Portfolio Performance Measurement: a No Arbitrage Bounds Approach

Dong-Hyun Ahn
Department of Economics, Seoul National University, San 56-1, Sillim-Dong, Kwanak-Gu, Seoul, 151-742, Korea
E-mail: ahnd@snu.ac.kr

H. Henry Cao
Department of Finance, Cheung Kong Graduate School of Business, Oriental Plaza 3/F, Tower E3, 1 East Chang An Avenue, Beijing, PR China, 100738
E-mail: hncao@ckgsb.edu.cn

Stéphane Chrétien
Department of Finance and Insurance, Faculty of Business Administration, Laval University and CIRPÉE, Pavillon Palasis-Prince, Quebec City, QC, Canada, G1V 0A6
E-mail: stephane.chretien@fsa.ulaval.ca

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.

Special thanks to an anonymous referee, Devraj Basu, Geert Bekaert, Jacob Boudoukh, Jennifer Conrad, Bob Dittmar, Wayne Ferson, Don Fraser, Ron Gallant, Hossein Kazemi, Bob Korkie, Phil Sang Lee, Lionel Martellini (the editor), Matthew Richardson, Anil Shivdansani, Steve Slezk, Mike Stutzer, Robert Whitelaw and seminar participants at the AFA and NFA Meetings, the EFM-EDHEC Symposium on Risk and Asset Management, 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 the Institut de Finance Mathématique de Montréal and the Faculty of Business Administration at Laval University.

© 2009 The Authors
Journal compilation © 2009 Blackwell Publishing Ltd.
Keywords: portfolio performance measurement, mutual funds

JEL classification: G12, G23

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) summarise 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, emphasised 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 advocated by Long (1990). This approach is developed into and culminates in the minimum-variability SDF-based measures of Chen and Knez (1996), which are innovative since the performance measures are admissible, by construction, in the passive portfolios from which the SDF with minimum volatility is derived.

The measures of Chen and Knez (1996) are not however without limitations. The authors demonstrate that there is in general an infinite number of admissible SDFs which assign different performance measures. This ambiguity in performance measures

---


2 For performance evaluation using the CAPM, see also Roll (1978), Dybvig and Ross (1985a, 1985b) and Green (1986).
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 performance measures. Therefore, one particular choice of SDFs, like the minimum-variability SDF proposed by Chen and Knez (1996), provides only one of the infinite performance measures. It may not be admissible in a larger set of the market and may lead to inference errors in performance measurement if another choice of SDFs is more appropriate.

This paper extends the study of Chen and Knez (1996) by considering the infinite number of admissible performance measures available in an incomplete market to find the range of, or bounds on, performance values. Specifically, we show how to extract from the data the admissible SDFs that give the best and worst values to a mutual fund. We use only one minimal condition to restrict the set of admissible SDFs into a closed and convex set that guarantees the existence of the performance bounds: the no arbitrage condition, which excludes non positive SDFs. In doing so, we avoid any auxiliary assumptions inherent in existing studies (assumptions on preference systems or budget constraints in the parametric measures, or an implicit assumption in Chen and Knez (1996) to use the minimum-variability SDF). As such, our measures are free from the aforementioned potential ‘bad model’ problem innate in existing performance measures.

One can interpret different admissible SDFs as the marginal utilities of different classes of investors who invest a small amount in the managed fund. Our upper (lower) bound can be thought as the performance assessment of the fund’s most (least) favourable investor class. Intuitively, it is found by using an admissible SDF that gives high (low) marginal utilities to states where the fund returns are high and low (high) marginal utilities to states where the fund returns are low. While a positive lower bound indicates that all investors value the fund favourably, leading to a positive performance evaluation without inference error, a positive upper bound indicates that at least some investors value the fund favourably. Such finding is important as the fund usually has a target investor class in mind, with its mission to serve this particular clientele. A positive best performance confirms that the fund caters adequately to at least one clientele’s preferences.

We consider three applications of our admissible performance bounds. First, we look at performance evaluation. We assign a positive performance to a mutual fund by at least some investors if its best performance value (its upper bound) is positive, and by all investors if its worst performance value (its lower bound) is positive. Second, we develop a new methodology to rank mutual funds. We define three dominance rules that depend on the relationship between the bounds of two mutual funds to rank them. Third, we propose a diagnostic instrument to evaluate the admissibility of candidate performance measures, in the same vein that the Hansen and Jagannathan (1991) bound is used to assess the validity of candidate SDFs. We use the restriction that if candidate performance models are admissible, their performance values should be inside the performance bounds that capture an entire set of admissible measures.

To empirically illustrate the three applications, we estimate the bounds using a monthly sample of 320 equity mutual funds during the period from 1984 to 1997. This exercise

---

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.
produces a number of interesting observations and conclusions, which are robust to a small sample analysis. First, it is often not possible to sign the performance of mutual funds or rank them precisely when considering the infinite set of admissible performance measures. 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 et al. (1993), Grinblatt and Titman (1994), Ferson and Schadt (1996), Carhart (1997) and Chan et al. (2008)).

Second, we cannot rule out the possibility that more than 80% of mutual funds are evaluated positively by some investors. This observation 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 of different investor classes implicit in the infinite admissible SDFs can potentially explain the disagreement in mutual fund valuation. Third, the diagnosis of a rich menu of popular parametric models, including the CAPM and the models of Fama and French (1993) and Ferson and Schadt (1996), show that between 8% and 50% of their performance values are outside the admissible performance bounds. This result further highlights the importance of the ‘bad model’ problem in performance evaluation, complementing studies by Kothari and Warner (2001), Farnsworth et al. (2002) and Coles et al. (2006), which look at simulations for such purpose.

The rest of the paper is organised 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. 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 introduce a finite state economy in which we define and characterise the set of admissible SDFs. Then, we derive and interpret our main theoretical results, the performance evaluation bounds, and present the related literature. Next, we discuss the performance ranking of mutual funds with our performance bounds. Furthermore, we show how to obtain conditional version of the bounds. Finally, we present an extension of our analysis to a more general market economy. First, however, we explore an example, using a simple economy, to illustrate the ideas of the paper.

2.1. Example

Assume an economy with \( K = 4 \) states, \( \Omega = \{ \omega_1, \omega_2, \omega_3, \omega_4 \} \), and the corresponding probabilities \( P = (0.3 \ 0.2 \ 0.2 \ 0.3)' \). Suppose that an econometrician has \( N = 3 \) basis assets with payoffs (in gross returns) represented by the following matrix:

\[
x = \begin{pmatrix}
1.4 & 0.9 & 1.05 \\
1.2 & 1.2 & 1.05 \\
0.8 & 0.7 & 1.05 \\
0.9 & 1.3 & 1.05
\end{pmatrix}.
\]
This economy is incomplete (as \( K > N \)) and has a risk-free asset with an interest rate of 5%.

To determine the performance bounds, the first task is to determine the set of admissible SDFs. In this economy, it consists of the positive solutions for \( M = (M_1 \ M_2 \ M_3 \ M_4)' \) such that \( 1_3 = E^F[M \cdot x] \):

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

To solve this system of three equations with four unknowns, we set \( M_1 \) to an arbitrary value \( \lambda \) and find the unique solution for \( M \) given \( \lambda \). Then, we find the restrictions on \( \lambda \) that ensure that \( M \) is positive in all states, so that the no arbitrage condition holds. Simple algebra shows that

\[
M = (\lambda \ 2.8822 - 2.6842\lambda \ 1.5038 - 0.5526\lambda \ 0.2506 + 1.1579\lambda)',
\]

where \( 0 < \lambda < 1.0738 \), and the range of admissible values for the SDFs is

\[
0 < M_1 < 1.0738 \\
0 < M_2 < 2.8822 \\
0.9103 < M_3 < 1.5038 \\
0.2506 < M_4 < 1.4940.
\]

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. Following the same logic, the third state is where the investors appear the most homogenous.

Once the set of admissible SDFs is determined, the admissible performance values for a mutual fund with gross returns \( x_{mf} \) can be computed as \( \alpha_s(x_{mf}) = E^F[M \cdot x_{mf}] - 1 \). Since there is an infinite number of admissible SDFs, obtained by varying \( \lambda \), there is an infinite number of performance values. But since there are restrictions on the minimum and maximum values of \( \lambda \), there are corresponding restrictions, or bounds, on the performance values. Let’s examine the performance of three mutual funds in this incomplete economy.

First consider a mutual fund with payoffs \( x_{mf1} = (1.085 \ 1.215 \ 0.705 \ 1.165)' \). The payoffs are such that this mutual fund is a portfolio of the basis assets: \( x_{mf1} = x \theta \), where \( \theta = (0.4 \ 0.7 \ -0.1)' \). Hence, all admissible SDFs assign a unique performance value, \( \alpha_s(x_{mf1}) = E^F[M \cdot x_{mf1}] - 1 = 1 - 1 = 0 \). 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
As these problems are linear in $\lambda$, we obtain corner solutions that involve $\lambda$ 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 $x_{mf2} = (0.7 \ 1.5 \ 2.1 \ 0.3)'$. It is easy to show that $-0.2577 < \alpha_s(x_{mf2}) < 0.5188$. Some investors assign a negative performance to this mutual fund, while others assign a positive performance. Furthermore, there is an equilibrium which designates $\alpha_s(x_{mf2}) = 0$. Hence, we cannot reject the hypothesis that the performance of this fund is equivalent to the market after an adjustment for risk. Finally consider a mutual fund with payoffs $x_{mf3} = (1.3 \ 1.3 \ 0.6 \ 0.5)'$. For this fund, $-0.2479 < \alpha_s(x_{mf3}) < -0.0326$. All admissible SDFs assign a negative performance measure to this fund, indicating that it is not valuable for all classes of investors. So we can conclude that this mutual fund is underperforming. We next develop more formally the ideas presented in this example.

2.2. Finite state economy

Our setup is similar to Harrison and Kreps (1979), Hansen and Jagannathan (1991, 1997) and Chen and Knez (1995, 1996). We consider a market economy represented by a probability space triplet $(\Omega, \mathcal{F}, P)$ on which the space $L^2$ of all random variables with finite second moment is defined. We endow $L^2$ with its inner product $\langle x | y \rangle = E[P[x \cdot y]]$ for $x$ and $y \in L^2$ to make the $L^2$ 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 $x$ and prices $1_N$, where $1_N$ is an $N \times 1$ vector of ones. The set $A$ 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 $A$ 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 $(-\theta' 1_N, \theta' x) \succeq (0, 0)$,
where the inequality $\succeq$ is defined as: $x \succeq y$ 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 than 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 $A$ 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.4

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

$$M = \{ M \mid \theta' 1_N = E^P [M \cdot y] \forall y = \theta' x \in A \text{ and } M \geq 0 \},$$

where $M$ is the SDF, a market-wide random variable.

Proposition 1 states that there exists an infinite number of positive SDFs $M$ which assign
a unique price $\theta' 1_N$ to a payoff $y = \theta' x \in A$. Since the basis payoffs are gross returns,
they have unit prices by construction. Hence, the price of the synthesised 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

---

4 Proofs of propositions are made in Appendix A.
on the physical probability measure $P$. Therefore, an econometrician should investigate the viability of price system in submarket $\mathcal{A}$ first before assessing the performance of mutual funds.\(^5\)

### 2.3. 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:

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

If $x_{mf} \not\in \mathcal{A}$, there is a closed and compact interval of admissible performance values, $[\alpha_S(x_{mf}), \bar{\alpha}_S(x_{mf})]$, such that the minimum admissible value and the maximum admissible value are respectively defined as

$$\alpha_S(x_{mf}) = \inf_{M \in \mathcal{M}} E^P [M \cdot x_{mf}] - 1,$$

$$\bar{\alpha}_S(x_{mf}) = \sup_{M \in \mathcal{M}} E^P [M \cdot x_{mf}] - 1.$$ 

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. 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

\(^5\) We could proceed alternatively by finding the risk-neutral probability $Q$ which corresponds to each SDF, and then using risk-neutral pricing under $Q$. See Harrison and Kreps (1979) for the one-to-one correspondence between the SDF and the equivalent martingale measure.
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 admissible SDFs assign different admissible prices to the mutual fund’s gross return. Thereby, an infinite number of performance measures are admissible.\(^6\) 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 can be interpreted in other economically useful ways. Thinking of different SDFs as intertemporal marginal rate of substitution of different investor classes leads to one interpretation. The upper bound represents the performance value of the investor class the most favourable 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 favourable 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.

Focusing on the no arbitrage assumption leads to another way to interpret the bounds. The upper bound corresponds to the least expensive basis asset portfolio that has a payoff greater than or equal to the mutual fund payoff. The mutual fund cannot have a greater price than the cost of this least expensive portfolio as it would create an arbitrage trading strategy. Similarly, the lower bound corresponds to the most expensive basis asset portfolio that has a payoff smaller than or equal to the mutual fund payoff. This interpretation is related to the duality between admissible discount factors and efficient portfolios discussed in Hansen and Jagannathan (1991). Our bounds could be equivalently stated in terms of this dual representation.

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

\[
\alpha_r = E^P[x_{mf}] - \frac{E^P[x_{mf}]}{E^P[M \cdot x_{mf}]} = \alpha_s^P \cdot \frac{E^P[x_{mf}]}{1 + \alpha_s^P}.
\]

The performance measure \(\alpha_r\) 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).

### 2.4. Related literature

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. The NA measure proposed in Chen and Knez (1996), which we denote \(\alpha_{MINNA}^s\), is

---

\(^6\) In the terminology of Harrison and Kreps (1979), there is an infinite number of valid pricing extensions to the pricing functional which assigns a unique price to an attainable payoff \(y = \theta'x\) and a different price to other payoffs.
based on a particular choice of $M$ in $\mathcal{M}$, $M^{\text{MINNA}} = \inf_{M \in \mathcal{M}} \sqrt{E[P[M^2]]}$. Therefore, their measure is admissible, $\alpha^{\text{MINNA}}(x_{mf}) \in [\alpha_S(x_{mf}), \bar{\alpha}_S(x_{mf})]$, but it represents only one among the infinite number of admissible measures.\footnote{Notice that $M^{\text{MINNA}}$ is not identical to $M$ leading to $\alpha_S(x_{mf})$ because of the difference in underlying metrics. The former is based on the second norm whereas the latter is based on the inner product.} There is no reason why $M^{\text{MINNA}}$ should be favoured over other admissible SDFs. In fact, there is a strong likelihood that $M^{\text{MINNA}}$ may not be admissible in an extended economy. When we enlarge the set of basis assets, the set of admissible SDFs 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.\footnote{See Bekaert and Urias (1996).} Therefore, given the limited number of basis assets included in any econometric analysis, the minimum second norm SDF $M^{\text{MINNA}}$ estimated by econometricians may not be admissible in a larger economy.

Our performance bounds are similar to ‘good deal’ asset price bounds pioneered by Cochrane and Sáa-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 of admissible SDFs by imposing an upper bound on the volatility of $M$ such that

$$\sigma(M) \leq \frac{h}{R_f} \Rightarrow \bar{\sigma}(M) = \frac{h}{R_f},$$

where $h$ is the maximum Sharpe ratio (defining ‘good deals’) and $\bar{\sigma}(M)$ is the exogenously specified upper bound on the standard deviation of $M$.

The underlying motivation of Cochrane and Sáa-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 assumption. 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

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

The ‘good deal’ restriction will be binding only if $\bar{\sigma}(M) < \sigma(M^{\text{MAXNA}})$, or $h < \sigma(M^{\text{MAXNA}})R_f$. Thus, the restriction will not provide tighter bounds unless we specify a relatively low Sharpe ratio $h$.$^{10}$
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.

**Definition 2:** Consider two mutual funds: Fund A and Fund B. The corresponding performance bounds are \([α_\text{S}(x_A), \bar{α}_\text{S}(x_A)]\) and \([α_\text{S}(x_B), \bar{α}_\text{S}(x_B)]\), respectively.

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

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

  The necessary condition for this type of dominance is \(\bar{α}_\text{S}(x_A) \geq \bar{α}_\text{S}(x_B)\) and \(α_\text{S}(x_A) \geq α_\text{S}(x_B)\).

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

  \[\bar{α}_\text{S}(x_A) > \bar{α}_\text{S}(x_B)\]

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

  \[α_\text{S}(x_A) > α_\text{S}(x_B)\]

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 \(A \text{ UD}_D > B\). We can rewrite the expression for Universal

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.

Dominance as
\[
\inf_{M \in \mathcal{M}} \mathbb{E}^P[M(x_A - x_B)] = \sup_{M \in \mathcal{M}} \mathbb{E}^P[M(x_A - x_B)]
\]

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. 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 favourable 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

\[12\text{ Notice that a sufficient (but not necessary) condition for } A^{UD} > B \text{ is that the lower bound on Fund } A \text{ is greater than the upper bound of Fund } B. \text{ Thus, it is possible to establish Universal Dominance directly from the performance bounds when the lower bound of a fund is greater than the upper bound of another one.} \]
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 the performance bounds.

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.

2.6. Including conditioning information

It is possible to extend the above unconditional analysis to a conditional asset pricing framework. 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 to a conditional framework is straightforward. First, we extend our probability space triplet to \((\Omega, F, P)\), where \(F = \{F_t\}_{0 \leq t < T}\), a filtration. Then, the bounds on admissible performance measures, \([\bar{\alpha}_x(x_{mf}), \underline{\alpha}_x(x_{mf})]\), is determined by the extended set of basis assets, \(x^c = x \otimes z\), where \(z = \tilde{z}/E[\tilde{z}]\) is a standardised 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 normalised 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, \(A\), 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.

2.7. Extension to the general market economy

It is finally possible to 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 bounds, we impose an additional smoothness condition on the SDF.

Assumption 3: Let \(\omega_1, \omega_2\) 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 realised
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:

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

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

$$\bar{\alpha}_S(x_{mf}) = \inf_{M \in \mathcal{M}} E^P [M \cdot x_{mf}] - 1,$$

$$\tilde{\alpha}_S(x_{mf}) = \sup_{M \in \mathcal{M}} E^P [M \cdot x_{mf}] - 1.$$  

Moreover, the bounds are finite.

### 3. Performance Bounds: Estimation and Asymptotic Properties

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 discuss their asymptotic properties.

#### 3.1. Estimation problems and solution technique

Let $x$ be the $N$-dimensional random vector of basis asset payoffs, which can also include payoffs on dynamically managed portfolios based on information variables. Then, we can rewrite the bounds as the following problems:

**Problem 1-1: The lower bound problem**

$$\bar{\alpha}_S(x_{mf}) = \max_M E^P (M \cdot x_{mf}) - 1 \text{ s.t. } 1_N = E^P (M \cdot x); M \geq 0.$$  

**Problem 2-1: The upper bound problem**

$$\tilde{\alpha}_S(x_{mf}) = \min_M E^P (M \cdot x_{mf}) - 1 \text{ s.t. } 1_N = E^P (M \cdot x); M \geq 0.$$  

The goal of the problems is to solve for the minimum or maximum price of the mutual fund payoffs, subject to the constraints that the stochastic discount factor 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 $M = (M_1 \cdots M_K)'$. Let $D_{tk}$ denote the dummy variable such that it has value 1 if the realised state at time $t$ is $k$ and zero otherwise and $D$ denote the $K \times T$ matrix composed of $D_{tk}$. Let $x_{mf} = (x_{mf1} \cdots x_{mfT})'$, and $x = (x_1 \cdots x_T)'$. Then, we can rewrite the problems as follow:

**Problem 1-2: The lower bound problem**

$$\underline{\alpha}_S(x_{mf}) = \min_{\{M\}} \frac{1}{T} M'Dx_{mf} - 1 \text{ s.t. } 1_N = \frac{1}{T} (M'Dx)' ; M \geq 0.$$  

© 2009 The Authors
Journal compilation © 2009 Blackwell Publishing Ltd
Problem 2-2: The upper bound problem

\[ \tilde{\alpha}_S(x_{mf}) = \max_{\{M_t\}} \frac{1}{T} \mathbf{M}' \mathbf{D} x_{mf} - 1 \text{ s.t. } \mathbf{1}_N = \frac{1}{T} (\mathbf{M}' \mathbf{D})'; \mathbf{M} \geq 0. \]

Problems 1-2 and 2-2 are linear programming optimisation 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

\[ \alpha_S(x_{mf}) = \min \max_{\{M_t\}} \frac{1}{T} \mathbf{M}' \mathbf{D} x_{mf} - 1 - \lambda' [(1/T)(\mathbf{M}' \mathbf{D})' - \mathbf{1}_N] - (1/T) \delta ' \mathbf{M}, \]

with the complementary slackness conditions for all \( k \)

\[ \delta_k \geq 0 \text{ for } M_k = 0, \]

\[ \delta_k = 0 \text{ for } M_k > 0. \]

Problem 2-3: The upper bound problem

\[ \tilde{\alpha}_S(x_{mf}) = \max \max_{\{M_t\}} \frac{1}{T} \mathbf{M}' \mathbf{D} x_{mf} - 1 + \lambda' [(1/T)(\mathbf{M}' \mathbf{D})' - \mathbf{1}_N] + (1/T) \delta ' \mathbf{M}, \]

with the complementary slackness conditions for all \( t \)

\[ \delta_t \geq 0 \text{ for } M_k = 0, \]

\[ \delta_t = 0 \text{ for } M_k > 0. \]

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

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

The optimisation 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. 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).

\[ \text{See Gill et al. (1981, Section 3.3.2) for optimality conditions and sketches of proofs.} \]
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 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, $P \equiv (P_1, \ldots, P_K)$. Let $\bar{x}_k$ denote the conditional expected payoff of $x$ in state $k$ and $\bar{x} = (\bar{x}_1, \ldots, \bar{x}_K)'$ the vector of conditional payoffs. Let $b(P, \bar{x})$ denote the solution of the following problem:

$$b(P, \bar{x}_{mf}) = \min_{M, \lambda, \delta} \max_k \sum_k P_k M_k \bar{x}_{mf,k} - 1 - \lambda' \sum_k [P_k \bar{x}_k M_k - 1_N] - \sum_k P_k \delta_k M_k = 0$$

where

$$\delta_k M_k = 0,$$

for all states $k$. Since the maximand is a continuous function, $b$ 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 $\bar{x}_{mf,k}(T) = \sum_{t,k=s_t} x_{mf,t} / N_k(T)$ the sample conditional mean and $P_k(T) = N_k(T)/T$ the sample frequency of state $k$. Denote $\bar{x}_{mf}(T) = (\bar{x}_{mf,1}(T), \ldots, \bar{x}_{mf,K}(T))'$ and $P(T) = (P_1(T), \ldots, P_K(T))'$. We have

$$\alpha^*_k(\bar{x}_{mf}) = b(P(T), \bar{x}_{mf}(T))$$

As $T$ goes to infinity, by the law of large numbers, $P_k(T) \rightarrow P_k$, $\bar{x}_{mf,k}(T) \rightarrow E_k^P [x_{mf} \mid k]$. Finally, assuming that $b$ is differentiable at $(P, \bar{x}_{mf})$, the convergence rate is of the order of $1/T$ as $P(T)\bar{x}_{mf}(T)$ converges at the rate of $1/T$ 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 discretisation. Let $I_{n,j,L} = [\frac{j-1}{2^L}, \frac{j}{2^L}]$, $n = 1, \ldots, N$, $j = 0, \ldots, 4^L$. Let $k(j_1, \ldots, j_N, L) \equiv \{\omega, x_{j_n}(\omega) \in I_{n,j,L}\}$. 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 discretised economy as $\bar{x}_{mf}^*(x_{mf}, L)$ and $\alpha^*_k(x_{mf}, L)$. Then, since the set of SDFs is increasing in $L$, the upper bound $\bar{x}_{mf}^*(x_{mf}, L)$ must be an increasing function in $L$ and the lower bound $\alpha^*_k(x_{mf}, L)$ 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.
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 $y$.

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 et al. (2002) and Coles et al. (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 $\lambda^f$ 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
\[
\alpha_f^f(x_{mf}) = E^P[r_{mf} - r_f] - \beta_f'\lambda_f^f,
\]
where \( r \) denotes a simple return (a gross return \( x \) minus one), and \( \beta \) 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:
\[
r_{mf,t} - r_{f,t} = \alpha_f^f(x_{mf}) - \beta_f'\lambda_t^f + \epsilon_t,
\]
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 prespecified variables \( z = [z_1 \ldots z^Z]' \). We can express the resulting conditional performance measure as
\[
\alpha_f^{f-C}(x_{mf}) = E^P[r_{mf} - r_f | z] - \beta(z)'\lambda^f(z).
\]
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:
\[
r_{mf,t} - r_{f,t} = \alpha_f^{f-C}(x_{mf}) + b_0 \lambda_{t}^f + (b_1 z_{t-1})'\lambda_t^f + \cdots + (b_Z z_{t-1}^Z)'\lambda_t^f + \epsilon_t,
\]
with \( E^P[\epsilon_t | z_{t-1}] = E^P[\epsilon_t \lambda_t^f | z_{t-1}] = 0 \).

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

- **The CAPM**: \( \lambda_{t}^{CAPM} = r_{m}^t - r_{f,t} \)
  where \( r_{m} \) is the return on the market portfolio.
- **The Fama-French Model**: \( \lambda_{t}^{FFM} = [r_{m}^t - r_{f,t} \quad r_{smb}^t \quad r_{hml}^t]' \)
  where \( r_{smb} \) and \( r_{hml} \) are returns on the mimicking portfolios of size and book-to-market respectively.
- **The Ferson-Schadt Model**: \( \lambda_{t}^{FSM} = [r_{ls}^t - r_{f,t} \quad r_{ss}^t - r_{f,t} \quad r_{lgb}^t - r_{f,t} \quad r_{lgcb}^t - r_{f,t}]' \)
  where \( r_{ls} \), \( r_{ss} \), \( r_{lgb} \) 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 criticised 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:
\[
M_{t+1}^C = \beta \frac{u'(C_{t+1}; \theta)}{u'(C_t; \theta)} \frac{P_{L_t}}{P_{L_{t+1}}},
\]
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 \( \theta \), and \( P_{L_t} \) is the price level at time \( t \). The performance measure implied by the consumption-based models is given by
\[
\alpha_s^C(x_{mf}) = E^P[M^C \cdot x_{mf}] - 1.
\]
We estimate the performance measure of the consumption-based model by following a two-step approach. In the first step, using the generalised method of moment procedure of Hansen (1982), we estimate the parameters $\theta$ by minimising a quadratic form of the average pricing error on the benchmark assets:

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

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

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

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

$$M_{t+1}^{\text{POWER}} = \beta \left(\frac{C_{t+1}}{C_t} - \theta C_{t-1} \right)^{-\gamma} \frac{P_L_t}{P_L_{t+1}}.$$

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

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

### 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

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

where $M^{\text{MINLOP}}$ solves the following problem:

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

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

$$\alpha_S^{\text{MINLOP}}(x_{mf}) = E^P [x(E^P[xx'])^{-1} 1_N \cdot x_{mf}] - 1,$$

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

$$\alpha_S^{\text{MINNA}}(x_{mf}) = E^P [M^{\text{MINNA}} \cdot x_{mf}] - 1.$$
where $M^{\text{MINNA}}$ solves the following problem:

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

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

As discussed previously, $M^{\text{MINNA}}$ 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:

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

where $M^{\text{MAXNA}}$ solves the following problem:

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

Again, this model is estimated numerically using its sample counterpart.\(^{14}\)

Notice that $\alpha^{\text{MINNA}}$ and $\alpha^{\text{MAXNA}}$ 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.\(^{15}\) 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 US equity market.\(^{16}\) 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 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.\(^ {17}\)

\(^{14}\) 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 optimisation techniques available.

\(^{15}\) 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.

\(^{16}\) Our sample contains one mutual fund specialised in real estate investments.

\(^{17}\) Brown et al. (1992), Brown and Goetzmann (1995), Malkiel (1995) and Elton et al. (1996) discuss the upward bias created when measuring the performance of portfolios of surviving
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.

### Table 1: Summary statistics for the mutual funds

Panel A: Individual Mutual Funds

<table>
<thead>
<tr>
<th>Statistics</th>
<th>Ret Avg</th>
<th>Std Dev</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Sharpe</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>1.01164</td>
<td>0.04613</td>
<td>0.77612</td>
<td>1.14170</td>
<td>0.16475</td>
</tr>
<tr>
<td>Std Dev</td>
<td>0.00307</td>
<td>0.01277</td>
<td>0.05650</td>
<td>0.04770</td>
<td>0.06679</td>
</tr>
<tr>
<td>Min</td>
<td>0.99236</td>
<td>0.01805</td>
<td>0.50379</td>
<td>1.05829</td>
<td>-0.20499</td>
</tr>
<tr>
<td>1%</td>
<td>0.99879</td>
<td>0.02484</td>
<td>0.65048</td>
<td>1.07285</td>
<td>-0.0938</td>
</tr>
<tr>
<td>5%</td>
<td>1.00663</td>
<td>0.03006</td>
<td>0.68998</td>
<td>1.09102</td>
<td>0.03885</td>
</tr>
<tr>
<td>10%</td>
<td>1.00911</td>
<td>0.03387</td>
<td>0.70865</td>
<td>1.09933</td>
<td>0.09585</td>
</tr>
<tr>
<td>25%</td>
<td>1.01082</td>
<td>0.03867</td>
<td>0.74024</td>
<td>1.11570</td>
<td>0.13945</td>
</tr>
<tr>
<td>Median</td>
<td>1.01215</td>
<td>0.04378</td>
<td>0.77892</td>
<td>1.13438</td>
<td>0.17357</td>
</tr>
<tr>
<td>75%</td>
<td>1.01318</td>
<td>0.05057</td>
<td>0.80776</td>
<td>1.15263</td>
<td>0.20777</td>
</tr>
<tr>
<td>90%</td>
<td>1.01405</td>
<td>0.06020</td>
<td>0.84998</td>
<td>1.17894</td>
<td>0.23032</td>
</tr>
<tr>
<td>95%</td>
<td>1.01510</td>
<td>0.07020</td>
<td>0.87608</td>
<td>1.22723</td>
<td>0.24209</td>
</tr>
<tr>
<td>99%</td>
<td>1.01724</td>
<td>0.08680</td>
<td>0.91148</td>
<td>1.33951</td>
<td>0.29832</td>
</tr>
<tr>
<td>Max</td>
<td>1.01802</td>
<td>0.12761</td>
<td>0.94443</td>
<td>1.45161</td>
<td>0.31546</td>
</tr>
</tbody>
</table>

Panel B: Investment Objective Portfolios

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

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 annualised numbers, the mutual funds have a mean return of 13.97%, with a mean standard deviation of 0.17357.

In cases where survivorship bias might be an issue, we will use estimates provided by Elton et al. (1996) to examine its effect on our results.
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 specialising in financials provide the best average return and Sharpe ratio, while the mutual funds specialising 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 annualised 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 specialise in a specific sector, like the specialty funds included in our sample. Chen and Knez (1996) and Dahlquist and S"oderlind (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 Portfolios Returns File. The annualised 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 annualised 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.\footnote{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}
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.

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

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 Sáa-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.
5.3. Variables for the parametric models

To implement the linear factor models, we need proxies for the market prices of risk $\lambda_t$. 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) US 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 et al. (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 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 annualised 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.
Fig. 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 $E(m) = 0.99545$.

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.\(^\text{19}\) First, 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

\(^\text{19}\) See Gallant et al. (1990), Ferson and Siegel (2003), Bekaert and Liu (2004) and Abhyankar et al. (2007) for related discussions.
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 annualised 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 et al. (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 summarises 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 annualised \(\alpha_r\). 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 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

---

\(^{20}\) 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.

\(^{21}\) 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.
This table presents the lower and upper performance bounds on mutual funds 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**

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

**Panel B: Investment Objective Portfolios**

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

The 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
Fig. 5. Lower and upper bounds on performance measurement

Lower and Upper performance bounds 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.

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, 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.
Empirical distributions of the simulated bounds 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.

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 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: the specialty portfolio, the natural resources portfolio, the utilities portfolio and the real estate portfolio. The performances of the other 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 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

23 We might expect that the performance of the objective portfolios is too optimistic because only surviving funds are included in each portfolio. Elton et al. (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 2.12 to 5.99 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.
Table 4
Statistical significance of the performance measurement bounds

This table examines the statistical significance of the lower and upper performance bounds. 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

<table>
<thead>
<tr>
<th>Classification</th>
<th>For Significance Level of</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>5%</td>
</tr>
<tr>
<td>( LB &gt; 0 + UB &gt; 0 )</td>
<td>0</td>
</tr>
<tr>
<td>( LB &gt; 0 + UB &lt; 0 )</td>
<td>1</td>
</tr>
<tr>
<td>( LB &lt; 0 + UB &gt; 0 )</td>
<td>0</td>
</tr>
<tr>
<td>( LB &lt; 0 + UB = 0 )</td>
<td>49</td>
</tr>
<tr>
<td>( LB = 0 + UB &gt; 0 )</td>
<td>33</td>
</tr>
<tr>
<td>( LB = 0 + UB = 0 )</td>
<td>237</td>
</tr>
</tbody>
</table>

### Panel B: Investment Objective Portfolios

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

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 bounds to make new observations on two 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. Thus, there exists at least a SDF that values 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 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 USA as of October 2007. In our sample, although Fidelity Magellan has a slightly higher average return, Vanguard 500 Index presents 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 favouring 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 et al. (1993), Grinblatt and Titman (1994), Ferson and Schadt (1996), Carhart (1997), Gruber (2001), and Chan et al. (2008)).

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

---

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

<table>
<thead>
<tr>
<th>Objectives</th>
<th>Best Rank</th>
<th>Worst Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aggressive Growth</td>
<td>1</td>
<td>12</td>
</tr>
<tr>
<td>Growth</td>
<td>1</td>
<td>12</td>
</tr>
<tr>
<td>Growth and Income</td>
<td>1</td>
<td>12</td>
</tr>
<tr>
<td>Equity Income</td>
<td>1</td>
<td>12</td>
</tr>
<tr>
<td>Small Company</td>
<td>1</td>
<td>12</td>
</tr>
<tr>
<td>Specialty</td>
<td>2</td>
<td>14</td>
</tr>
<tr>
<td>Specialty–Health</td>
<td>1</td>
<td>10</td>
</tr>
<tr>
<td>Specialty–Financial</td>
<td>1</td>
<td>14</td>
</tr>
<tr>
<td>Specialty–Nat Res</td>
<td>10</td>
<td>14</td>
</tr>
<tr>
<td>Specialty–Prec Metals</td>
<td>1</td>
<td>12</td>
</tr>
<tr>
<td>Specialty–Technology</td>
<td>1</td>
<td>14</td>
</tr>
<tr>
<td>Specialty–Utilities</td>
<td>2</td>
<td>12</td>
</tr>
<tr>
<td>Specialty–Real Estate</td>
<td>10</td>
<td>14</td>
</tr>
<tr>
<td>Specialty–Comm</td>
<td>1</td>
<td>12</td>
</tr>
</tbody>
</table>

**Panel B: Scenario Dominance**

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

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
Dowmiminate 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 favourable 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. 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.
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

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.
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 below the lower bound (% < *LB*) or above the upper bound (% > *UB*). 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 risk-free 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

<table>
<thead>
<tr>
<th>Statistics</th>
<th>CAPM</th>
<th>CAPM-C</th>
<th>FFM</th>
<th>FFM-C</th>
<th>FSM</th>
<th>FSM-C</th>
<th>POWER</th>
<th>HABIT</th>
<th>MINLOP</th>
<th>MINNA</th>
<th>MAXNA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>−0.0012</td>
<td>−0.0011</td>
<td>−0.0002</td>
<td>−0.0002</td>
<td>0.0008</td>
<td>0.0008</td>
<td>0.0019</td>
<td>−0.0012</td>
<td>−0.0013</td>
<td>−0.0023</td>
<td></td>
</tr>
<tr>
<td>Std Dev</td>
<td>0.0028</td>
<td>0.0027</td>
<td>0.0028</td>
<td>0.0027</td>
<td>0.0025</td>
<td>0.0028</td>
<td>0.0030</td>
<td>0.0029</td>
<td>0.0034</td>
<td>0.0053</td>
<td></td>
</tr>
<tr>
<td><em>t</em>-stat</td>
<td>(−4.76)</td>
<td>(−4.70 )</td>
<td>(−0.95 )</td>
<td>(−1.01 )</td>
<td>(3.72)</td>
<td>(3.18)</td>
<td>(7.16)</td>
<td>(4.85)</td>
<td>(−5.05 )</td>
<td>(−4.34 )</td>
<td>(−5.03 )</td>
</tr>
<tr>
<td>Min</td>
<td>−0.0217</td>
<td>−0.0227</td>
<td>−0.0205</td>
<td>−0.0224</td>
<td>−0.0187</td>
<td>−0.0202</td>
<td>−0.0174</td>
<td>−0.0179</td>
<td>−0.0171</td>
<td>−0.0183</td>
<td>−0.0259</td>
</tr>
<tr>
<td>1%</td>
<td>−0.0116</td>
<td>−0.0114</td>
<td>−0.0116</td>
<td>−0.0103</td>
<td>−0.0093</td>
<td>−0.0118</td>
<td>−0.0106</td>
<td>−0.0103</td>
<td>−0.0096</td>
<td>−0.0114</td>
<td>−0.0134</td>
</tr>
<tr>
<td>5%</td>
<td>−0.0054</td>
<td>−0.0050</td>
<td>−0.0044</td>
<td>−0.0034</td>
<td>−0.0024</td>
<td>−0.0029</td>
<td>−0.0031</td>
<td>−0.0037</td>
<td>−0.0050</td>
<td>−0.0057</td>
<td>−0.0083</td>
</tr>
<tr>
<td>10%</td>
<td>−0.0041</td>
<td>−0.0035</td>
<td>−0.0026</td>
<td>−0.0021</td>
<td>−0.0012</td>
<td>−0.0017</td>
<td>−0.0007</td>
<td>−0.0013</td>
<td>−0.0040</td>
<td>−0.0047</td>
<td>−0.0069</td>
</tr>
<tr>
<td>25%</td>
<td>−0.0022</td>
<td>−0.0022</td>
<td>−0.0012</td>
<td>−0.0011</td>
<td>0.0001</td>
<td>0.0011</td>
<td>0.0004</td>
<td>0.0023</td>
<td>−0.0029</td>
<td>−0.0045</td>
<td></td>
</tr>
<tr>
<td>Median</td>
<td>−0.0008</td>
<td>−0.0008</td>
<td>0.0000</td>
<td>−0.0001</td>
<td>0.0011</td>
<td>0.0012</td>
<td>0.0024</td>
<td>0.0017</td>
<td>−0.0010</td>
<td>−0.0011</td>
<td>−0.0026</td>
</tr>
<tr>
<td>75%</td>
<td>0.0006</td>
<td>0.0005</td>
<td>0.0011</td>
<td>0.0009</td>
<td>0.0022</td>
<td>0.0021</td>
<td>0.0034</td>
<td>0.0026</td>
<td>0.0003</td>
<td>0.0002</td>
<td>−0.0005</td>
</tr>
<tr>
<td>90%</td>
<td>0.0014</td>
<td>0.0015</td>
<td>0.0023</td>
<td>0.0023</td>
<td>0.0028</td>
<td>0.0033</td>
<td>0.0043</td>
<td>0.0036</td>
<td>0.0015</td>
<td>0.0018</td>
<td>0.0014</td>
</tr>
<tr>
<td>95%</td>
<td>0.0020</td>
<td>0.0021</td>
<td>0.0034</td>
<td>0.0031</td>
<td>0.0035</td>
<td>0.0039</td>
<td>0.0053</td>
<td>0.0045</td>
<td>0.0026</td>
<td>0.0033</td>
<td>0.0039</td>
</tr>
<tr>
<td>99%</td>
<td>0.0042</td>
<td>0.0039</td>
<td>0.0063</td>
<td>0.0053</td>
<td>0.0047</td>
<td>0.0052</td>
<td>0.0075</td>
<td>0.0066</td>
<td>0.0050</td>
<td>0.0075</td>
<td>0.0178</td>
</tr>
<tr>
<td>Max</td>
<td>0.0046</td>
<td>0.0044</td>
<td>0.0073</td>
<td>0.0067</td>
<td>0.0061</td>
<td>0.0066</td>
<td>0.0083</td>
<td>0.0075</td>
<td>0.0109</td>
<td>0.0213</td>
<td>0.0397</td>
</tr>
</tbody>
</table>

### Panel B: Comparison with the Performance Measurement Bounds

<table>
<thead>
<tr>
<th>Statistics</th>
<th>CAPM</th>
<th>CAPM-C</th>
<th>FFM</th>
<th>FFM-C</th>
<th>FSM</th>
<th>FSM-C</th>
<th>POWER</th>
<th>HABIT</th>
<th>MINLOP</th>
<th>MINNA</th>
<th>MAXNA</th>
</tr>
</thead>
<tbody>
<tr>
<td>% &lt; <em>LB</em></td>
<td>2.19</td>
<td>2.50</td>
<td>1.25</td>
<td>1</td>
<td>0.63</td>
<td>1.25</td>
<td>1.56</td>
<td>1.56</td>
<td>0.31</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>% &gt; <em>UB</em></td>
<td>6.56</td>
<td>8.44</td>
<td>9.38</td>
<td>7.50</td>
<td>27.19</td>
<td>31.25</td>
<td>50.63</td>
<td>36.56</td>
<td>0.31</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
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 overestimate 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 SDF and secondly verify the existence of an infinite number of such SDFs. Denote the set of arbitrage trading opportunities by \( \mathcal{H} = \{(-cy, y) \geq^+ (0, 0)\} \), which is a subspace of \( \mathbb{R} \times L^2 \), where \( cy \) denotes the cost of trading to get the random payoff \( y \). The premise of viable price system means that \( (\mathbb{R} \times \mathcal{A}) \cap \mathcal{H} = \emptyset \). The Separating Hyperplane Theorem implies the existence of continuous linear functions \( f: \mathbb{R} \times L^2 \to \mathbb{R} \) such that \( f = 0 \) if \( (-cy, y) \in (\mathbb{R}_- \times \mathcal{A}) \) and \( f > 0 \) if \( (-cy, y) \in \mathcal{H} \) since both \( \mathcal{A} \) and \( \mathcal{H} \) are closed and convex sets. Let \( f(-cy, y) = -cy + E^p[M \cdot y] \), after normalisation. Since the basis assets are achievable,

\[
 f(-1_N, x) = -1_N + E^p[M \cdot x] = 0.
\]

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

\[
 f(0, y) = 0 + E^p[M \cdot y] > 0.
\]

The fact that the above inequality holds for any \( y \geq^+ 0 \) yields \( M > 0 \). Finally, since the market is incomplete, the prices of additional assets with payoffs not spanned by the basis assets will not be uniquely determined. Thus, there must exist more than two positive SDFs. Since any convex combination of SDFs is a SDF, there exists an infinite number of SDFs. \( \square \)

Proof of Proposition 2: Since the payoffs are all positive and there are only finite states, the SDFs are bounded from above and below and are thus in a closed compact set. As a result, the upper bound and lower bound exist and are attainable. Moreover,

\[26\] 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.
since SDFs 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, $E^p[x_{mf}^2] < \infty$, $\forall j$. Let $\omega_0$ be an arbitrary state. We have:

$$E^p[M(\omega)^2] = E^p[(M(\omega) - M(\omega_0))^2 + 2M(\omega)M(\omega_0) - M(\omega_0)^2]$$

$$= 2M(\omega_0)/R_f - M(\omega_0)^2 + \int_\omega (M(\omega) - M(\omega_0))^2 dP(\omega)$$

$$\leq 2M(\omega_0)/R_f - M(\omega_0)^2 + B^2 \sum_{j} x_j(\omega) - x_j(\omega_0)^2 dP(\omega)$$

$$= 2M(\omega_0)/R_f - M(\omega_0)^2 + B^2 \sum_{j} E^p[(x_j(\omega) - x_j(\omega_0))^2].$$

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

$$E^p[M(\omega)^2] \leq 2/R_f^2 - E^p[M(\omega_0)^2] + B^2 \sum_{j} E^p[(x_j(\omega) - x_j(\omega_0))^2].$$

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

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

which implies that

$$E^p[M(\omega)^2] \leq 1/R_f^2 + \frac{1}{2} B^2 \sum_{j} \sigma_j^2,$$

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

$$-\tilde{U} \sqrt{E^p[x_{mf}^2]} \leq \alpha_S(x_{mf}) \leq \tilde{U} \sqrt{E^p[x_{mf}^2]}.$$

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

**Appendix B: Two-Phase Revised Simplex Method**

The simplex method is a ‘feasible-point’ method: given initial feasible points $M_0$, all subsequent iterates $M_k$ 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. Readers interested in more details and references can consult Gill et al. (1981, hereafter GMW).

In phase one, the procedure finds a basic initial feasible point $M_0$ to the problem. The technique to find such a point is called phase 1 simplex (see GMW, Section 5.7).
It considers an artificial linear objective function made of the sum of infeasibilities (the violated constraints) at \( M \). A feasible point is then found by minimising 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 (see GMW, Sections 5.3.1 and 5.6.1). Let the working set of constraints be the \( T \) binding constraints in current iteration, and let \( M_k \) denote the current iterate. \( M_k \) 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 \( \delta_k \) when \( M_k = 0 \). If \( \delta_k > 0 \) when \( M_k = 0 \), then the third necessary and sufficient conditions is met and \( M_k = M^* \). However, if any \( \delta_k \) 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 factorisation (see GMW, Section 2.2.5.1) 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 factorisation is known as the Bartels-Golub scheme, which ensure numerical stability by using row interchanges during the updating.

References


