$(x_{mf})$ because of the difference in underlying metrics. The former is based on the second norm whereas the latter is based on the inner product.

[7] See Bekaert and Urias (1996).

bound on the standard deviation of M.

The underlying motivation of Cochrane and Saá-Requejo (2000) is to eliminate auxiliary assumptions and rely upon only one condition, "good deals." In contrast, we even eliminate this condition, allowing us to focus on asset pricing with no auxiliary assumptions. The cost of relaxing this condition is not necessarily important. It is possible to determine the maximum Sharpe ratio available in the economy *endogenously* by finding

$$\sigma(M^{MAXNA}) = \sup_{M \in \mathcal{M}} \sigma(M).^{8}$$

The "good deal" restriction will be binding only if $\overline{\sigma}(M) < \sigma(M^{MAXNA})$, which yields $h < \sigma(M^{MAXNA})R_f$. Thus, the "good deal" restriction will not provide tighter bounds unless we specify a relatively low Sharpe ratio h.[9]

Finally, we can represent the performance measures in conventional return form:

$$\alpha_r = E^P[x_{mf}] - \frac{E^P[x_{mf}]}{\psi(x_{mf})} = \alpha_{\$} \frac{E^P[x_{mf}]}{1+\alpha_{\$}}.$$

The performance measure α_r is now comparable to a performance measure obtained from the intercept in a linear regression of excess mutual fund returns on the market prices of risk associated with a linear factor model (i.e. a Jensen's Alpha).

2.3 Example

We explore an example, using a simple finite state economy, to illustrate the formal analysis derived above. Assume an economy with four states, $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\}$ and with physical probability measure $P = \begin{pmatrix} 0.3 & 0.2 & 0.2 & 0.3 \end{pmatrix}'$. Suppose that an econometrician has three basis assets with payoffs represented by the following matrix:

$$\mathbf{x} = \begin{pmatrix} 1.4 & 0.9 & 1.05 \\ 1.2 & 1.2 & 1.05 \\ 0.8 & 0.7 & 1.05 \\ 0.9 & 1.3 & 1.05 \end{pmatrix}.$$

This economy is incomplete (as $K = 4 > N = 3$) and has a risk-free asset with an interest rate of 5%. In this economy, the set of admissible SDFs are the positive solutions for $M = \begin{pmatrix} M_1 & M_2 & M_3 & M_4 \end{pmatrix}'$ such

[8] To show the theoretical existence of M^{MAXNA}, notice that $\sigma(M) = \|M\|$ is a continuous function mapping $\mathcal{M}$ (the closed convex set of positive stochastic discount factors that price correctly the basis assets) to $\mathcal{R}$. Thus, a maximum exists for $\sigma(M)$.

[9] Bernardo and Ledoit (2000) also propose asset price bounds, using restriction on the gain-loss ratio of assets. They show that their restriction is equivalent to a Sharpe ratio restriction if returns are normally distributed, while it might be superior to a Sharpe ratio restriction in ruling out attractive investments (near-arbitrage opportunities) when returns are not normal. The discussion of our approach compared to the bounds in Cochrane and Saá-Requejo (2000) could be paralleled for the bounds in Bernardo and Ledoit (2000).

that

$$\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 1.4 & 0.9 & 1.05 \\ 1.2 & 1.2 & 1.05 \\ 0.8 & 0.7 & 1.05 \\ 0.9 & 1.3 & 1.05 \end{pmatrix}' \begin{pmatrix} 0.3 & & M_1 \\ 0.2 & \cdot & M_2 \\ 0.2 & & M_3 \\ 0.3 & & M_4 \end{pmatrix}.$$

To solve this system of three equations with four unknowns, we set M_1 to an arbitrary value λ and find the unique solution for M given λ. Then, we find the restrictions on λ that ensure that M is positive in all states. Simple algebra shows that

$$M = \begin{pmatrix} \lambda & 2.8822 - 2.6842\lambda & 1.5038 - 0.5526\lambda & 0.2506 + 1.1579\lambda \end{pmatrix}' \text{ where } 0 < \lambda < 1.0738,$$

and the range of admissible values for the SDFs is

$$\begin{array}{ccccc} 0 & < & M_1 & < & 1.0738 \\ 0 & < & M_2 & < & 2.8822 \\ 0.9103 & < & M_3 & < & 1.5038 \\ 0.2506 & < & M_4 & < & 1.4940. \end{array}$$

Interpreting the SDFs as the marginal utilities of different classes of investors, our solutions indicate that the second state is where investors are the most heterogenous: some investors have a marginal utility close to zero, giving no value to a payoff in state two, while others have a high marginal utility, giving a value of 2.88 to a unit payoff in that state (and its corresponding Arrow-Debreu security). Following the same logic, the third state is where the investors appear the most homogenous.[10]

We examine the performance of three mutual funds in this incomplete economy. First consider a mutual fund with payoffs $x_{mf1} = \begin{pmatrix} 1.085 & 1.215 & 0.705 & 1.165 \end{pmatrix}'$. The payoffs of this mutual fund are such that $x_{mf1} = \mathbf{x}\theta$, where $\theta = \begin{pmatrix} 0.4 & 0.7 & -0.1 \end{pmatrix}'$. Since $x_{mf1} \in \mathcal{A}$, all admissible SDFs assign a unique performance measure, $\alpha_{\$}(x_{mf1}) = \psi(x_{mf1}) - 1 = 1 - 1 = 0$. Put differently, as the investors are able to replicate its payoffs with the available basis assets, they give zero performance to this mutual fund.

The other two funds we consider are more realistic in the sense that their payoffs are not achievable from the basis assets. Hence, different investors assign different performance to these funds. To find out the lowest and highest admissible performance measures for any fund with $x_{mf} \notin \mathcal{A}$, we solve for λ in the following

[10]For completeness, as $M = \frac{Q/P}{R_f}$, the solution to the equivalent martingale measure is

$$Q = \begin{pmatrix} 0.3150\lambda & 0.6053 - 0.5637\lambda & 0.3158 - 0.1160\lambda & 0.0789 + 0.3647\lambda \end{pmatrix}'.$$

problems

$$
\begin{aligned}
\underline{\psi}(x_{mf}) &= \inf_{0<\lambda<1.0738} x'_{mf} \begin{pmatrix} 0.3 & & \lambda \\ 0.2 & \cdot & 2.8822 - 2.6842\lambda \\ 0.2 & & 1.5038 - 0.5526\lambda \\ 0.3 & & 0.2506 + 1.1579\lambda \end{pmatrix} \\
\overline{\psi}(x_{mf}) &= \sup_{0<\lambda<1.0738} x'_{mf} \begin{pmatrix} 0.3 & & \lambda \\ 0.2 & \cdot & 2.8822 - 2.6842\lambda \\ 0.2 & & 1.5038 - 0.5526\lambda \\ 0.3 & & 0.2506 + 1.1579\lambda \end{pmatrix}
\end{aligned}
$$

As these problems are linear in λ, we obtain corner solutions that involve λ being set at its minimum or its maximum. When $\lambda = 0$, the admissible SDF reflects the preferences of investors with marginal utilities of a payoff at their lowest values in the first and fourth states and at their highest values in the second and third states. When $\lambda = 1.0738$, we obtain an admissible SDF implying opposite preferences.

Now consider a mutual fund with payoffs $x_{mf2} = (\ 0.7 \quad 1.5 \quad 2.1 \quad 0.3\ )' \notin \mathcal{A}$. It is easy to show that $0.7423 < \psi(x_{mf2}) < 1.5188$ and $-0.2577 < \alpha_{\$}(x_{mf2}) = \psi(x_{mf2}) - 1 < 0.5188$. Some investors assign a negative performance to this mutual fund, while others assign a positive performance. Furthermore, there is a supporting equilibrium which designates $\alpha_{\$}(x_{mf2}) = 0$. Hence, we cannot reject the hypothesis that the performance of this fund is equivalent to the market after an adjustment of risk. Finally consider a mutual fund with payoffs $x_{mf3} = (\ 1.3 \quad 1.3 \quad 0.6 \quad 0.5\ )' \notin \mathcal{A}$. For this fund, $0.7521 < \psi(x_{mf3}) < 0.9674$ and consequently $-0.2479 < \alpha_{\$}(x_{mf3}) = \psi(x_{mf3}) - 1 < -0.0326$. All admissible SDFs assign a negative performance measure to this fund, indicating that this fund is not valuable for all classes of investors. So we can conclude that this mutual fund is underperforming.

2.4 Extension to the General Market Economy

Here we extend the model to the case in which the dimension of the state space can be infinite. As before, we assume that the market is incomplete and there is no arbitrage. In order to obtain finite performance measure bounds, we impose an additional smoothness condition on the SDF:

Assumption 4: *Let ω_1, ω_2 denote two arbitrary states. The SDF satisfies a smoothness condition: there exist a constant B such that*

$$|M(\omega_1) - M(\omega_2)| \leq B \sqrt{\sum_{j=1}^{N} (x_j(\omega_1) - x_j(\omega_2))^2}.$$

The smoothness condition implies that the SDF is relatively smooth and the distance of the SDF across two states is bounded by a constant times the distance in the realized payoffs of the N basis assets. With the addition of assumption 3, we obtain performance bounds in the general market economy.

Proposition 3: *Let x_{mf} be the gross return on a mutual fund. If $x_{mf} \in \mathcal{A}$, the performance value of the fund is a singleton:*

$$\alpha_{\$}(x_{mf}) = \psi(x_{mf}) - 1 = 0 \text{ where } \psi(x_{mf}) \stackrel{\triangle}{=} E^P[M \cdot x_{mf}] \text{ for any } M \in \mathcal{M}.$$

If $x_{mf} \notin \mathcal{A}$, there is a closed and convex set of admissible performance values, $\mathcal{PM} = [\underline{\alpha}_{\$}(x_{mf}), \overline{\alpha}_{\$}(x_{mf})]$, such that the minimum admissible value and the maximum admissible value are respectively defined as

$$\begin{aligned} \underline{\alpha}_{\$}(x_{mf}) &= \underline{\psi}(x_{mf}) - 1 \text{ where } \underline{\psi}(x_{mf}) \stackrel{\triangle}{=} \inf_{M \in \mathcal{M}} E^P[M \cdot x_{mf}], \\ \overline{\alpha}_{\$}(x_{mf}) &= \overline{\psi}(x_{mf}) - 1 \text{ where } \overline{\psi}(x_{mf}) \stackrel{\triangle}{=} \sup_{M \in \mathcal{M}} E^P[M \cdot x_{mf}]. \end{aligned}$$

Moreover, the bounds are finite.

2.5 Performance Ranking of Mutual Funds

An important application of measuring portfolio performance is to rank mutual funds. Magazines and newspapers regularly report the ranking of funds (such as the "top ten" fund managers) based on their past performance. Since there is, arguably, some evidence of persistence in fund performance[11], the ranking of mutual funds might as well have crucial importance for investors' decision making. Roll (1978), Dybvig and Ross (1985a), Green (1986), Lehmann and Modest (1987) and Chen and Knez (1996) show that the ranking of mutual funds can change significantly from one model to another. Thus, ranking is highly sensitive to the 'bad model' problem. In this section, we examine the ranking of mutual funds based on all admissible performance measures, thus avoiding inference errors due to the 'bad model' problem. We develop the following three alternative ranking rules: Universal Dominance, the Best Case Scenario Dominance, and finally the Worst Case Scenario Dominance.

[11] See Grinblatt and Titman (1992), Hendricks, Patel, and Zeckhauser (1993), Elton, Gruber, Das, and Hlavka (1993), Goetzmann and Ibbotson (1994), Brown and Goetzmann (1995), Elton, Gruber, Das, and Blake (1996), Carhart (1997) and Christopherson, Ferson, and Glassman (1998).

Definition 2: *Consider two mutual funds: Fund A and Fund B. The corresponding performance bounds are $\mathcal{PM}_A = [\underline{\alpha}_\$(x_A), \overline{\alpha}_\$(x_A)]$ and $\mathcal{PM}_B = [\underline{\alpha}_\$(x_B), \overline{\alpha}_\$(x_B)]$ respectively.*

- **Universal Dominance:** *Fund A dominates Fund B in the sense of* Universal Dominance*, denoted by $A \overset{UD}{>} B$, if the lower bound on the differential in performance measures of A and B evaluated with the same SDF is positive: i.e.,*

$$\inf_{M \in \mathcal{M}} E^P[M(x_A - x_B)] \quad > \quad 0.$$

The necessary condition for this type of dominance is $\overline{\alpha}_\$(x_A) \geq \overline{\alpha}_\$(x_B)$ and $\underline{\alpha}_\$(x_A) \geq \underline{\alpha}_\$(x_B)$.

- **Best Case Scenario Dominance:** *Fund A dominates Fund B in the sense of* Best Case Scenario Dominance*, denoted by $A \overset{BCSD}{>} B$, if the upper bound on the performance measure of A is greater than the upper bound on the performance measure of B: i.e.,*

$$\overline{\alpha}_\$(x_A) > \overline{\alpha}_\$(x_B).$$

- **Worst Case Scenario Dominance:** *Fund A dominates Fund B in the sense of* Worst Case Scenario Dominance*, denoted by $A \overset{WCSD}{>} B$, if the lower bound on the performance measure of A is greater than the lower bound on the performance measure of B: i.e.,*

$$\underline{\alpha}_\$(x_A) > \underline{\alpha}_\$(x_B).$$

Figure 1 illustrates the three dominance rules. Universal Dominance is more complicated. One needs to determine the lowest price among the set of admissible SDFs for a strategy which consists of buying Fund A and simultaneously short selling Fund B. If that lowest price is positive, then $A \overset{UD}{>} B$. We can rewrite the expression for Universal Dominance as

$$\begin{aligned}
\inf_{M \in \mathcal{M}} E^P[M(x_A - x_B)] &= \inf_{M \in \mathcal{M}} \left[E^P[M \cdot x_A] - 1\right] - \left[E^P[M \cdot x_B] - 1\right] \\
&= \inf_{M \in \mathcal{M}} \left[\alpha_\$(x_A) - \alpha_\$(x_B)\right].
\end{aligned}$$

The last equality suggests another way to interpret the Universal Dominance rule. The rule looks at the difference between the performance values of Fund A and Fund B under the same admissible SDF. If this difference is always positive, then Fund A is always preferred to Fund B for any given admissible SDF. The key point is that Universal Dominance narrows the analysis to evaluating two funds with *identical* admissible SDFs.[12] This means that all investors would prefer Fund A to Fund B.

The second rule, the Best Case Scenario Dominance, is based on the notion that when we examine two funds at their upper bounds, Fund A has a higher performance value than Fund B. In this case, among all types of investors, the most one is willing to pay for Fund A is higher than that for Fund B. The Best Case Scenario Dominance hence give a ranking on how the funds add value to its most favorable clientele. The third rule, the Worst Case Scenario Dominance, is based on the notion that when we examine two funds at their lower bounds, Fund A has a higher performance value than Fund B. The Worse Case Scenario rule is similar to the Hansen and Jagannathan (1997) ranking of the performance of asset pricing models according to the worst pricing error they generate in a set of portfolios. A computational strength of the second and the third rules is that they can be established directly from the estimation of $\mathcal{PM}_A$ and $\mathcal{PM}_B$.

In summary, the full-fledged inference error-free ranking is the one based on Universal Dominance. As with the bounds for performance evaluation, the Universal Dominance rule leads to performance ranking bounds as opposed to a precise ranking. When a precise ranking is preferred however, the Best Case Scenario Dominance and the Worst Case Scenario Dominance rules can provide convenient alternative schemes since they are easy to implement and still based on admissible performance measures.

3 Performance Bounds: Estimation and Conditional Version

In the previous section, we provide a theoretical analysis on the admissible performance bounds of mutual funds. In this section, we examine how to estimate the performance bounds and show how to obtain conditional version of the bounds.

3.1 Estimation Problems and Solution Technique

Let $\mathbf{x}$ be the N-dimensional random vector of basis asset payoffs. Then, we can rewrite the bounds as the following problems:

[12]Notice that a sufficient (but not necessary) condition for $A \overset{UD}{>} B$ is that $\underline{\alpha}_{\$}(x_A) > \overline{\alpha}_{\$}(x_B)$. Thus, it is possible to establish Universal Dominance directly from the performance bounds when the lower bound of a fund is greater than the upper bound of another one.

Problem 1-1: The lower bound problem

$$\underline{\psi}(x_{mf}) = \min_{M \in \mathcal{M}} E^P(M \cdot x_{mf}) \;\; s.t. \;\; \mathbf{1} = E^P(M \cdot \mathbf{x}); \; M \geq 0.$$

Problem 2-1: The upper bound problem

$$\overline{\psi}(x_{mf}) = \max_{M \in \mathcal{M}} E^P(M \cdot x_{mf}) \;\; s.t. \;\; \mathbf{1} = E^P(M \cdot \mathbf{x}); \; M \geq 0.$$

Problems 1-1 and 2-1 state that the goal is to solve for the minimum or maximum price of the mutual fund payoffs, subject to the constraints that the stochastic discount factor M is positive and prices correctly the basis asset payoffs.

To examine an empirically interesting sample, we assume that $N < T$ (so that the sample represents an incomplete market), and that the observed payoffs of the N basis assets are linearly independent (so that the payoffs are not redundant). Let $\mathbf{M} = \begin{pmatrix} M_1 & \cdots & M_K \end{pmatrix}'$. Let D_{tk} denote the dummy variable such that it has value 1 if the realized state at time t is k and zero otherwise and $\mathbf{D}$ denote the $K \times T$ matrix composed of D_{tk}. Let $\mathbf{x_{mf}} = \begin{pmatrix} x_{mf1} & \cdots & x_{mfT} \end{pmatrix}'$, and $\mathbf{x} = \begin{pmatrix} \mathbf{x}_1 & \cdots & \mathbf{x}_T \end{pmatrix}'$. Then, we can rewrite the problems as follow:

Problem 1-2: The lower bound problem

$$\underline{\psi}^*(x_{mf}) = \min_{\{\mathbf{M}\}} \frac{1}{T}\mathbf{M}'\mathbf{D}\mathbf{x_{mf}} \;\; s.t. \;\; \mathbf{1} = \frac{1}{T}(\mathbf{M}'\mathbf{D}\mathbf{x})'; \; \mathbf{M} \geq 0.$$

Problem 2-2: The upper bound problem

$$\overline{\psi}^*(x_{mf}) = \max_{\{\mathbf{M}\}} \frac{1}{T}\mathbf{M}'\mathbf{D}\mathbf{x_{mf}} \;\; s.t. \;\; \mathbf{1} = \frac{1}{T}(\mathbf{M}'\mathbf{D}\mathbf{x})'; \; \mathbf{M} \geq 0.$$

Problems 1-2 and 2-2 are linear programming optimization problems with equality constraints and boundaries. Although they cannot be solved analytically, they can easily be solved numerically. Introducing Lagrange multipliers, the problems can be written as:

Problem 1-3: The lower bound problem

$$\underline{\psi}^*(x_{mf}) = \min_{\{\mathbf{M}\}} \max_{\{\lambda,\delta\}} (1/T)\mathbf{M}'\mathbf{D}\mathbf{x_{mf}} - \lambda'[(1/T)(\mathbf{M}'\mathbf{D}\mathbf{x})' - \mathbf{1}] - (1/T)\delta'\mathbf{M},$$

with the complementary slackness conditions for all k

$$\begin{aligned}\delta_k &\geq 0 \text{ for } M_k = 0,\\ \delta_k &= 0 \text{ for } M_k > 0.\end{aligned}$$

Problem 2-3: The upper bound problem

$$\overline{\psi}^*(x_{mf}) = \max_{\{\mathbf{M}\}} \max_{\{\lambda,\delta\}} (1/T)\mathbf{M}'\mathbf{D}\mathbf{x}_{\mathbf{mf}} + \lambda'[(1/T)(\mathbf{M}'\mathbf{D}\mathbf{x})' - \mathbf{1}] + (1/T)\delta'\mathbf{M},$$

with the complementary slackness conditions for all t

$$\begin{aligned}\delta_k &\geq 0 \text{ for } M_k = 0,\\ \delta_k &= 0 \text{ for } M_k > 0.\end{aligned}$$

Taking the first derivative of problems 1-3 and 2-3, the first order conditions for an optimum are:[13]

- $(1/T)(\mathbf{M}'\mathbf{D}\mathbf{x})' = \mathbf{1}$, $\mathbf{M} \geq 0$;
- $\mathbf{D}\mathbf{x}_{\mathbf{mf}} = \mathbf{D}\mathbf{x}\lambda + \delta$ for the lower bound; $\mathbf{D}\mathbf{x}_{\mathbf{mf}} = -\mathbf{D}\mathbf{x}\lambda - \delta$ for the upper bound;
- $\delta_k > 0$ when $M_k = 0$.

The optimization technique implemented to solve the standard form linear programming problems 1-3 and 2-3 is referred to as the *simplex method*. We briefly describe the procedure used in this paper, a variant of the simplex method known as the two-phase revised simplex method, in Appendix B.[14] Intuitively, the solution to the upper (lower) bound problem gives high values to SDFs corresponding to high (low) mutual fund payoffs and low values to SDFs corresponding to low (high) mutual fund payoffs, while ensuring that the constraints on correct basis asset pricing and SDF positivity are met. Thus, as expected, an investor who values a fund at its upper (lower) bound as high marginal utility in states where the fund returns are high (low) and low marginal utility in states where the fund returns are low (high).

3.2 Asymptotics and Consistency

The above estimation technique provides point estimates of the performance evaluation bounds. By themselves, the point estimates represent useful information as they examine extreme performance possibilities

[13]See Gill, Murray, and Wright (1981, Section 3.3.2) for optimality conditions and sketches of proofs.
[14]Readers interested in more details and references can consult Gill, Murray, and Wright (1981).

of mutual funds in an economy with no arbitrage opportunities. In this subsection, we briefly provide the asymptotic property and consistency of the bounds. Given the potential presence of sampling error, it is also desirable to obtain the statistical significance of the performance bounds. In our empirical implementation, we will obtain the empirical small sample distributions of the performance bound estimates from Monte Carlo simulations, and use these distributions for hypothesis testing.

Let P_k denote the probability that state k occurs and P denote the vector of probabilities, $P \equiv (P_1, ..., P_K)'$. Let $\bar{x}_k$ denote the conditional expected payoff of x in state k and $\bar{x} = (\bar{x}_1, ..., \bar{x}_K)'$ the vector of conditional payoffs. Let $\underline{b}(P, \bar{x})$ denote the solution of the following problem:

$$\underline{b}(P, \bar{x}) = \min_{M} \max_{\lambda,\delta} \sum_k P_k M_k \bar{x}_k - \lambda' \sum_k [P_k \bar{x}_k M_k - 1] - \sum_k P_k \delta_k M_k = 0$$

where

$$\delta_k M_k = 0,$$

for all states k. Since the maximand is a continuous function, $\underline{b}$ is also a continuous function.

Let $N_k(T)$ denote the number of times that state k occurred in a sample of size T, and s_t the state at time t. Denote $\bar{x}_{mf,k}(T) = \sum_{t,s_t=k} \frac{x_{mft}}{N_k(T)}$ the sample conditional mean and $P_k(T) = N_k(T)/T$ the sample frequency of state k. Denote $\bar{x}_{mf}(T) = (\bar{x}_{mf,1}(T), ..., \bar{x}_{mf,K}(T))', P(T) = (P_1(T), ..., P_K(T))'$. We have

$$\underline{\psi}^*(x_{mf}) = \underline{b}(P(T), \bar{x}_{mf}(T))$$

As T goes to infinity, by the law of large numbers, $P_k(T) \to P_k, \bar{x}_{mf,k}(T) \to E_k^P[x_{mf}|k]$. Finally, assuming that $\underline{b}$ is differentiable at $(P, \bar{x}_{mf})$, the convergence rate is of the order of $\frac{1}{T}$ as $P(T)\bar{x}_{mf}(T)$ converges at the rate of $\frac{1}{T}$ by the central limit theorem. Asymptotic property and consistency of the upper bound can be obtained similarly.

The previous analysis considers only the case with finite states. When the dimension of the states is infinite, we can examine a series of economies with finite states that converge to the infinite state economy. In particular, we can consider the following discretization. Let $I_{(n,j,L)} = [\frac{j}{2^L}, \frac{j+1}{2^L}]$, $n = 1, ..., N$, $j = 0, ..., 4^N$. Let $k(j_1, ...j_N, L) \equiv \{\omega, x_{j_n}(\omega) \in I_{(n,j,L)}\}$. We have partitioned the states in two finite partitions. For each L, we can consider the set of SDFs that give the same value on states in the same partition. Denote the bounds for a particular mutual fund in such a discretized economy as $\overline{\psi}^*(x_{mf}, L)$ and $\underline{\psi}^*(x_{mf}, L)]$. Then, since the set of SDFs is increasing in L, the upper bound $\overline{\psi}^*(x_{mf}, L)$ must be an increasing function in L and the lower bound $\underline{\psi}^*(x_{mf}, L)$ must be a decreasing function in L. In addition, the upper bound is bounded from above while the lower bound is bounded from below. By the monotone convergence theorem, as L goes to infinity, the economy converges to the infinite state economy and the bounds converges to a finite number.

3.3 Including Conditioning Information

It is possible to implement the above methodology under both unconditional and conditional asset pricing frameworks. Unconditional models arise either from one-period static models or from explicit (albeit dynamic) discount factor models with constant parameters over time. As Chen and Knez (1996) and Ferson and Schadt (1996) note, unconditional models presume a simple buy-and-hold trading strategy. If expected returns and risk premia, however, change over time, performance evaluation should incorporate dynamic trading strategies as well. Otherwise, unconditional performance measures may simply capture gains or losses of dynamic trading strategies. Given this concern, we examine conditional performance measures following Chen and Knez (1996), Ferson and Schadt (1996) and Dahlquist and Söderlind (1999).

The extension of the unconditional approach to a conditional framework is straightforward. First, we extend our probability space triplet to $(\Omega, \mathcal{F}, F, P)$, where $F = \{F_t\}_{0 \le t < \mathcal{T}}$, a filtration. Then, the bounds on admissible performance measures, $\mathcal{PM} = [\overline{\alpha}_{\$}(x_{mf}), \underline{\alpha}_{\$}(x_{mf})]$ will be determined by the extended set of basis assets:

$$x^c = x \otimes z,$$

where $z \stackrel{\triangle}{=} \tilde{z}/E[\tilde{z}]$ is a standardized predetermined instrumental variables $\in F_t$. Thus, following the convention in Ferson (1989), the payoffs based on dynamic trading make use of publicly available information, $\tilde{z}$. Information variables are normalized to z to make the cost of dynamic trading strategy a unit dollar. As discussed by Cochrane (1996), we can interpret $x \otimes z$ as dynamically managed portfolios which are based on information variables. The inclusion of conditioning information enlarges the set of basis assets, and its corresponding achievable set, $\mathcal{A}$, which tightens the bounds on performance measures. This desirable feature occurs since the SDFs are enforced to satisfy more restrictions, namely to assign zero performance measures to dynamically managed portfolios.

4 Diagnosis of Performance Evaluation Models

The literature on performance evaluation proposes a large number of parametric performance measures. These measures provide a single point estimate, which results in a precise performance evaluation. However, they suffer from three problems discussed earlier. First, they require strong economic assumptions to be admissible under their null. Second, they empirically give non-zero performance to passive portfolios. Third, they result in performance evaluation that can change significantly from one measure to another. How these problems affect the performance evaluation exercise? Which parametric measures suffer the least from these problems? Answers to these questions requires the comparison of different parametric performance measures. Such a comparison is difficult because the mutual funds' true performance measures are not known. Recent studies by Kothari and Warner (2001), Farnsworth, Ferson, Jackson, and Todd (2002) and Coles, Daniel, and Nardari (2006) overcome this unobservability by using artificial mutual funds.

Our performance bounds provide an alternative way to compare parametric performance measures. The bounds represent the entire set of admissible performance values. Therefore, if a parametric performance measure is admissible, its performance value must reside inside the bounds. Figure 2 illustrates this idea by showing an inadmissible candidate performance measure. Since our bounds are based on a particular choice of basis assets, residing inside the bounds is not a sufficient condition, but a necessary condition that parametric performance measures should meet. In that sense, our bounds, in the context of performance measurement, play the role of a diagnostic tool similar to the role played by the Hansen and Jagannathan (1991) variance bound in the context of asset pricing models.

In the empirical section, we use our bounds as a diagnostic tool to investigate a rich menu of alternative performance measures considered in existing studies. This section presents a brief overview of the theories and estimation techniques used to obtain these candidate performance measures. We classify them into three categories: linear factor models, consumption-based models and nonparametric models.

4.1 Linear Factor Models

Arguably, the most widely used models for the assessment of portfolio performance are linear factor models. These models can either be seen as versions of the intertemporal asset pricing theory of Merton (1973) or the arbitrage pricing theory of Ross (1976). The SDF implied in these models is a linear function of the state variables:

$$M^f = \omega_0 + \omega_1' f,$$

where f is a vector of factors or state variables. Let λ^f be a vector of the corresponding market prices of systematic risk or expected risk premia. Then, the performance measure based on a linear factor model can be expressed as

$$\alpha_r^f(x_{mf}) = E^P[r_{mf} - r_f] - \beta' \lambda^f,$$

where r denotes a simple return (a gross return x minus one), and β is the vector of factor loadings or sensitivities. We estimate the linear factor model performance measure using a regression analysis, assuming the following statistical model:

$$r_{mf,t} - r_{f,t} = \alpha_r^f(x_{mf}) - \beta' \lambda_t^f + \epsilon_t,$$

with $E^P[\epsilon_t] = E^P[\epsilon_t \lambda_t^f] = 0$.

Ferson and Schadt (1996) also propose an extension to include the information contained in some pre-specified variables $z = [\ z^1 \ \cdots \ z^Z \ ]'$. We can express the resulting conditional performance measure as

$$\alpha_r^{f\text{-}C}(x_{mf}) = E^P[r_{mf} - r_f | z] - \beta(z)' \lambda^f(z).$$

To estimate the conditional linear factor model performance measure, we assume that the conditional betas are affine functions of the information variables: $\beta(z) = b_0 + b_1 z^1 + \cdots + b_Z z^Z$. Then, we use the following linear regression model:

$$r_{mf,t} - r_{f,t} = \alpha_r^{f\text{-}C}(x_{mf}) + b_0' \lambda_t^f + (b_1 z_{t-1}^1)' \lambda_t^f + \cdots + (b_Z z_{t-1}^Z)' \lambda_t^f + \varepsilon_t,$$

with $E^P[\varepsilon_t | z_{t-1}] = E^P[\varepsilon_t \lambda_t^f | z_{t-1}] = 0$.

We implement the conditional and unconditional version of the following three linear factor models with implied market prices of risk such that:

- *The CAPM:* $\lambda_t^{CAPM} = r_t^m - r_{f,t}$

  where r^m is the return on the market portfolio.

- *The Fama-French Model:* $\lambda_t^{FFM} = [\ r_t^m - r_{f,t} \quad r_t^{smb} \quad r_t^{hml} \ ]'$

  where r^{smb} and r^{hml} are returns on the mimicking portfolios of size and book-to-market respectively.

- *The Ferson-Schadt Model:* $\lambda_t^{FSM} = [\ r_t^{ls} - r_{f,t} \quad r_t^{ss} - r_{f,t} \quad r_t^{ltgb} - r_{f,t} \quad r_t^{lgcb} - r_{f,t}\ ]'$

  where r^{ls}, r^{ss}, r^{ltgb} and r^{lgcb} are the returns on large stocks, small stocks, long-term government bonds and low-grade corporate bonds.

4.2 Consumption-Based Models

The second category of models we consider is consumption-based models. While linear factor models are popular because of their relatively good pricing performance, they are often criticized because the factor selection is usually not guided by theory, which raises the issue of data snooping (see Lo and MacKinlay (1990)). The consumption-based models do not suffer from this drawback. The SDF implied by these models is a function of the marginal utility of the representative agent:

$$M_{t+1}^C = \beta \frac{u'(C_{t+1};\theta)}{u'(C_t;\theta)} \frac{PL_t}{PL_{t+1}},$$

where $u'(C_t;\theta)$ is the marginal utility of the representative agent as a function of his consumption C_t at time t and a vector of parameters θ, and PL_t is the price level at time t. The performance measure implied by the consumption-based models is given by

$$\alpha_{\$}^C(x_{mf}) = E^P[M^C \cdot x_{mf}] - 1.$$

We estimate the performance measure of the consumption-based model by following a two-step approach. In the first step, using the generalized method of moment procedure of Hansen (1982), we estimate the parameters θ by minimizing a quadratic form of the average pricing error on the benchmark assets:

$$\hat{\theta} = \arg\min_{\{\theta\}} g(\theta)' W(\theta) g(\theta),$$

where $g(\theta) = (1/T)\sum_{t=1}^T M_t^C(\theta)\mathbf{x}_t - \mathbf{1}$ and $W(\theta)$ is the inverse of a consistent estimate of the variance-covariance matrix of $g(\theta)$. In the second step, we compute $\hat{\alpha}_{\$}^C(x_{mf})$ as its sample counterpart:

$$\hat{\alpha}_{\$}^C(x_{mf}) = (1/T)\sum_{t=1}^T M_t^C(\hat{\theta}) x_t^{mf} - 1.$$

We implement two consumption-based models. The first model assumes that the representative agent has time-separable power utility. The resulting SDF is

$$M_{t+1}^{POWER} = \beta \left(\frac{C_{t+1}}{C_t}\right)^{-\gamma} \frac{PL_t}{PL_{t+1}}.$$

The second model is an external habit-formation preference specification similar to the one used in Campbell and Cochrane (1999). The habit is assumed exogenous to the agent's decision and linear in lagged aggregate consumption. The implied SDF is

$$M_{t+1}^{HABIT} = \beta \left(\frac{C_{t+1} - \theta C_t}{C_t - \theta C_{t-1}} \right)^{-\gamma} \frac{PL_t}{PL_{t+1}}.$$

4.3 Nonparametric Models

The linear factor models and the consumption-based models are parametric models. No restriction in parametric models guarantees the correct pricing of the basis assets. Thus, these models are usually not admissible empirically. To avoid this problem, Chen and Knez (1996) advocate the use of a nonparametric approach.

Their first performance measure assumes that the law of one price (LOP) holds. The performance measure implied by this assumption, which we denote by MINLOP, is given by

$$\alpha_{\$}^{MINLOP}(x_{mf}) = E^P[M^{MINLOP} \cdot x_{mf}] - 1,$$

where M^{MINLOP} solves the following problem:

$$M^{MINLOP} = \arg\min_{\{M\}} \sigma(M) \;\; s.t. \;\; \mathbf{1} = E^P[M \cdot \mathbf{x}],$$

with $\sigma(M) = E^P[(M - E^P[M])^2]$. This problem offers the following analytical solution:

$$\alpha_{\$}^{MINLOP}(x_{mf}) = E^P[\mathbf{x}'(E^P[\mathbf{x}\mathbf{x}'])^{-1}\mathbf{1} \cdot x_{mf}] - 1,$$

which we estimate by using the sample counterpart to the solution.

As noted by Chen and Knez (1996), a problem with the MINLOP measure is that the model can admit arbitrage trading strategies since the SDF is allowed to be negative. To eliminate this problem, Chen and Knez (1996) propose a no arbitrage (NA) measure (denoted by MINNA), which is defined as

$$\alpha_{\$}^{MINNA}(x_{mf}) = E^P[M^{MINNA} \cdot x_{mf}] - 1,$$

where M^{MINNA} solves the following problem:

$$M^{MINNA} = \arg\min_{\{M\}} \sigma(M) \;\; s.t. \;\; \mathbf{1} = E^P[M \cdot \mathbf{x}]; \; M > 0.$$

We estimate the measure numerically by using the sample counterpart to the moments in the problem.

As discussed previously, M^{MINNA} is only one of the infinite number of admissible SDFs defined in proposition 1. To examine the particularity of this specific choice of nonparametric measure, we use an alternative NA measure (denoted by MAXNA). This measure uses the SDF which has the maximum standard deviation among all admissible SDFs. It is defined as:

$$\alpha_{\$}^{MAXNA}(x_{mf}) = E^P[M^{MAXNA} \cdot x_{mf}] - 1$$

where M^{MAXNA} solves the following problem:

$$M^{MAXNA} = \arg\max_{\{M\}} \sigma(M) \;\; s.t. \;\; \mathbf{1} = E^P[M \cdot \mathbf{x}]; \; M > 0.$$

Again, this model is estimated numerically using its sample counterpart.[15]

Notice that α^{MINNA} and α^{MAXNA} are two specific choices of performance measures among the infinite set we consider implicitly in the performance bounds. Although both performance measures will always be between the performance measurement bounds, their performance evaluation can differ widely for any specific mutual funds.

5 Data

We now turn to an empirical implementation of the performance measurement bounds to illustrate their applicability. For this implementation, we use our number of observations T as the number of states K in the economy.[16] This section describes the datasets of mutual funds, basis assets and variables for the parametric models.

5.1 Mutual Fund Sample

The sample consists of 320 open-end mutual funds that invest primarily in the U.S. equity market.[17] For each fund, we obtain monthly returns from January 1984 to December 1997, a total of 168 observations. The returns include reinvestment of all distributions and are net of management fees, incentive fees, and other fund expenses, but disregard load charges and exit fees. Morningstar, Inc is the data source. Even though

[15] To find the solution to the problem $\max_{\{M\}} \sigma(M)$, notice that it is a quadratic programming problem with linear constraints and boundaries. Such a problem can be solved using one of the many quadratic optimization techniques available.

[16] Although not necessary, choosing $K = T$ is typical in most empirical studies. Implicitly, it assumes that the data generating process is stationary and ergodic, thus allowing state-averaging to be replaced by time-averaging.

[17] Our sample contains one mutual fund specialized in real estate investments.

the 320 funds exist for the duration of our sample, survivorship bias is not a considerable issue since most of our analysis is concerned with the performance of individual funds.[18]

Table 1 presents summary statistics on the gross monthly returns of the mutual funds. Panel A looks at the cross-sectional distribution of the 320 mutual funds, while panel B examine the funds grouped by their Morningstar investment objectives. In annualized numbers, the mutual funds have a mean return of 13.97%, with a mean standard deviation of 55.36%. The average returns range from -9.17% to 21.62% while the standard deviations range from 21.66% to 153.13%. The average Sharpe ratio is 0.165, with minimum and maximum values of -0.205 and 0.315, respectively. The portfolios of funds grouped by investment objectives are obtained from equally-weighted portfolios using the 320 funds in our sample, and are thus subject to survivorship bias. The mutual funds specializing in financials provide the best average return and Sharpe ratio, while the mutual funds specializing in precious metals show the worst average return and Sharpe ratio. Overall, our mutual fund sample offers widely distributed return characteristics.

5.2 Basis Asset Payoffs

We select basis assets that reflect the returns and risks available to investors and fund managers. Table 2 presents summary statistics of the variables used to construct the basis assets. To represent the stock market, we choose 20 industry portfolios following Moskowitz and Grinblatt (1999). The industry portfolios are formed monthly using CRSP returns and SIC codes, which allow for time-variation in industrial classification. The annualized average returns on the industry portfolios vary from 7.84% for Apparel to 17.68% for Chemical. King (1966) shows that industry groupings are important in capturing the common variation in stock returns. Furthermore, industry portfolios are relevant in examining funds that specialize in a specific sector, like the specialty funds included in our sample. Chen and Knez (1996) and Dahlquist and Söderlind (1999), for example, also use industry portfolios in their analysis of the performance of mutual funds.

We also select two bond portfolios formed from assets of different maturity. The short-term bond portfolio contains bonds with maturity less than one year and represents the returns available in the money market. The long-term bond portfolio includes bonds with maturity greater than ten years and represents the returns available from the fixed income market. We obtain the bond portfolio returns from the CRSP Fama Maturity

[18] Brown, Goetzmann, Ibbotson, and Ross (1992), Brown and Goetzmann (1995), Malkiel (1995) and Elton, Gruber, and Blake (1996)) discuss the upward bias created when measuring the performance of portfolios of surviving funds. In cases where survivorship bias might be an issue, we will use estimates provided by Elton, Gruber, and Blake (1996) to examine its effect on our results.

Portfolios Returns File. The annualized average return is 6.77% on the short-term bond portfolio and 12.19% on the long-term bond portfolio. Wermers (2000, table I, panel C) shows that the average percentages of non-stock holdings from equity mutual funds are between 15% and 20% during our sample period. The bond portfolios represent the opportunities offered by those holdings.

We also use two predetermined information variables to take into account the information available to investors and fund managers. The first variable is a lagged credit spread, measured as the difference between the lagged yields on Baa corporate bonds and long-term Treasury bonds. The yields are from CITIBASE. The second variable is the lagged monthly dividend yield on the CRSP value-weighted index. The credit spread and dividend yield have annualized average values of 1.64% and 3.10% respectively. Fama and French (1989) and Ferson and Harvey (1991), among others, have argued that these information variables are correlated with time-variation in expected returns. Using the two information variables, we form 44 managed portfolios from the previous 22 portfolio returns, giving a total of 66 basis assets.[19]

5.3 Variables for the Parametric Models

To implement the linear factor models, we need proxies for the market prices of risk λ_t. For the CAPM, we use the return on the CRSP value-weighted index as the market portfolio return. For the FFM, we obtain the three factors described in Fama and French (1993). For the FSM, we replicate the proxies presented in Ferson and Schadt (1996). We use the return on CRSP S&P 500 index as return on large stocks. The returns on a small cap index and on a long-term (approximatively 20-year) U.S. government bond, taken from Ibbotson Associates, represent the returns on small stocks and long-term government bond, respectively. The return on low-grade corporate bonds is from the series presented in Blume, Keim, and Patel (1991) until December 1989, and from the Merrill Lynch High Yield Composite Index (obtained from Datastream) thereafter. Finally, the return on the one-month Treasury bill is used as the risk-free return.

To implement the conditional version of the linear factor models, we choose the five instruments adopted by Ferson and Schadt (1996). Specifically, we use the lagged level of the one-month Treasury bill yield, the

[19]Given the large number of potentially relevant portfolio returns and information variables, along with the relatively limited guidance offered by theory, we construct our set of basis assets with two objectives in mind. First, the resulting price system must be viable, i.e. the basis assets must allow the set of admissible SDFs to exist. Second, the set of admissible SDFs must be small, in the sense that the minimum and maximum SDF standard deviations must be close to each other. From Hansen and Jagannathan (1991) and Cochrane and Saá-Requejo (2000), this second objective can be interpreted as selecting a price system where the maximum Sharpe ratio allowed is close to the maximum Sharpe ratio achievable from the basis assets. A set of SDFs with such characteristic should generate bounds within an economically interesting range.

lagged dividend yield of the CRSP value-weighted index, a lagged measure of the slope of the term structure, a lagged quality spread in the corporate bond market, and a dummy variable for the month of January. The level of the Treasury bill yield is the 30-day annualized Treasury bill yield from the CRSP RISKFREE file. The dividend yield is the price level at the end of the previous month on the CRSP value-weighted index, divided in the previous twelve months of dividend payments for the index. The term spread is the ten-year Treasury bond yield minus the three-month Treasury bill yield. The quality spread is Moody's BAA- rated corporate bond yield less the AAA-rated corporate bond yield. The bond yields are from CITIBASE.

To implement the consumption-based asset pricing models, we use the seasonally-adjusted personal consumption expenditures on non-durable and service, their respective consumption deflator and the resident population to construct a proxy of aggregate per capita consumption. Finally, we use the CPI (not seasonally-adjusted) for the price level. CITIBASE is the data source.

6 Empirical Results

This section presents our empirical results. We first examine the admissible SDFs in our sample. Then, we present the performance measurement bounds, along with their empirical small sample distributions. Finally, we consider three applications of the bounds: performance evaluation, performance ranking and diagnostics of parametric evaluation models.

6.1 Set of Admissible Stochastic Discount Factors

Our results on portfolio performance measurement are based on a set of SDFs that correctly price the basis assets and preclude arbitrage trading strategies. As a first step, we now describe some characteristics of the infinite number of admissible SDFs. As discussed earlier, this step corresponds to the important task of assessing the viability of the price system under consideration. Figure 3 illustrates, in the mean-standard deviation space advocated by Hansen and Jagannathan (1991), the sets of admissible SDFs with and without conditioning information, assuming different values for the mean of the SDFs. Admissible SDFs under the law of one price (LOP) and under no arbitrage (NA) are both provided.

The graph shows that the inclusion of conditioning information results in a notable reduction (in mean-standard deviation space) of the set of SDFs. We make four observations on the effect of conditioning information. First, as discussed by Gallant, Hansen, and Tauchen (1990) and Bekaert and Liu (2004) among others, an increase in the lower standard deviation bounds (LOP and NA) implies that the information variables are helpful in predicting returns. Second, as seen by the distances between the LOP and NA lower standard deviation bounds, the NA condition is more restrictive when including conditioning information. Third, the decrease in the upper standard deviation bound caused by including conditioning information is very pronounced. Finally, the bounds on the lowest and highest means of the SDFs are much tighter when conditioning information is included.

To examine portfolio performance, we further restrict the set of admissible models by fixing the mean of the SDFs. Specifically, we include an additional basis asset that has a constant return equal to the average one-month T-bill return over our sample period.[20] The annualized return on this asset is 5.485%, which implies that the mean of the SDFs is equal to 0.99545. A dotted line in figure 3 indicates the new set resulting from this restriction. This restriction not only provides tighter bounds, but it also ensures that the mean of the SDFs is tied to a reasonable value. Dahlquist and Söderlind (1999) and Farnsworth, Ferson, Jackson, and Todd (2002) discuss the importance of identifying the mean of the SDFs. Considering the no arbitrage condition, the conditioning information and their fixed mean, the admissible SDFs have a minimum standard deviation of 1.514 and a maximum standard deviation of 2.389.

6.2 Performance Measurement Bounds

Table 3 summarizes the performance measurement bounds, presented in monthly abnormal return form. Panel A presents statistics on the cross-sectional distribution of the results for the 320 mutual funds. Figure 4 illustrates the distribution for the lower and upper bounds by presenting an histogram of the annualized α_r. The lower and upper bounds have a monthly mean of -0.578% and 0.316%, respectively. Although more than 90% of the values of the bound are within 0.5% of their mean, some mutual funds have very extreme performance measures; the minimum and maximum values are respectively -2.852% and 0.243% for the lower bounds, and -1.181% and 5.447% for the upper bounds. The bound differences examine the tightness of the bounds. The mean of the bound differences is 0.895%, indicating an economically important divergence of values on mutual fund performance. The minimum bound difference is 0.311%, indicating that no mutual fund has payoffs spanned by the basis assets. The correlation coefficient between the bounds (not reported) is 0.35 (p-value < 0.0001). The t-statistics confirm the impression offered by figure 4 that the bounds are

[20] Thus, we are estimating the unconditional mean of the SDFs as the inverse of the average one-month T-bill gross return, which is considered as a good proxy for the conditional riskfree rate.

significantly different from zero and significantly different from each other.[21] Finally, the bound averages give an indication of the location of the performance values. Under the assumption that the performance values are distributed symmetrically between the bounds, more than 75% of the mutual funds are assigned a negative performance by a least 50% of their admissible performance measures.

To gain more insight on the lower and upper bounds, figure 5 shows the bounds for the 320 mutual funds sorted in increasing order of their mean return (fig. 5a), their standard deviation of returns (fig. 5b) and their Sharpe ratio (fig. 5c). Figure 5a reveals that, except for some funds with the lowest average return, the upper and (especially) the lower bounds are generally increasing with average returns. Thus, a fund with higher average return is generally given higher performance bounds. However, the less than perfect relation indicates that there is a considerable 'risk adjustment' implicit in our measures. Figure 5b shows that the bounds widen as the standard deviation of returns increased. As expected, the basis assets have more difficulty replicating the payoffs of mutual funds with larger variation. Finally, figure 5c shows that Sharpe ratios are generally increasing with the lower bounds, but have little relation with the upper bounds. Our bounds and the Sharpe ratios provide related, but different 'risk adjustments'.

While the point estimates of the bounds are informative by themselves, it is also important to examine the effect of sampling error on the estimates. We conduct Monte Carlo simulations to obtain an empirical small sample distribution for the lower and upper bounds of each mutual funds and investment objective portfolios.[22] The main result from the simulations is that the precision of the bound estimates is strongly inversely related to the variability of fund returns. Figure 6 illustrates this finding by presenting the distribution for the mutual funds located at the 10th, 25th, 50th, 75th and 90th percentile of the 320 funds sorted by their standard deviation of returns. As can be seen, the bounds for the 10th percentile fund (fig. 6a) are considerably more precisely estimated than the bounds for the 90th percentile fund (fig. 6e). This relation is highly significant; the correlation between the standard deviations of returns and the standard deviations of the bound estimates is 0.89 (p-value < 0.0001) for the lower bound and 0.90 (p-value < 0.0001) for the upper

[21] The t-statistics for the lower and upper bounds are computed by assuming that the cross-sectional distribution of the bounds is multivariate normal with a mean of zero, a standard deviation as reported in Table 3, and a correlation between any two lower or any two upper bounds of 0.68, which corresponds to the average correlation between the returns of the investment objective portfolios. For the bound difference and bound average t-statistics, we also assume a correlation between the lower and upper bounds of 0.35.

[22] For each mutual fund, 168 observations of its returns and of the variables needed to construct the basis assets (20 industry portfolio returns, two bond portfolio returns and two lagged information variables) were generated assuming a multivariate normal distribution. The bounds were then computed from the simulated data. The process was repeated 5000 times providing the empirical distribution of the estimates.

bound. Thus, similar to more familiar performance measures, the bounds are more reliable when measuring the performance of mutual funds with smaller return variation. In the next section, we will use the small sample distributions to examine the statistical significance of the bounds.

6.3 Performance Evaluation

6.3.1 Individual Mutual Funds and Investment Objective Portfolios

For how many funds can we assign a positive or negative performance without incorrect inference? As discussed previously, a fund has a positive (negative) performance without inference error if there is no admissible SDF that gives a negative (positive) performance to the fund. A closer look at the cross-sectional distribution of the bounds described in panel A of table 3 indicates that 55 individual mutual funds have a negative upper bound, while only 8 funds have a positive lower bound. So, out of 320 funds, 17.2% of the funds have a negative performance, while 2.5% of the funds have a positive performance, and the performance sign of 80.3% of the funds cannot be determined. Thus, 80.3% of the funds have performance that depends critically on the specific choice of SDFs (or asset pricing models). Furthermore, there exists an admissible SDF that can change the sign of the performance for these funds. More positively, we cannot rule out that more than 80% of the funds could be valued positively by some investors in incomplete markets.

Looking at the bounds presented in panel B of table 3, we can also evaluate the performance of the investment objective fund portfolios. Out of the 14 investment objectives, four portfolios are assigned a negative performance by our admissible measures: the specialty portfolio, the natural resources portfolio, the utilities portfolio and the real estate portfolio. The performances of the other investment objective portfolios are 'gray'.[23] If we are not willing to make auxiliary assumptions to increase the precision of our measures, it is difficult to sign the performance of the investment objective portfolios.

To evaluate the effect of sampling errors on our results, table 4 examines the significance of the bounds using the small sample distributions introduced previously. Panel A classifies the 320 mutual funds into mutually exclusive groups based on whether or not their upper and lower bounds are significantly different than zero. The results in panel A vary widely depending on the desired significance level. For example, only

[23]We might expect that the performances of the objective portfolios are too optimistic because only surviving funds are included in each portfolios. Elton, Gruber, and Blake (1996) present estimates of survivorship bias as a function of the number of years in the study. For a 14-year sample, they document survivorship bias varying from 25.4 basis points per year (2.117 basis points per month) to 71.9 basis points per year (5.992 basis points per month), depending on the model and reinvestment assumptions. Using their highest estimate, two additional investment objective portfolios received a negative performance by our bounds: the growth and income portfolio and the equity income portfolio.

about a quarter of the funds have one significant bound at the 5% level, while all funds have at least one significant bound at the 20% level. Focusing on a significance level of 15%, 191 funds have a lower bound smaller than zero, while 158 funds have an upper bound greater than zero. Signing the performance of mutual funds without incorrect inference is difficult, as just one fund is given a significantly positive performance by its lower bound, while three funds are given a significantly negative performance by their upper bound. Moreover, a number of funds are significantly valued positively by some investors, while negatively by others. Hence, 34 funds have both a lower bound smaller than zero and an upper bound larger than zero, a number increasing to 139 at the 20% significance level. Panel B presents the p-values of the bounds for the investment objective portfolios. Even though it is not possible to significantly sign the performance of any portfolio, respectively nine and six of the 14 portfolios have their lower and upper bounds different from zero at the 15% significance level. Overall, although the bounds are not always precisely estimated, our general conclusions about the difficulty associated with signing the performance of mutual funds without incorrect inference and about the potentially positive valuation of many mutual funds by some investors remain.

6.3.2 Mutual Fund Debates

We can use our performance bounds to make new observations on two highly debated and still ongoing issues regarding the mutual fund industry. The first issue, which arose from the negative performance results of Jensen (1968), is whether or not the mutual fund industry provides valuable services. As an indication of the performance of the universe of mutual funds, we compute the bounds for an equally-weighted portfolio of all funds in our sample. The worst and best performance values on this portfolio are -0.337% and 0.091%, respectively.[24] Thus, there exist at least a SDF that value positively the universe of mutual funds. Furthermore, our results do not rule out the 'efficiency with costly information' argument advanced by Grossman and Stiglitz (1980).

The second issue, brought forward by proponents of the 'Efficient Market Hypothesis', is whether mutual funds should be managed actively or passively. The actively managed fund *Fidelity Magellan* and the passively managed fund *Vanguard 500 Index* represent well the debate as they are, with roughly $45 and $67 billions under management respectively, two of the largest funds in the U.S. as of October 2007. In our sample, although *Fidelity Magellan* has a slightly higher average return, *Vanguard 500 Index* presents

[24] Even after considering reasonable estimates of survivorship bias, it is not possible to sign the performance of this portfolio.

slightly higher performance bounds. Overall, the performance of both funds is 'gray', indicating that the debate could never end since the specific choice of SDF will determine the winner. In fact, if markets are incomplete, there could potentially be rational investors preferring active management, while others favoring passive management.

6.3.3 Discussion

Our results show that it is difficult to evaluate precisely the performance of mutual funds. Without making auxiliary assumptions, it is often not possible to sign the performance of mutual funds. Our findings show the importance of the benchmark model choice, and illustrate that inference errors can have a strong effect on the measurement of portfolio performance. They complement the existing empirical literature on the sensitivity of performance to the benchmark chosen (see Lehmann and Modest (1987), Elton, Gruber, Das, and Hlavka (1993), Grinblatt and Titman (1994), Ferson and Schadt (1996), Carhart (1997), Chan, Dimmock, and Lakonishok (2006)).

More positively, our results support the casual observation that, given the number of investors and amount of money involved, mutual funds must be valuable to some. As mutual funds could cater to the investor class that values their services the most, the results of our upper bounds indicate than a large number of the mutual funds could add value to their target clientele. If markets are truly incomplete, then heterogeneous preferences underlying the infinite number of admissible SDFs could provide wide-ranging performance measures, explaining current disagreement on the value of mutual funds. In incomplete markets, our results suggest that the features, presented by Gruber (1996), on the growth in actively managed mutual funds might not be puzzling.

6.4 Performance Ranking

To illustrate the three performance ranking rules of section 2.5, table 5 presents the bounds on the performance ranking for the investment objective portfolios. Universal Dominance (panel A) is established by finding price bounds on the differential of payoffs between two funds. For example, the lower and upper price bounds on the payoffs of the growth portfolio minus the payoffs of the natural resources portfolio are 0.00188 and 0.00965, respectively. Hence, the growth portfolio dominates the natural resources portfolio in the Universal Dominance sense. Our ranking indicates that the best rank of the natural resource portfolio is tenth. This means that nine other portfolios Universally Dominate the natural resource portfolio. Also, the worst rank of the health portfolio is tenth. Thus, the health portfolio dominates four other portfolios, namely the specialty, natural resources, utilities and real estate portfolios.

The Best Case Scenario and the Worst Case Scenario rules (panel B) rank funds according to the upper bounds and the lower bounds respectively. For example, the health portfolio is ranked first according to the Worst Case Scenario ranking while it is ranked fourth according to the Best Case Scenario. Both rules give the same ranking to three portfolios: the aggressive growth portfolio (sixth), the real estate portfolio (13th), and the natural resources portfolio (14th). The largest discrepancy comes from the technology portfolio which is ranked third according to the Best Case Scenario rule, but 11th according to the Worst Case Scenario rule. If what matters most to mutual funds is their value to their most favorable investor class, then the precious metals portfolio is the most valuable to its clientele as it is ranked first according to the Best Case Scenario rule.

Overall, our results suggest that investors with wide-ranging preferences could rank mutual funds very differently in incomplete markets and highlight the difficulty of ranking mutual funds while avoiding incorrect inference. Not only the ranking of mutual funds can be altered greatly from one model to another, but a ranking that attempts to consider the preferences of the fund's target clientele could be significantly different from one with no such distinction.

6.5 Diagnosis of Performance Evaluation Models

We now examine the performance measures obtained by the 11 models presented in Section 4. The models were chosen because of their popularity and their potential to shed light on the sources of inference errors. Table 6 presents the results. Panel A gives statistics on the cross-sectional distribution of the abnormal returns for the 320 mutual funds. The t-statistics test the hypothesis that the mean is equal to zero.[25] The overall performance of the mutual funds (as given by the mean performance measure) is significantly negative for the CAPM, the conditional CAPM (CAPM-C) and the three nonparametric models (MINLOP, MINNA, MAXNA), while significantly positive for the Ferson-Schadt model (FSM), the conditional Ferson-Schadt model (FSM-C), and the power utility (POWER) and habit-formation (HABIT) consumption-based models.

[25]The t-statistics are computed by assuming that the cross-sectional distribution of the abnormal returns is multivariate normal with a mean of zero, a standard deviation as reported in Table 6, and a correlation between any two abnormal returns of 0.68, which corresponds to the average correlation between the returns of the investment objective portfolios.

The overall performance of the mutual funds is not significantly different than zero for Fama-French (FFM) and conditional Fama-French (FFM-C) models.

To provide a diagnostic of the models, we investigate their inference errors by comparing their performance values with the bounds obtained using the infinite set of admissible measures. Panel B of table 6 gives the percentage of performance values that fall outside the bounds and hence are not admissible. Except for MINNA and MAXNA (constructed to be in the admissible set), panel B reveals the presence of inference errors that vary widely across models. The percentage of non-admissible values goes from 0.62% for MINLOP to 52.19% for POWER. Furthermore, all models except MINLOP present an upward bias: they have a higher percentage of non-admissible values above the upper bounds than below the lower bounds. We now examine the inference errors for models by category.

The first category of models are linear factor models. These models are widely used, and arguably the most successful parametric models in pricing equities. However, they are not admissible empirically since they do not price the basis assets correctly and give rise to negative SDFs. Furthermore, Ghysels (1998) argues that these problems might be worse for the conditional models than their unconditional counterparts. How important are these shortcomings in the context of performance evaluation? CAPM, CAPM-C, FFM and FFM-C present a relatively low percentage of non-admissible performance values, approximately 10%. This is not the case for FSM and FSM-C, which have around 30% of non-admissible values. Given their important upward bias, the performance measures of FSM and FSM-C also appear to overestimated considerably the performance of mutual funds in our sample.

The second category of models are consumption-based models. These models do not imply negative SDFs, but generally perform poorly in pricing financial assets. The consumption-based models present the most inference errors and have a large upward bias. HABIT generates a smaller percentage of non-admissible values than POWER, and is, in this sense, more comparable to FSM-C. The last category of models are nonparametric models. These models price correctly the basis assets by construction. MINNA and MAXNA are two of the infinite admissible models. Their performance values could be used if a precise estimate of performance is required. Interestingly, the choice between both measures represents an economically significant dilemma, since MAXNA gives an overall performance almost two times more negative than MINNA. MINLOP is not an admissible model since it does not impose the positivity constraint on SDFs.

Its results show inference errors for only two mutual funds. Comparing the results for all three categories of models, our diagnostic suggests that pricing correctly the basis assets is more important than imposing the positivity of SDFs in reducing the percentage of non-admissible measures.

In summary, our results show that parametric models can present a large percentage of non-admissible performance values. Furthermore, some models have an intriguing tendency to obtain more non-admissible values above the upper bounds than below the lower bounds, suggesting an upward bias in their performance evaluation. As falling inside the bounds is a necessary but not a sufficient condition for a performance measure to be admissible, our results represent a conservative look at the 'bad model' problem in performance evaluation. However, our analysis remains only a diagnostic of selected models. Formal tests of whether candidate measures fall outside admissible performance bounds are left for future research.

7 Conclusion

This paper addresses the following critical question, 'What is the admissible set of performance measures?' Instead of attempting to pursue a point estimate, we take a diametrically opposite position by examining a potentially infinite admissible set of performance measures. In that regard, our approach is in the spirit of Hansen and Jagannathan (1991). Whereas the Hansen-Jagannathan bound can be used as a guideline for asset pricing theories, our performance bounds can be used as a yardstick for the development of better performance measures. Furthermore, the bounds themselves are useful in evaluating mutual fund performance if they are relatively tight. In that sense, our approach is comparable with Cochrane and Saá-Requejo (2000). Therefore, our bounds can be counted as a double-edge sword: a diagnostic tool for evaluating alternative parametric performance measures and a stand-alone performance measure of mutual funds.

Empirically, our results demonstrate that measuring the performance of mutual funds is a difficult exercise in incomplete markets. In fact, without making auxiliary assumptions on SDFs, it is often possible to obtain an economically important range of performance values, justifying the casual observation that mutual funds are valuable to some agents willing to invest large sums of money. Furthermore, our results show that the potential for inference errors is large in performance evaluation and ranking of mutual funds, justifying the performance sensitivity documented in the literature. Finally, they suggest that some existing parametric performance models present a considerable percentage of non-admissible performance values.

The performance bounds that we develop herein can be used for evaluating the performance of other trading strategies in the field of investments as well as corporate finance. For example, our performance bounds can be used to determine abnormal returns after corporate events, such as seasoned equity offerings. In addition, bounds could be developed to evaluate segmentation across borders, or markets, extending the work of Chen and Knez (1995). Finally, our diagnostic instrument on the admissibility of performance measures could be expanded to a formal testing procedure of whether performance values fall outside the bounds. These applications are left for future research.

Appendix A: Proofs of Propositions

Proof of Proposition 1: We first prove the existence of a positive stochastic discount factor and secondly verify the existence of an infinite number of such stochastic discount factors.[26] Denote the set of arbitrage trading opportunities by $\mathcal{H} \stackrel{\triangle}{=} \{(-cy, y) \stackrel{+}{>} (0, 0)\}$, which is a subspace of $R \times \mathcal{L}^2$, where cy denotes the cost of trading to get the random payoff y. The premise of viable price system means that $(R \times \mathcal{A}) \cap \mathcal{H} = \emptyset$. The Separating Hyperplane Theorem implies the existence of continuous linear functions $f : R \times \mathcal{L}^2 \to R$ such that $f = 0$ if $(-cy, y) \in (R_- \times \mathcal{A})$ and $f > 0$ if $(-cy, y) \in \mathcal{H}$ since both $\mathcal{A}$ and $\mathcal{H}$ are closed and convex sets. Let $f(-cy, y) = -cy + \langle M|y\rangle$, after normalization. Since the basis assets are achievable,

$$f(-1_N, x) = -1_N + \langle M|x\rangle = 0.$$

First $E^P[M] = 1/R_f$ if the risk-free asset exists. Second, we claim that $M > 0$. For any $y \stackrel{+}{>} 0$, $(0, y) \in \mathcal{H}$. Thus

$$f(0, y) = 0 + \langle M|y\rangle > 0.$$

The fact that the above inequality holds for *any* $y \stackrel{+}{>} 0$ yields $M > 0$. Finally, since the market is incomplete, the prices of additional assets with unspanned payoffs will not be uniquely determined. Thus, there must exist more than two positive pricing kernels. Since any convex combination of pricing kernels is a pricing kernel, there exists an infinite number of pricing kernels. □

Proof of Proposition 2: Since the payoffs are all positive and there are only finite states, the pricing kernels are bounded from above and below and are thus in a closed compact set. As a result, the upper bounds and lower bounds exist and are attainable. Moreover, since pricing kernels form a convex set, the performance measures span the whole interval between the lower bound and the upper bound. □

Proof of Proposition 3: First, we show that $\sigma(M)$ has an upper bound. We assume that all assets have finite first moment and second moment. Therefore, $E^P[x_j^2] < \infty, \forall j$. Let ω_0 be an arbitrary state. We

[26] The proof of proposition 1 is not new. For more details, see Harrison and Kreps (1979) and Duffie (1996) among others. Here we provide the proof for the paper to be self-contained.

have:

$$\begin{aligned} E^P[M(\omega)^2] &= E^P[(M(\omega)-M(\omega_0))^2 + 2M(\omega)M(\omega_0) - M(\omega_0)^2] \\ &= 2M(\omega_0)/(1+r_f) - M(\omega_0)^2 + \int_\omega (M(\omega)-M(\omega_0))^2 dP(\omega) \\ &\leq 2M(\omega_0)/(1+r_f) - M(\omega_0)^2 + B^2 \int_\omega \sum_{j=1}^N (x_j(\omega)-x_j(\omega_0))^2 dP(\omega) \\ &= 2M(\omega_0)/(1+r_f) - M(\omega_0)^2 + B^2 \sum_{j=1}^N E^P[(x_j(\omega)-x_j(\omega_0))^2]. \end{aligned}$$

Taking the expectation with respect of ω_0, notice that $E^P[M(\omega_0)] = 1/(1+r_f)$, we get

$$E^P[M(\omega)^2] \quad \leq \quad 2/(1+r_f)^2 - E^P[M(\omega_0)^2] + B^2 \sum_{j=1}^N E^P[(x_j(\omega)-x_j(\omega_0))^2].$$

Since ω and ω_0 represent separate and independent sampling from the same sample space, we have $E^P[M(\omega)^2] = E^P[M(\omega_0)^2]$. Consequently,

$$2E^P[M(\omega)^2] \leq 2/(1+r_f)^2 + B^2 \sum_{j=1}^N E^P[x_j^2] - (E^P[x_j])^2 = 2/(1+r_f)^2 + B^2 \sum_{j=1}^N \sigma_j^2,$$

which implies that

$$E^P[M(\omega)^2] \leq 1/(1+r_f)^2 + \frac{1}{2} B^2 \sum_{j=1}^N \sigma_j^2,$$

where σ_j^2 is the variance of asset j. Since all assets have finite first moment and second moment, $E^P[M^2]$ is bounded. Consequently all Ms have finite first moment and second moment bounded from above. Let $\bar{U} \equiv \sup_{M \in \mathcal{M}} E^P[M^2]$, then by the Schwartz inequality, we have

$$-\bar{U}||x_{mf}|| \leq \alpha_\$(x_{mf}) \leq \bar{U}||x_{mf}||.$$

Thus, $\alpha_\$(x_{mf})$ is bounded and the set of $\alpha_\$(x_{mf})$ belongs to a bounded, closed and convex set. The upper bound and lower bound of $\alpha_\$(x_{mf})$ exist and are attainable. $\square$.

Appendix B: Two-Phase Revised Simplex Method

The simplex method is a "feasible-point" method: given initial feasible points $\mathbf{M}_0$, all subsequent iterates $\mathbf{M}_k$ are also feasible. It is part of a larger family of methods known as active set methods. The simplex method is readily available in most numerical procedure packages.

In phase one, the procedure finds a basic initial feasible point $\mathbf{M}_0$ to the problem. The technique to find such a point is called phase 1 simplex.[27] It considers an artificial linear objective function made of the sum of infeasibilities (the violated constraints) at $\mathbf{M}$. A feasible point is then found by minimizing this objective function, subject to the constraints that the non-violated constraints remain that way.

Since any basic feasible solution has T binding or active constraints, the difficulty of the problem is really to find what are the T optimal binding constraints. This is done in phase two, which can be described as follows.[28] Let the working set of constraints be the T binding constraints in current iteration, and let $\mathbf{M}_k$ denote the current iterate. $\mathbf{M}_k$ is optimal for the equality-constrained subproblem defined by the working set, and thus the first two necessary and sufficient conditions for an optimum are satisfied.

The next step is to check the sign of the Lagrange multiplier δ_k when $M_k = 0$. If $\delta_k > 0$ when $M_k = 0$, then the third necessary and sufficient conditions is met and $\mathbf{M}_k = \mathbf{M}^\star$. However, if any δ_k is negative when $M_k = 0$ (say, $\delta_k < 0$), then objective function can be improved by stepping in a direction that makes inactive the constraint $M_s \geq 0$ and keeps the other active constraints identical. This produces a unique search direction, and a maximum feasible step to the nearest constraint not in the working set is taken. The process is then repeated with the new working set.

Methods for linear programming differ mainly in the way in which the Lagrange multipliers and the search directions are computed. Finding the Lagrange multipliers or the search directions each involve solving a system of linear equations. The two-phase revised simplex method used the LU factorization[29] of the matrix that needs to be inverted for solving the system, a very efficient method for large-scale linear programming. The technique for updating the LU factorization is known as the Bartels-Golub scheme, which ensure numerical stability by using row interchanges during the updating.

[27] See Gill, Murray, and Wright (1981, Section 5.7).
[28] See Gill, Murray, and Wright (1981, Sections 5.3.1 and 5.6.1).
[29] See Gill, Murray, and Wright (1981, Section 2.2.5.1).

References

Bekaert, Geert, and Jun Liu, 2004, Conditioning information and variance bounds on pricing kernels, *Review of Financial Studies* 17, 339–378.

Bekaert, Geert, and Michael S. Urias, 1996, Diversification, integration and emerging market closed-end funds, *The Journal of Finance* 51, 835–869.

Bernardo, Antonio E., and Olivier Ledoit, 2000, Gain, loss and asset pricing, *Journal of Political Economy* 108, 144–172.

Blume, Marshall, Donald B. Keim, and S. A. Patel, 1991, Returns and volatility of low-grade bonds, 1977-1989, *Journal of Finance* 46, 49–74.

Brown, Stephen J., and William N. Goetzmann, 1995, Performance persistence, *Journal of Finance* 50, 679–698.

Brown, Steven J., William N. Goetzmann, Roger G. Ibbotson, and Stephen A. Ross, 1992, Survivorship biases in performance studies, *Review of Financial Studies* 5, 553–580.

Campbell, John Y., and John H. Cochrane, 1999, By force of habit: A consumption based explanation of aggregate stock market behavior, *Journal of Political Economy* 107, 205–251.

Carhart, Mark M., 1997, On persistence in mutual fund performance, *Journal of Finance* 52, 57–82.

Chan, Louis K.C., Stephen G. Dimmock, and Josef Lakonishok, 2006, Benchmarking money manager performance: Issues and evidence, NBER working paper no. 12461.

Chen, Zhiwu, and Peter J. Knez, 1995, Measurement of market integration and arbitrage, *Review of Financial Studies* 8, 287–325.

——— , 1996, Portfolio performance measurement: Theory and applications, *Review of Financial Studies* 9, 511–555.

Christopherson, on A., Wayne E. Ferson, and Debra A. Glassman, 1998, Conditioning manager alphas on economic information: Another look at the persistence of performance, *The Review of Financial Studies* 11, 111–142.

Cochrane, John H., 1996, A cross-sectional test of an investment-based asset pricing model, *Journal of Political Economy* 104, 572–621.

——— , and Jesús Saá-Requejo, 2000, Beyond arbitrage: "good-deal" asset price bounds in incomplete markets, *Journal of Political Economy* 108, 79–119.

Coles, Jeffrey L., Naveen D. Daniel, and Federico Nardari, 2006, Does the choice of model or benchmark affect inference in measuring mutual fund performance?, Arizona State University working paper.

Connor, Gregory, and Robert A. Korajczyk, 1986, Performance measurement with the arbitrage pricing theory: A new framework for analysis, *Journal of Financial Economics* 17, 255–289.

Dahlquist, Magnus, and Paul Söderlind, 1999, Evaluating portfolio performance with stochastic discount factors, *Journal of Business* 72, 347–383.

Duffie, Darrell, 1996, *Dynamic Asset Pricing Theory* (Princeton University press: New Jersey) second edn.

Dybvig, Philip H., and Stephen A. Ross, 1985a, The analytics of performance measurement using a security market line, *Journal of Finance* 40, 401–416.

——— , 1985b, Differential information and performance measurement using a security market line, *Journal of Finance* 40, 383–399.

Elton, Edwin. J., Martin. J. Gruber, and Christopher R. Blake, 1996, Survivorship bias and mutual fund performance, *Review of Financial Studies* 9, 1097–1120.

Elton, Edwin. J., Martin. J. Gruber, Sanjiv Das, and Christopher R. Blake, 1996, The persistence of risk-adjusted mutual fund performance, *Journal of Business* 69, 133–157.

Elton, Edwin. J., Martin. J. Gruber, Sanjiv Das, and Matthew Hlavka, 1993, Efficiency with costly information: A reinterpretation of evidence from managed portfolios, *Review of Financial Studies* 6, 1–22.

Fama, Eugene F, 1998, Market efficiency, long-term returns, and behavioral finance, *Journal of Financial Economics* 49, 283–306.

Fama, Eugene F., and Kenneth R. French, 1989, Business conditions and expected returns on stocks and bonds, *Journal of Financial Economics* 25, 23–49.

——— , 1993, Common risk factors in the returns on stocks and bonds, *Journal of Financial Economics* 33, 3–56.

Farnsworth, Heber, Wayne E. Ferson, David Jackson, and Steven Todd, 2002, Performance evaluation with stochastic discount factors, *Journal of Business* 75, 473–503.

Ferson, Wayne, 1989, Changes in expected security returns, risk, and the level of interest rates, *The Journal of Finance* 44, 1191–1217.

Ferson, Wayne E., and Campbell R. Harvey, 1991, The variation of economic risk premiums, *Journal of Political Economy* 99, 385–415.

Ferson, Wayne E., and Rudi W. Schadt, 1996, Measuring fund strategy and performance in changing economic conditions, *Journal of Finance* 51, 425–461.

Gallant, A. Ronald, Lars Peter Hansen, and George Tauchen, 1990, Using conditional moments of asset payoffs to infer the volatility of intemporal rates of substitution, *Journal of Econometrics* 45, 141–179.

Ghysels, Eric, 1998, On stable factor structure in the pricing of risk: Do time-varying betas help or hurt?, *Journal of Finance* 53, 549–573.

Gill, Philip E., Walter Murray, and Margaret H. Wright, 1981, *Practical Optimization* (Academic Press, Inc.: San Diego) first edn.

Goetzmann, William N., and Roger G. Ibbotson, 1994, Do winners repeat: Predicting mutual fund performance, *Journal of Portfolio Management* 20, 9–18.

Green, Richard C., 1986, Benchmark portfolio inefficiency and deviations from the security market line, *Journal of Finance* 41, 295–312.

Grinblatt, Mark, and Sheridan Titman, 1989, Portfolio performance evaluation: Old issues and new insights, *Review of Financial Studies* 2, 393–422.

——— , 1992, The persistence of mutual fund performance, *Journal of Finance* 47, 1977–1984.

——— , 1994, A study of monthly mutual fund returns and performance evaluation techniques, *Journal of Financial and Quantitative Analysis* 29, 419–444.

Grossman, Sanford J., and Joseph E. Stiglitz, 1980, On the impossibility of informationally efficient markets, *American Economic Review* 70, 393–408.

Gruber, Martin. J., 1996, Another puzzle: The growth in actively managed mutual funds, *Journal of Finance* 51, 783–810.

Hansen, Lars P., 1982, Large sample properties of generalized method of moments estimators, *Econometrica* 50, 1029–1054.

——— , and Ravi Jagannathan, 1991, Implications of security market data for models of dynamic economies, *Journal of Political Economy* 99, 225–262.

——— , 1997, Assessing specification errors in stochastic discount factor models, *Journal of Finance* 52, 557–590.

Harrison, J. Michael, and David M. Kreps, 1979, Martingales and arbitrage in multiperiod securities markets, *Journal of Economic Theory* 20, 381–408.

Hendricks, Darryll, Jayendu Patel, and Richard Zeckhauser, 1993, Hot hands in mutual funds: Short-run persistence of relative performance, 1974-1988, *Journal of Finance* 48, 93–130.

Jensen, Michael C, 1968, The performance of mutual funds in the period 1945-1964., *Journal of Finance* 23, 389–416.

———, 1969, Risk, the pricing of capital assets, and the evaluation of investment portfolios, *Journal of Business* 42, 167–247.

King, Benjamin F., 1966, Market and industry factors in stock price behavior, *Journal of Business* 39, 139–190.

Kothari, S.P., and Jerold B. Warner, 2001, Evaluating mutual fund performance, *Journal of Finance* 56, 1985–2010.

Lehmann, Bruce, and David Modest, 1987, Mutual fund performance evaluation: A comparison of benchmarks and benchmark comparisons, *Journal of Finance* 42, 233–265.

Lo, Andrew W., and Craig A. MacKinlay, 1990, Data-snooping biases in tests of financial asset pricing models, *The Review of Financial Studies* 3, 431–467.

Long, John B., 1990, The numeraire portfolio, *Journal of Financial Economics* 26, 29–70.

Malkiel, Burton G., 1995, Returns from investing in equity mutual funds 1971 to 1991, *Journal of Finance* 50, 549–572.

Merton, Robert C., 1973, An intertemporal capital asset pricing model, *Econometrica* 41, 867–887.

Moskowitz, Tobias J., and Mark Grinblatt, 1999, Do industries explain momentum?, *Journal of Finance* 54, 1249–1290.

Roll, Richard W., 1978, Ambiguity when performance is measured by the securities market line, *Journal of Finance* 33, 1051–1069.

Ross, Stephen A., 1976, The arbitrage theory of capital asset pricing, *Journal of Economic Theory* 13, 341–360.

Sharpe, William F., 1966, Mutual fund performance, *Journal of Business* 39, 119–138.

Treynor, Jack, 1965, How to rate management of investment funds, *Harvard Business Review* 43, 63–75.

Wermers, Russ, 2000, Mutual fund performance: An empirical decomposition into stock-picking talent, style, transactions costs, and expenses, *Journal of Finance* 55, 1655–1695.

Table 1: Summary Statistics for the Mutual Funds

This table gives summary statistics for the mutual fund gross returns using monthly data from January 1984 to December 1997. Panel A gives statistics on the distribution of the averages, standard deviations, minimum, maximum and Sharpe ratios for the sample of 320 mutual funds. Panel B gives the number of funds per portfolio, the average return, the standard deviation of returns and the Sharpe ratio for equally-weighted portfolios of funds grouped by investment objectives.

Panel A: Individual Mutual Funds

Statistics	Ret Avg	Std Dev	Minimum	Maximum	Sharpe
Mean	1.01164	0.04613	0.77612	1.14170	0.16475
Std Dev	0.00307	0.01277	0.05650	0.04770	0.06679
Min	0.99236	0.01805	0.50379	1.05829	-0.20499
1%	0.99879	0.02484	0.65048	1.07285	-0.06938
5%	1.00663	0.03006	0.68998	1.09102	0.03885
10%	1.00911	0.03387	0.70865	1.09933	0.09859
25%	1.01082	0.03867	0.74024	1.11570	0.13945
Median	1.01215	0.04378	0.77892	1.13438	0.17357
75%	1.01318	0.05057	0.80776	1.15263	0.20777
90%	1.01405	0.06020	0.84998	1.17894	0.23032
95%	1.01510	0.07020	0.87608	1.22723	0.24209
99%	1.01724	0.08680	0.91148	1.33951	0.29832
Max	1.01802	0.12761	0.94443	1.45161	0.31546

Panel B: Investment Objective Portfolios

Objectives	N	Ret Avg	Std Dev	Sharpe
Aggressive Growth	23	1.01140	0.05371	0.12715
Growth	135	1.01227	0.04248	0.18124
Growth and Income	81	1.01157	0.03609	0.19394
Equity Income	16	1.01135	0.03158	0.21467
Small Company	30	1.01193	0.04738	0.15532
Specialty	1	1.00926	0.03833	0.12234
Specialty–Health	3	1.01479	0.04994	0.20463
Specialty–Financial	3	1.01536	0.04615	0.23378
Specialty–Nat Res	5	1.00868	0.04385	0.09371
Specialty–Prec Metals	9	1.00195	0.08677	-0.03020
Specialty–Technology	6	1.01336	0.06208	0.14158
Specialty–Utilities	6	1.01052	0.02652	0.22433
Specialty–Real Estate	1	1.00962	0.02907	0.17369
Specialty–Comm	1	1.01436	0.04299	0.22771

Table 2: Summary Statistics for the Benchmark Assets

This table gives summary statistics for the monthly gross returns of the industry and bond portfolios, and for the instrumental variables. The 20 industry portfolios are formed by grouping stocks according to the SIC codes given in parentheses. The stock returns and SIC codes are obtained from CRSP. The bond portfolio returns are taken from the CRSP Fama Maturity Portfolios Returns File. The credit spread variable is calculated as the difference between the yield on Baa corporate bonds and the yield on long-term Treasury bonds. The yields are taken from CITIBASE. The Dividend Yield variable is calculated as the difference between the CRSP value-weighted index returns with and without dividends. The data cover the period from January 1984 to December 1997, for a total of 168 observations.

Benchmark Assets	Avg	Std Dev	Minimum	Maximum
Industry Portfolios (SIC Codes)				
Mining (10-14)	1.00738	0.05357	0.72656	1.19512
Food (20)	1.01099	0.04155	0.73397	1.09777
Apparel (22-23)	1.00653	0.05407	0.67825	1.15815
Paper (26)	1.01219	0.04934	0.75492	1.19440
Chemical (28)	1.01473	0.06032	0.68366	1.18566
Petroleum (29)	1.01026	0.04749	0.75807	1.14175
Construction (32)	1.01227	0.05032	0.71086	1.13380
Primary Metals (33)	1.00935	0.05413	0.67422	1.14533
Fabricated Metals (34)	1.01375	0.04816	0.70787	1.16559
Machinery (35)	1.01027	0.06001	0.67859	1.17365
Electrical Equipment (36)	1.01146	0.06193	0.68915	1.20473
Transport Equipment (37)	1.01019	0.05012	0.69074	1.14404
Manufacturing (38-39)	1.01061	0.05811	0.68664	1.23006
Railroads (40)	1.01282	0.04433	0.77309	1.15482
Other Transportation (41-47)	1.01063	0.05017	0.69737	1.14556
Utilities (49)	1.01276	0.03185	0.85756	1.09277
Department Stores (53)	1.00950	0.06494	0.69292	1.15281
Retail (50-52, 54-59)	1.00755	0.05217	0.70616	1.15466
Financial (60-69)	1.01247	0.03765	0.79402	1.13201
Other	1.01076	0.05475	0.69982	1.16685
Bond Portfolios (Maturity)				
Short-Term (< 1 year)	1.00564	0.00239	1.00018	1.01459
Long-Term (> 10 years)	1.01016	0.02762	0.94545	1.10438
Instrumental Variables				
Credit Spread (Annual %)	1.64387	0.37719	1.14000	2.60000
Dividend Yield (Annual %)	3.10308	1.23744	1.28760	7.05960

Table 3: Performance Measurement Bounds

This table presents the lower and upper bounds on portfolio managers' α_r using monthly data from January 1984 to December 1997. Panel A gives statistics on the distribution of the results for the sample of 320 mutual funds. t-stat are the values of the t-statistic for the hypotheses that the mean is equal to zero assuming that the cross-sectional distribution of the bounds is multivariate normal with a mean of zero, a standard deviation as given by Std Dev, and a correlation between any two lower or any two upper bounds of 0.68. For the Bound Diff and Bound Avg t-statistics, a correlation between the lower and upper bounds of 0.35 is also assumed. Panel B gives the results for equally-weighted portfolios of funds grouped by investment objectives.

Panel A: Individual Mutual Funds

Statistics	Lower Bound	Upper Bound	Bound Diff	Bound Avg
Mean	-0.00578	0.00316	0.00895	-0.00131
Std Dev	0.00372	0.00543	0.00539	0.00379
(t-stat)	(-18.0767)	(6.7705)	(19.2747)	(-4.0211)
Min	-0.02852	-0.01181	0.00311	-0.02017
1%	-0.02064	-0.00453	0.00393	-0.01140
5%	-0.01161	-0.00169	0.00441	-0.00636
10%	-0.00953	-0.00088	0.00520	-0.00468
25%	-0.00713	0.00072	0.00610	-0.00306
Median	-0.00532	0.00232	0.00754	-0.00129
75%	-0.00355	0.00405	0.00976	-0.00002
90%	-0.00224	0.00710	0.01323	0.00218
95%	-0.00118	0.00994	0.01791	0.00347
99%	0.00090	0.02335	0.02859	0.01031
Max	0.00243	0.05447	0.05362	0.02767

Panel B: Investment Objective Portfolios

Objectives	Lower Bound	Upper Bound
Aggressive Growth	-0.00432	0.00232
Growth	-0.00394	0.00103
Growth and Income	-0.00381	0.00040
Equity Income	-0.00315	0.00042
Small Company	-0.00380	0.00189
Specialty	-0.00645	-0.00001
Specialty–Health	-0.00258	0.00608
Specialty–Financial	-0.00905	0.00245
Specialty–Nat Res	-0.01154	-0.00318
Specialty–Prec Metals	-0.00495	0.02515
Specialty–Technology	-0.00721	0.00790
Specialty–Utilities	-0.00503	-0.00037
Specialty–Real Estate	-0.01069	-0.00305
Specialty–Comm	-0.00510	0.00992

Table 4: Statistical Significance of the Performance Measurement Bounds

This table examines the statistical significance of the lower and upper bounds on portfolio managers' α_r. For each individual mutual funds and investment objective portfolios, we obtain the empirical small sample distribution from Monte Carlo techniques. The returns of the basis assets and the funds were generated assuming a multivariate normal distribution. The significance is based on 5000 simulations of 168 observations for each fund. Panel A classifies the 320 individual mutual funds in mutually exclusive groups based on their results, at various significance levels, in tests of the null hypothesis that the bounds are equal to zero. The classifications, given in the first column, are based on joint results on the lower bounds (LB) and upper bounds (UB). For example, $LB < 0 + UB > 0$ are funds with a significantly negative lower bound *and* a significantly positive upper bound, while $LB < 0 + UB = 0$ are funds with a significantly negative lower bound *and* an insignificant upper bound. Panel B gives the p-values for the null hypothesis that the lower or upper bounds on the equally-weighted portfolios of funds grouped by investment objectives are equal to zero.

Panel A: Individual Mutual Funds

# of Funds in	For Significance Level of			
Classification	5%	10%	15%	20%
$LB > 0 + UB > 0$	0	0	1	2
$LB < 0 + UB < 0$	1	2	3	3
$LB < 0 + UB > 0$	0	0	34	139
$LB < 0 + UB = 0$	49	112	154	98
$LB = 0 + UB > 0$	33	93	123	78
$LB = 0 + UB = 0$	237	113	5	0

Panel B: Investment Objective Portfolios

Objectives	LB p-value	UB p-value
Aggressive Growth	0.17391	0.07971
Growth	0.16154	0.14615
Growth and Income	0.06618	0.25000
Equity Income	0.07843	0.26797
Small Company	0.29787	0.04965
Specialty	0.01887	0.33019
Specialty–Health	0.21333	0.09333
Specialty–Financial	0.14388	0.16547
Specialty–Nat Res	0.01449	0.38406
Specialty–Prec Metals	0.07752	0.15504
Specialty–Technology	0.14400	0.07200
Specialty–Utilities	0.02516	0.38365
Specialty–Real Estate	0.02113	0.45775
Specialty–Comm	0.29688	0.07031

Table 5: Performance Ranking

This table presents the ranking of investment objective portfolios using monthly data from January 1984 to December 1997. Panel A gives bounds on ranking using the Universal Dominance rule, established by finding positive price bounds on the differential of payoffs between two funds. Panel B gives the ranking using two Scenario Dominance rules. The Worst Case Scenario Dominance rule is based uniquely on the lower bound of each fund. The Best Case Scenario Dominance is based uniquely on the upper bound of each fund. The investment objective portfolios are equally-weighted portfolios of funds constructed from the 320 mutual funds.

Panel A: Universal Dominance

Objectives	Best Rank	Worst Rank
Aggressive Growth	1	12
Growth	1	12
Growth and Income	1	12
Equity Income	1	12
Small Company	1	12
Specialty	2	14
Specialty–Health	1	10
Specialty–Financial	1	14
Specialty–Nat Res	10	14
Specialty–Prec Metals	1	12
Specialty–Technology	1	14
Specialty–Utilities	2	12
Specialty–Real Estate	10	14
Specialty–Comm	1	12

Panel B: Scenario Dominance

Objectives	Worst Case Rank	Best Case Rank
Aggressive Growth	6	6
Growth	5	8
Growth and Income	4	10
Equity Income	2	9
Small Company	3	7
Specialty	10	11
Specialty–Health	1	4
Specialty–Financial	12	5
Specialty–Nat Res	14	14
Specialty–Prec Metals	7	1
Specialty–Technology	11	3
Specialty–Utilities	8	12
Specialty–Real Estate	13	13
Specialty–Comm	9	2

Table 6: Diagnosis of Performance Measurement Models

This table presents results on the performance measurement of alternative models using monthly data from January 1984 to December 1997. Panel A gives statistics on the distribution of the results for the sample of 320 mutual funds. t-stat are the values of the t-statistic for the hypotheses that the mean is equal to zero assuming that the cross-sectional distribution of the abnormal returns is multivariate normal with a mean of zero, a standard deviation as given by Std Dev, and a correlation between any two abnormal returns of 0.68. Panel B examines the percentage of the performance measures that falls outside the bounds on α_r. CAPM, FFM and FSM are the performance measures using the Capital Asset Pricing Model, the Fama-French three-factor model, and the Ferson-Schadt four-factor model, respectively. The conditional versions of these models (indicated by -C) assume that the factor coefficients are linear functions of the instrumental variables. The instruments are the lagged values of the one-month bill yield from the CRSP riskfree files, the lagged values of the dividend yield implied by the CRSP value-weighted index, the lagged term spread, the lagged corporate credit spread, and a January dummy. The results are obtained using a linear regression method with excess returns. POWER and HABIT are the performance measures using the consumption-based asset pricing models with power utility and with external habit-formation preference, respectively. MINLOP, MINNA and MAXNA are the performance measures using the stochastic discount factors with the minimum standard deviation under the law of one price, and the minimum and maximum standard deviation under no arbitrage, respectively.

Panel A: Results on Individual Mutual Funds

Statistics	CAPM	CAPM-C	FFM	FFM-C	FSM	FSM-C	POWER	HABIT	MINLOP	MINNA	MAXNA
Mean	-0.00115	-0.00110	-0.00023	-0.00023	0.00080	0.00077	0.00187	0.00120	-0.00119	-0.00126	-0.00227
Std Dev	0.00281	0.00272	0.00283	0.00266	0.00250	0.00282	0.00304	0.00288	0.00274	0.00338	0.00525
(t-stat)	(-4.76)	(-4.70)	(-0.946)	(-1.01)	(3.72)	(3.18)	(7.16)	(4.85)	(-5.05)	(-4.34)	(-5.03)
Min	-0.02165	-0.02268	-0.02045	-0.02243	-0.01865	-0.02024	-0.01742	-0.01792	-0.01706	-0.01829	-0.02588
1%	-0.01164	-0.01140	-0.01159	-0.01030	-0.00928	-0.01184	-0.01059	-0.01032	-0.00959	-0.01135	-0.01335
5%	-0.00535	-0.00497	-0.00437	-0.00335	-0.00242	-0.00286	-0.00314	-0.00374	-0.00502	-0.00573	-0.00831
10%	-0.00408	-0.00353	-0.00259	-0.00214	-0.00124	-0.00173	-0.00066	-0.00133	-0.00399	-0.00468	-0.00694
25%	-0.00222	-0.00223	-0.00122	-0.00108	0.00008	0.00005	0.00106	0.00037	-0.00234	-0.00287	-0.00451
Median	-0.00081	-0.00077	-0.00002	-0.00005	0.00111	0.00117	0.00238	0.00168	-0.00097	-0.00108	-0.00256
75%	0.00060	0.00052	0.00112	0.00087	0.00216	0.00211	0.00340	0.00264	0.00026	0.00020	-0.00049
90%	0.00139	0.00147	0.00230	0.00228	0.00282	0.00327	0.00429	0.00359	0.00148	0.00175	0.00137
95%	0.00202	0.00207	0.00342	0.00310	0.00352	0.00390	0.00533	0.00453	0.00261	0.00330	0.00390
99%	0.00417	0.00385	0.00631	0.00525	0.00466	0.00515	0.00749	0.00658	0.00496	0.00754	0.01775
Max	0.00456	0.00444	0.00732	0.00668	0.00609	0.00658	0.00826	0.00754	0.01091	0.02129	0.03966

Panel B: Comparison with the Performance Measurement Bounds

	CAPM	CAPM-C	FFM	FFM-C	FSM	FSM-C	POWER	HABIT	MINLOP	MINNA	MAXNA
% < Lower Bound	2.19	2.50	1.25	1	0.63	1.25	1.56	1.56	0.31	0	0
% > Upper Bound	6.56	8.44	9.38	7.50	27.19	31.25	50.63	36.56	0.31	0	0

Figure 1a: Universal Dominance

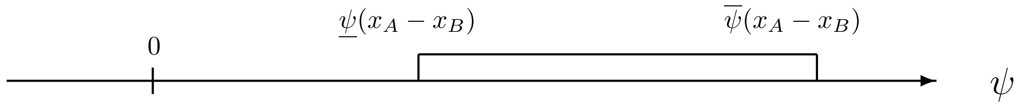

Figure 1b: Best Case Scenario Dominance

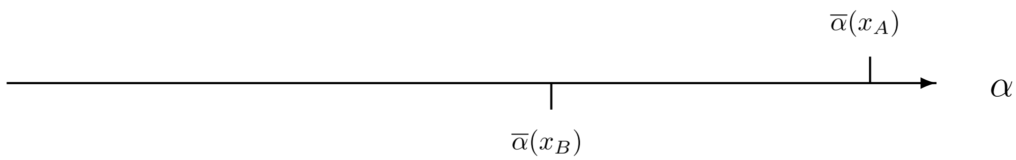

Figure 1c: Worst Case Scenario Dominance

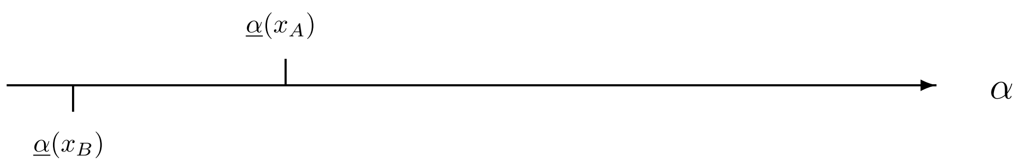

Figure 1: Dominance Rules

Figures 1a, 1b and 1c illustrate the Universal Dominance, the Best Case Scenario Dominance and the Worst Case Scenario Dominance rules, respectively. In each figure, Fund A dominates Fund B.

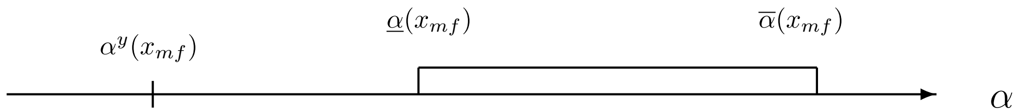

Figure 2: Candidate Performance Measure Not Admissible

The figure gives an example of a candidate performance measure that is not admissible. $\alpha^y(x_{mf})$ represents the performance measure for the candidate SDF y.

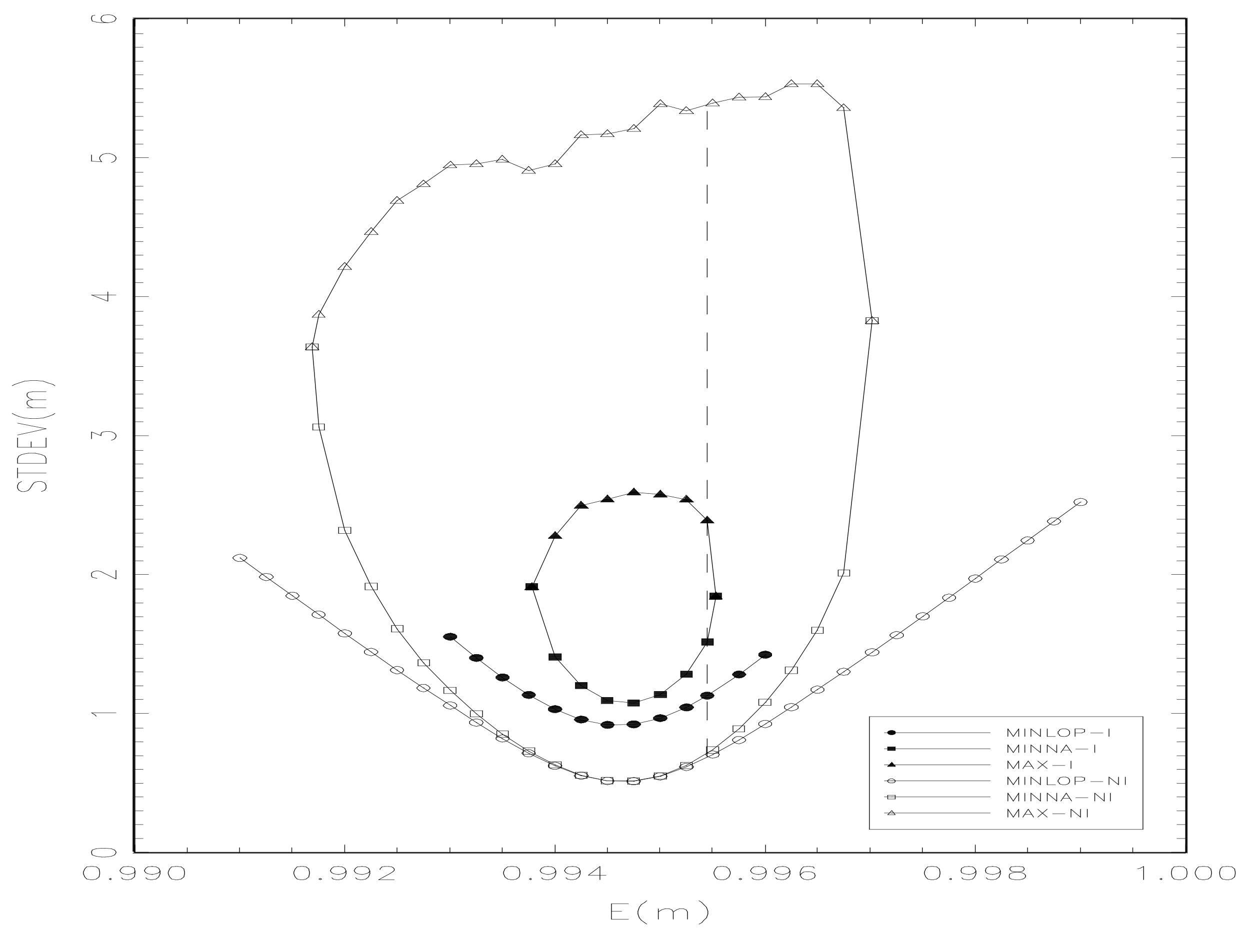

Figure 3: Mean and Standard Deviation Bounds on the Stochastic Discount Factors

MINLOP is the minimum standard deviation assuming the law of one price. MINNA and MAX are the minimum and maximum standard deviation assuming no arbitrage. -I indicates that managed portfolios constructed with the instrumental variables are used. -NI indicates that no managed portfolios are used. The dashed line represents $E(m) = 0.99545$.

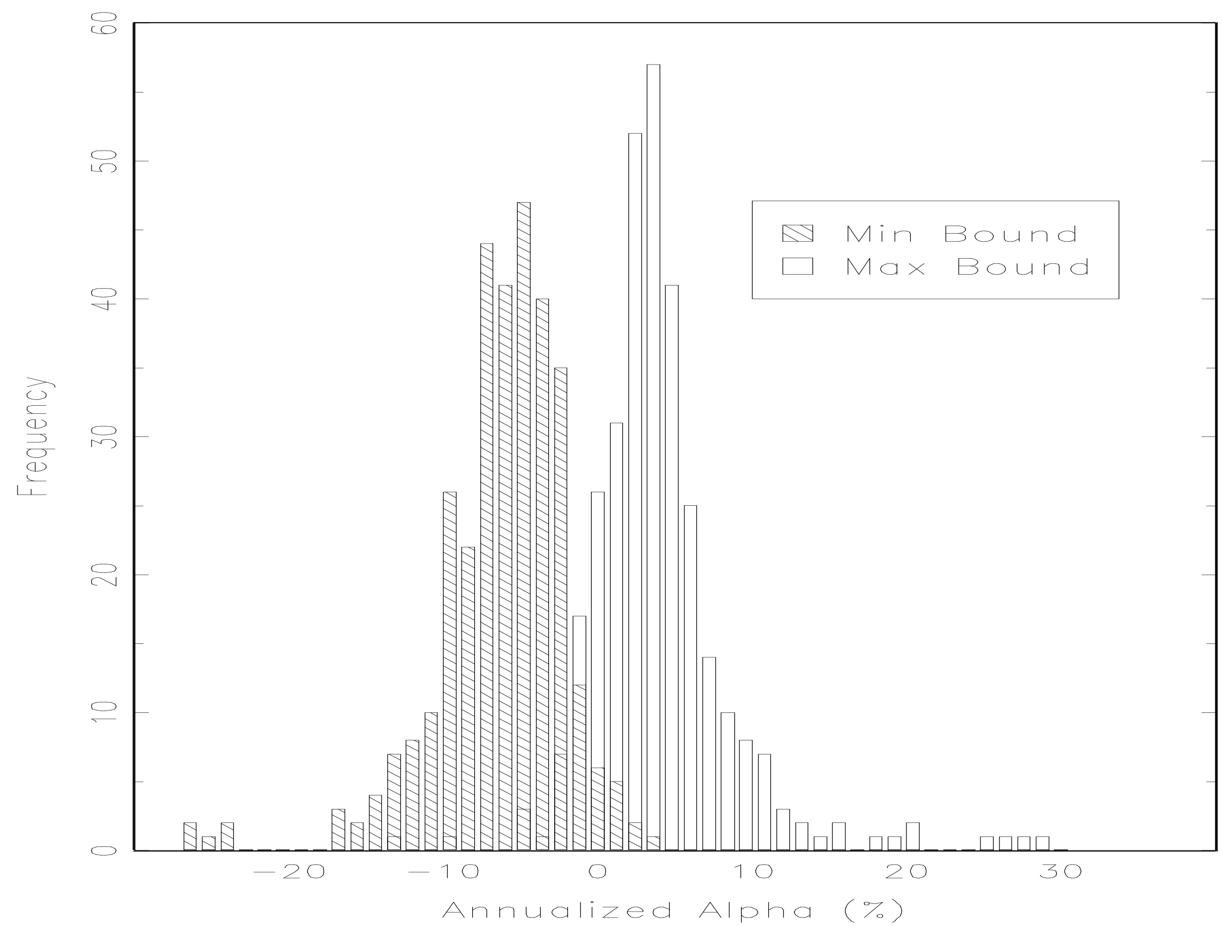

Figure 4: Histogram of the Lower and Upper Performance Measurement Bounds

Histogram of the lower (Min) and upper (Max) bounds on α_r for 320 mutual funds.

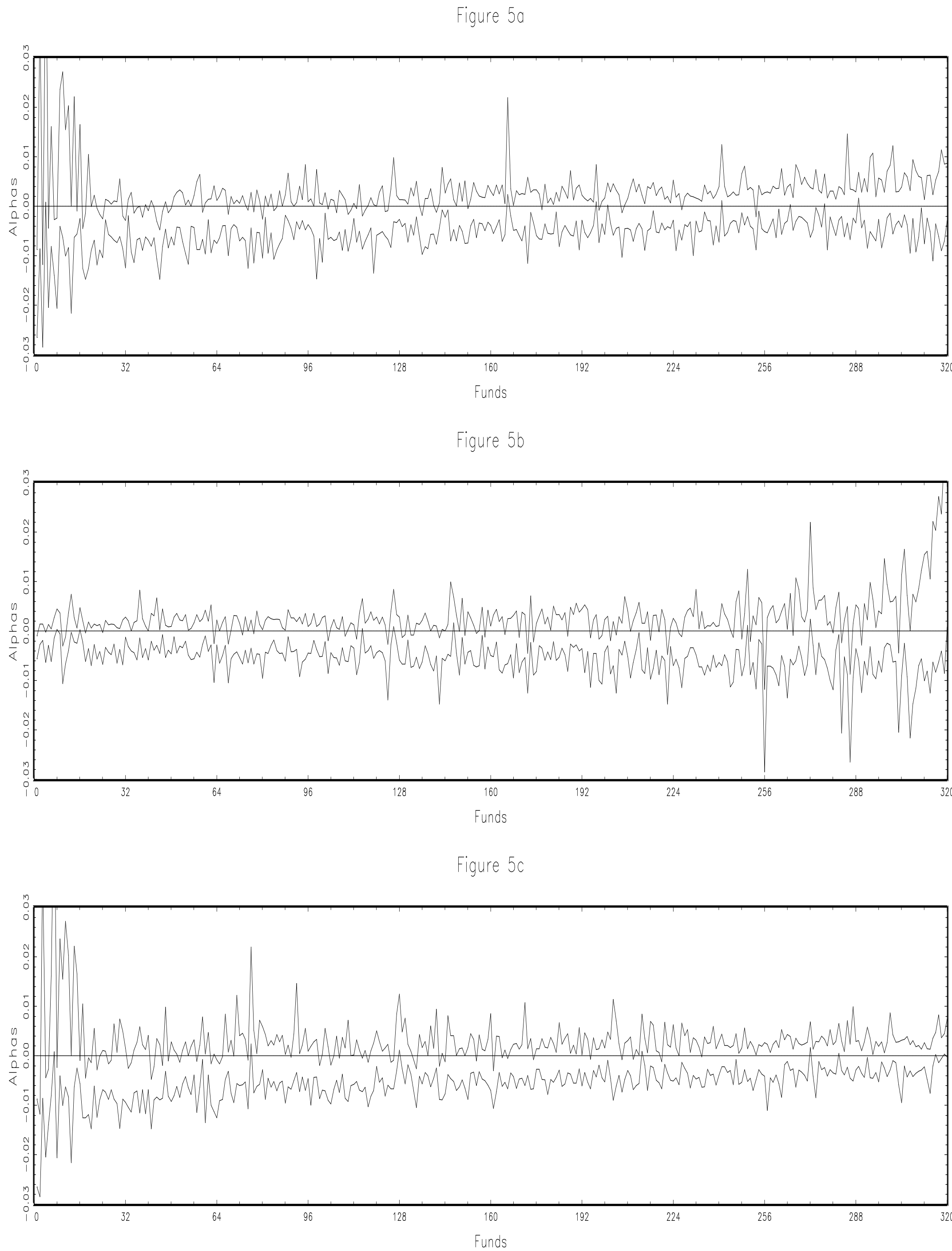

Figure 5: Lower and Upper Bounds on Performance Measurement

Lower and Upper bounds on α_r for 320 sorted mutual funds. In Figure 5a, the funds are sorted in increasing order of their average returns. In Figure 5b, the funds are sorted in increasing order of their standard deviation of returns. In Figure 5c, the funds are sorted in increasing order of their Sharpe ratio.

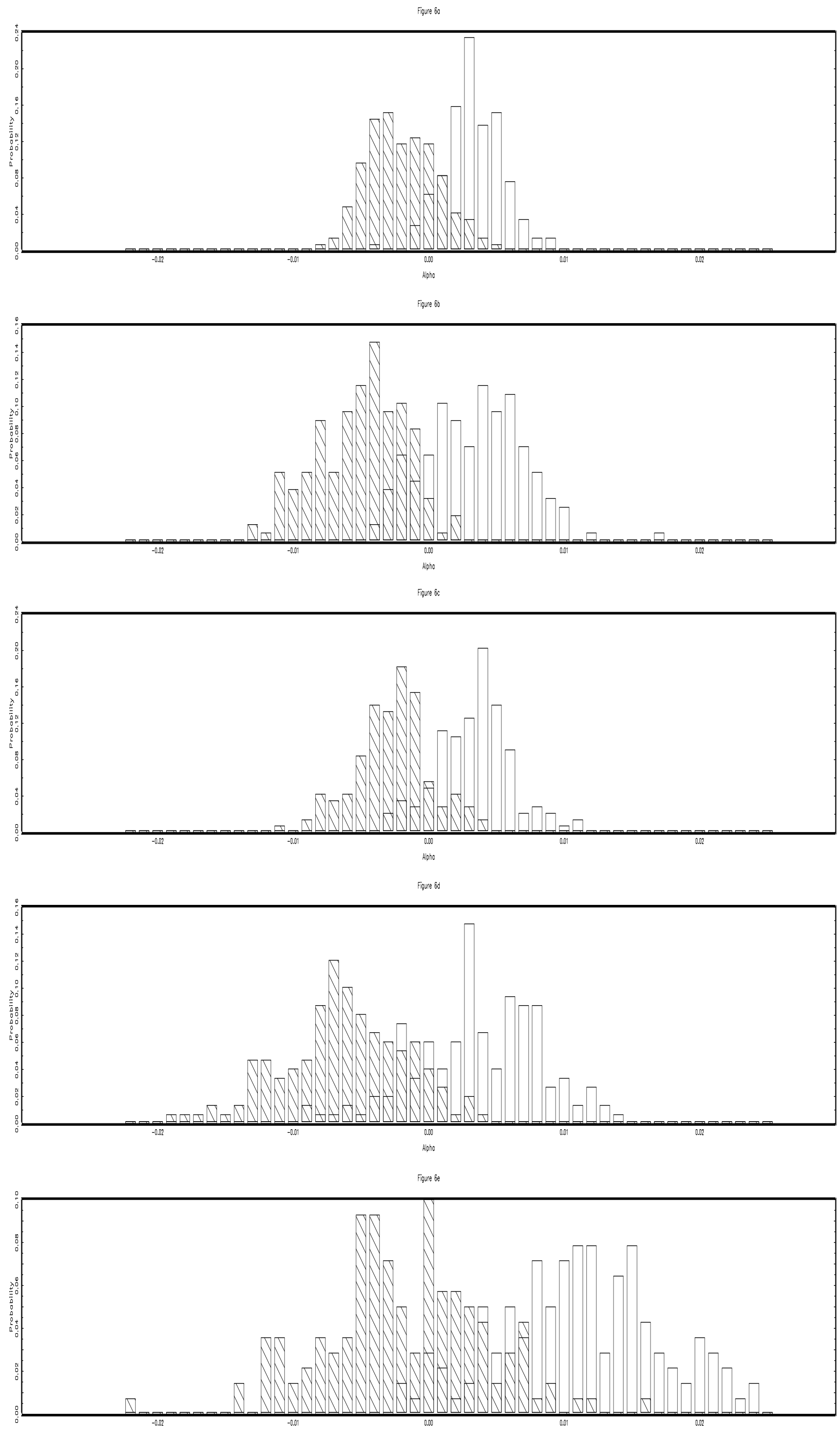

Figure 6: Empirical Distributions of the Simulated Bounds

Empirical Distributions of the simulated lower and upper bounds on α_r for 5 selected mutual funds. Figures 6a, 6b, 6c, 6d and 6e correspond respectively to the fund at the 10th, 25th, 50th, 75th and 90th percentile of the 320 funds sorted by their standard deviation of returns.