A Bayesian approach to diagnosis of asset pricing models

Michael Stutzer

Department of Finance, University of Minnesota, Minneapolis, MN 55455, USA

(Received July 1992; final version received May 1994)

Abstract

A large literature has arisen which exploits a particular portfolio on the mean–variance frontier, determining a minimum variance bound on the set of stochastic discount factors (state price to probability ratios). This paper proposes a new variational characterization of the closely related set of state price densities, based on minimization of the Kullback–Leibler Information Criterion. In contrast to the variance bound, the resulting information bound automatically satisfies an important positivity constraint. Furthermore, the information bound is determined by a portfolio which maximizes expected CARA utility. Several interpretations of the information bound are given, and empirical uses of it are illustrated.

Key words: Information-theoretic statistics; Asset pricing

JEL classification: C11; C14; G11; G12

1. Introduction

Arbitrage-free asset pricing models imply the existence of a convex set of pricing operators, sometimes called stochastic discount factors (SDFs) (Hansen and Jagannathan, 1991a). In environments detailed in Hansen and Richard (1987), variational characterizations of these sets may be used to diagnose the
nature of specific pricing models errors, as well as to aid in the construction and testing of other asset pricing theories. For example, a large literature has arisen which exploits a (Hansen–Jagannathan) minimum variance bound characterization of SDFs. This characterization identifies a variance minimizing benchmark SDF, whose variance places a lower bound on the variances of other SDFs. To provide motivation for the general reader, Section 2 describes the variance bound, highlights three interpretations of it, and provides an illustrative application.

Such variational characterizations of SDFs are natural to economists proposing utility-maximizing models of consumption and investment decisions, in which the intertemporal marginal rates of substitution (IMRS) of agents are SDFs. But variational characterizations of the closely related state price probability densities (SPDs) are also quite natural to financial theorists proposing contingent claims pricing models, in which the risk-neutral probabilities are SPDs. With no less generality than required for the minimum variance bound characterization of SDFs, a new minimum information bound characterization of SPDs is developed in Section 3. The information bound minimizes the Kullback–Leibler Information Criterion (KLIC), a fundamental tool of information theory and many of its econometric uses. Many appealing statistical and economic interpretations of the information bound are provided. For example, some uses of the variance bound are based on the fact that it also determines a mean–variance efficient portfolio. In the absence of distributional restrictions (e.g., multivariate normality) that discrete-time variational characterizations were meant to avoid, emphasis on mean–variance analysis rests on the assumption that investors possess quadratic utility functions. In contrast, the information bound will be seen to determine a portfolio which is optimal for investors possessing constant absolute risk aversion (CARA) utility. CARA utility does not have quadratic utility’s undesirable properties of satiation and increasing absolute risk aversion. In addition, it is shown that the information bound provides a ‘utility-based metric’ (Rossi and McCulloch, 1990) for evaluating the economic significance of changes in the information bound induced by different sets of assets or different time periods.

In Section 4, the information bound is compared and contrasted to the variance bound. The information bound is shown to provide a simple, economically interpretable way of incorporating the positivity restriction on SPDs implied by nonsatiation in both arbitrage-free consumption-based and contingent-claims pricing models. In an illustrative empirical application, the

---

1 Starting with Hansen and Jagannathan (1991a), there has been an explosion of papers which have developed and/or applied the diagnostics, including Backus et al. (1991), Becket and Hodrick (1992), Burnside (1991), Cecchetti et al. (1992), Chen and Knez (1992a, b), Cochrane and Hansen (1992), DeSantis (1993a, b), Epstein and Zin (1991), Ferson and Harvey (1992), He and Modest (1992), Luttmer (1991), and Snow (1991).
The aforementioned optimal portfolio interpretation is used in an examination of 'small-firm effects'.

We now turn to a review of SDFs and the variance bound characterization of them, to motivate readers unfamiliar with the literature as well as to establish aspects of it which are focused on later.

2. Stochastic discount factors and the variance bound: Review

The object under examination is the set of stochastic discount factors (SDFs) \( m \), defined by the following conditions in environments described in Hansen and Richard (1987):

\[
E[X_{i}^{t+1}m^{t+1}|I_t] = P^t_i, \quad i = 1, \ldots, N, \tag{1}
\]

where \( P^t_i \) denotes the price of risky asset \( i \) at time \( t \) and \( X_{i}^{t+1} \) denotes its real payoff the next period. Expectations are taken conditional on information \( I_t \) at time \( t \). If an observable riskless asset exists, I will add a payoff \( X_{r}^{t+1} = 1 \) to denote its end-of-period real payoff in all states, use \( P^t_0 \) to denote its price at \( t \), and add a corresponding equation to (1). The \( N \) assets in (1) are of interest to an analyst seeking to evaluate and/or construct theories intended to price those assets.

In dynamic consumption-portfolio choice problems, an agent's intertemporal marginal rate of substitution (IMRS) of consumption between period \( t \) and \( t + 1 \) must also satisfy (1) (see, e.g., Bansal and Viswanathan, 1993, p. 1234). The assumption of nonsatiation, consistent with most any dynamic utility-maximizing theory, implies the existence of positive SDFs. More generally, the absence of arbitrage opportunities implies that there is a convex set of positive SDFs (which may be just one element). The implications of this additional positivity constraint will be explored in Section 4.1.

Now divide the equations in (1) by their respective right-hand sides and apply the law of iterated expectations to produce the unconditional moment conditions:

\[
E[R_i m] = \int R_i m \, d\mu = 1, \quad i = 1, \ldots, N. \tag{2}
\]

where \( R_i \) denotes the gross real return from asset \( i \) and \( \mu \) is the state probability measure.

While (2) is implied by (1), the former implies more than the latter. Without making any assumptions about the nature and form of the conditional distributions, how can conditioning information be incorporated? The way in which variational characterizations handle this is by considering investors' typical use of conditioning information. Investors use information to conduct dynamic trading strategies, i.e., to manage portfolios of the assets. Consider a portfolio strategy investing \( z^t_i \) units in asset \( i, i = 1, \ldots, N, t = 1, 2, \ldots \) When the vector
\( z^t \) is constructed using only the information known at \( t \), the portfolio’s time-varying payoff \( X^{t+1}_z = \sum_i z_i X_i^{t+1} \) is adapted to the information structure. The value of the portfolio at time \( t \) is \( V_t^z = \sum_i z_i P_i^t \), which is known at time \( t \). Multiplying both sides of (1) by \( z_t^i \) and summing over \( i \) yields \( V_t^z = \mathbb{E}[X^{t+1}_z | I_t] \). Dividing both sides by \( V_t^z \) and denoting the portfolio’s gross return by \( R^{t+1}_z = X^{t+1}_z / V_t^z \) yields the following conditional moment condition for each adapted strategy \( z \):

\[
\mathbb{E}[R^{t+1}_z m^{t+1} | I_t] = 1. \tag{3}
\]

Take the unconditional expectation of (3) to produce the unconditional moment conditions:

\[
\mathbb{E}[R_z m] = 1, \quad \text{all adapted } z. \tag{4}
\]

Of course, what was once only \( N \) moment conditions (2) potentially become an infinity of conditions (4), one for each adapted strategy. And assuming that analysts observe less information than investors do, analysts could not specify all possible \( z \) anyway. As a result, it is possible to investigate the implications of only a subset of the possible conditions (4). In any event, (4) permits us to reinterpret the return from a managed portfolio as a member of the set of returns satisfying (2), so we need only refer to (2) in what follows.\(^2\)

### 2.1. The affine SDF and its variance bound

There are several ways to use measured asset returns to provide a variational characterization useful in the construction and testing of asset pricing theories. The most widely used way is find an SDF which has the minimum variance among those satisfying (2) and which have a fixed mean, denoted by \( c \). One motive for fixing the mean is to investigate the relationship between the variance bound and the mean price of a riskless asset with unit real payoff \( X_0 = 1 \), and possibly time-varying gross return (i.e., gross real interest rate) \( r \) and time varying price \( 1/r \). To see this, substitute \( X_0 = 1 \) into (1) and take the unconditional expectation to obtain

\[
\mathbb{E}[m] = \mathbb{E}[1/r] = c. \tag{5}
\]

For our purposes it will make sense to investigate a fixed mean \( c \) even when there is no riskless asset. To find a bound satisfying (2) and (5), aspects of the

\(^2\) Actually, there are more general implications of (1), which are not as economically interpretable. For example, with any adapted variable \( y \) and return \( R_y \), similar steps yield the implication that \( \mathbb{E}[y R_y / \mathbb{E}[y] m] = 1 \), where \( y R_y / \mathbb{E}[y] \) is a ‘pseudo-return’. As in the case of a managed portfolio return, a pseudo-return may also be treated as a member of the set of returns satisfying (2).
compact derivations in Ferson and Harvey (1992) and Cochrane (1992) are followed. Consider an affine combination of the returns in excess of their means:

\[ m^a = (R - E[R])'w^a + c, \tag{6} \]

where \( w^a \) is a (column) vector of coefficients. Note that this always satisfies (5). Substituting into (2) and factoring out \( w^a \) yields the linear equation system

\[ E[R(R - E[R])']w^a = 1 - cE[R], \]

where 1 denotes a vector of ones. When there is a solution to (2) of the form (6), we may solve this linear equation system to produce the unique solution:

\[ w^a = \text{cov}[R]^{-1}(1 - cE[R]) = -\text{cov}[R]^{-1}(cE[R] - 1), \tag{7} \]

where \( \text{cov}[R] \) is the covariance matrix \( E[RR'] - E[R]E[R]' \) of the returns vector. The affine benchmark \( m^a \) is produced by substituting (7) into (6).

2.2. Interpretation of the affine benchmark

Hansen and Jagannathan (1991a, b) give several interpretations of the affine benchmark of interest to analysts. Three are of interest here. First, it has the minimum variance among those \( m \) satisfying (2) and (5), to wit:

\[ \text{var}[m^a] = w^a \text{cov}[R]w^a \\
= (cE[R] - 1)'\text{cov}[R]^{-1}(cE[R] - 1) \leq \text{var}[m]. \tag{8} \]

In other words, (8) provides a lower variance bound on the variance of all SDFs \( m \) satisfying (2) and (5). Second, \( w^a \) determines the particular affine combination \( m^a \), among all possible combinations \( m(w) \) given by (6), which is closest to any SDF \( m \) satisfying (2) and (5) in the sense of mean square distance. Third, (7) determines a mean–variance efficient portfolio. To see this, divide (7) by \(-c\), producing:

\[ -w^a/c = \text{cov}[R]^{-1}\left(E[R] - \frac{1}{c}\right). \tag{9} \]

Interpreting the second term in (9) as a vector of mean excess returns, it is easy to verify (Huang and Litzenberger, 1988, Eq. 3.18.1) that (9) is the vector of risky asset weights in the mean–variance efficient portfolio (with a riskless asset) which has standard deviation equal to the Sharpe performance measure.

\[ ^3 \text{That is, there exists a unique solution to (2) of the form (6) when the returns' covariance matrix exists and is invertible.} \]
('Sharpe ratio') attained by the tangency portfolio of the risky assets. This Sharpe ratio is the highest attainable by a portfolio composed solely of the risky assets, and is precisely \( \sqrt{\text{var}[m^2]/c} \), where \( \text{var}[m^2] \) is the variance bound (8) (Huang and Litzenberger, 1988, pp. 77–78). In this way, the affine benchmark and variance bound characterize a mean–variance efficient portfolio and the maximal Sharpe ratio.\

The first interpretation will be used in the following illustrative application, while the others will be referred to later.

2.3. A common application of the variance bound

Suppose one wishes to investigate the potential ability of a model-based candidate for an SDF, which will generically be denoted as \( m_c \) with mean \( E[m_c] = c \), to 'price' the assets in (2). For example, in the time-separable, representative agent model of Lucas (1978) with constant relative risk aversion equal to \( \alpha \) and discount factor equal to \( \delta \), the (random) intertemporal marginal rate of substitution between time \( t \) and \( t + 1 \) is

\[
m_c = \delta(C_{t+1}/C_t)^{-\alpha},
\]

where \( C_t \) is (aggregate) consumption at time \( t \). One may wish to estimate the parameter vector (\( \delta, \alpha \)) and conduct formal tests for (10)'s compliance with (2). But before doing so, suppose one just used the sample variance from a time series of consumption to estimate \( \text{var}[m_c] \) for a pair of parameter values (\( \delta, \alpha \)). One could then use a time series of asset returns over the same period to estimate \( \text{var}[m'^2] \), and just check whether the population variance bound inequality \( \text{var}[m_c] \geq \text{var}[m'^2] \) holds in sample. By repeating this for a 'reasonable' range of parameter values, one can determine ranges of parameter values for which the variance bound inequality appears to be violated. Statistically significant violations in a range of parameters indicate that the candidate \( m_c \) can not 'price' one or more of the \( N \) assets in (2) when the parameters are in that range.\(^5\)

But even if statistical significance is not checked, the exercise is informative, and may serve to focus attention on some puzzling aspects of the candidate model.

A simplified presentation of one typical use of the variance bound (Cochrane and Hansen, 1992, Sect. 2.3) is now described, using a different data set. The data

\(^4\)See Hansen and Jagannathan (1991a, Sect. 3C) for a different, more detailed presentation of the duality between the square root (i.e., standard deviation) of the variance bound frontier and the mean–standard deviation frontier of the asset returns.

\(^5\)A referee noted that the bounds tests incorporate transversality (sufficiency) conditions of intertemporal models, in addition to the usual Euler equation (necessary) conditions. As such, they may provide more powerful tests.
employed here are inflation-adjusted monthly returns over the period 1959:7–1986:12 from three portfolios representing investments in indices of large stocks, small stocks, and government bonds, and three managed portfolios exploiting different term structure information to allocate funds between large stocks and one-month Treasury bills. Substituting sample counterparts in (8) produces estimates of the variance bound for different values of $c$, presented as the convex curve $(1/c, s^2[m])$ in Fig. 1. Monthly consumption data was used in conjunction with a few parameter pairs to produce some points $(1/\tilde{m}_c, s^2(m))$ which are also placed on Fig. 1. The figure also shows the point associated with

---

6 Thanks to Ravi Jagannathan for providing the data and his time. Hansen and Jagannathan (1991b, Table 1) describe the data base as follows: ‘The first three assets are: the value weighted portfolio of stocks in the smallest, and the largest decile in the NYSE, and a portfolio of long term government bonds. The 4th, 5th, and 6th assets consists of investing one unit in one month treasury bill and $z$ units in a zero investment portfolio that pays the excess return on the largest decile portfolio, where $z$ is either the currently known annualized yield difference between Aaa and Baa bonds (4th asset), or the yield difference between Baa bonds and one month t-bills (5th asset), or the yield difference between 1 year and 1 month t-bills (6th asset).'}
Hansen and Jagannathan's best-fitting parameter estimates (0.84, 0.87) using this data (Hansen and Jagannathan, 1991b). We immediately see in Fig. 1 that, in sample, the coefficient of relative risk aversion must be over 30 to satisfy the variance bound inequality over a reasonable range of the approximate, mean monthly real riskless interest rate $1/c = 1/E[1/r]$ for this period, i.e., between $-1$ and $+0.4$ percent per month. But Prescott and Mehra (1985) have argued that a coefficient of relative risk aversion in excess of 30 is too high to be compatible with information from other sources – information not incorporated into the GMM estimation process. Cochrane and Hansen (1992) discuss other asset pricing puzzles which are quickly discovered using this variational characterization.

While one could test for the statistical significance of the above findings (Cochrane and Hansen, 1992, Sect. 2.4), Fig. 1 itself has value, by providing a simple tool to uncover parameter regions which may be inconsistent with prior information, prior to the conduct of more formal econometric procedures. I share the view of Cochrane and Hansen (1992, p. 129) that the bounds tests 'are not a substitute for directly testing the pricing implications of a model'. In this vein, it is not too worrisome that only a small subset of the possible moment conditions are used in the variational characterization, nor that more precise ways of incorporating the implications of conditioning information are available. For a spirited defense of this position, see Cochrane (1992, Sects. 6.1–6.3).

3. The Gibbs benchmark SPD and information bound

3.1. The information bound problem and its solution

Again, we consider SDF's satisfying (2) and (5). Divide (2) by $E[m] = c > 0$ to produce the following conditions:

$$E\left[ R_i \frac{m}{E[m]} \right] = \int R_i \frac{m}{E[m]} d\mu = \frac{1}{c}, \quad (11)$$

which is more compactly written

$$E_i[R_i] \equiv \int R_i dv = \frac{1}{c} \quad i = 1, \ldots, N, \quad (12)$$

by utilizing the change of measure

$$dv = \frac{m}{E[m]} d\mu. \quad (13)$$

This section derives a variational characterization of the set of risk-neutral measures $v$ satisfying (12). Specifically, we minimize a function of measures $v$ which have a state price probability density (SPD) $dv/d\mu$. An SDF $m$ with mean $E[m] = c > 0$ will not produce an SPD $m/E[m]$ satisfying (11) unless it is
nonnegative. Thus, reformulating the SDF requirements (2) and (5) into the SPD requirement (12) automatically imposes an additional nonnegativity restriction. In Section 4.1, it is argued that there are some applications in which this restriction is desirable, and we will contrast the following way of imposing it with a very different construction used in deriving a variance bound over nonnegative SDFs.

As described in Section 2.2, the affine benchmark SDF (6) is chosen by minimization of variance over the set of SDFs. In choosing a benchmark SPD satisfying (12), we will minimize the Kullback-Leibler Information Criterion (KLIC) \( I(v, \mu) \) (White, 1982), given by

\[
I(v, \mu) = \int \log(dv/d\mu) dv. \tag{14}
\]

over the set of SPDs satisfying (12). Specifically, the benchmark SPD will solve the following convex program:

\[
v^\prime = \arg \min_v I(v, \mu) = \int \log(dv/d\mu) dv \quad \text{subject to (12)}. \tag{15}
\]

It is well-known that the solution \( v^\prime \) of problem (15) has the following Gibbs density, familiar from statistical mechanics and importance sampling,

\[
\frac{dv^\prime}{d\mu} = \frac{e^{\sum_{i=1}^N w^i R_i}}{E[e^{\sum_{i=1}^N w^i R_i}]} \tag{16}
\]

under regularity conditions guaranteeing that there is a measure satisfying the constraints (12), which has a density of the form (16) (Csiszar, 1975, Sect. 3(A)). Note that the Gibbs benchmark SPD (16) is parameterized by a specific vector of parameters \( w^\prime = (w^1, \ldots, w^N) \), which are the weights in a specific linear combination (i.e., index) of the asset returns determining the density. The benchmark’s parameters solve the following unconstrained, convex minimization problem:

\[
w^\prime = \arg \min_{w_1, \ldots, w_N} M(w) = E\left[ e^{\sum_{i=1}^N w^i R_i - 1/c} \right]. \tag{17}
\]

To verify that (17) is the vector of parameters needed in the Gibbs density (16), note that the first-order conditions for problem (17) are:

\[
E[(R_j - 1/c)e^{\sum_{i=1}^N w^i R_i - 1/c}] = 0, \quad j = 1, \ldots, N. \tag{18}
\]

Divide both sides of (18) by \( e^{-\sum w^i/c} E[e^{\sum_{i=1}^N w^i R_i}] \), and then solve for the constant \( 1/c \). The result shows that the Gibbs density (16) satisfies the constraints (12) for the information bound problem (15).

---

\( ^7 \) Much as in the case of the minimum norm linear projection producing the variance bound, the solution to the nonlinear projection problem (15) exists when the set of measures satisfying (12) is closed (in the variation metric) (Csiszar, 1975, Thm. 2.1). The KLIC (14) is sometimes called the relative entropy (Cover and Thomas, 1991), the discrimination information statistic (Kullback, 1959), or the directed divergence (Kapur, 1989).
It can be verified by substitution that the information bound \( I(v', \mu) \) may also be computed using the minimized value of problem (17):

\[
I(v', \mu) = -\log M(w').
\]  

(19)

3.2. Interpretation of the Gibbs benchmark and information bound

In parallel to Section 2.2, we will stress three interpretations of the Gibbs benchmark SPD, and also discuss interpretations which have no parallel in the variance bound literature.

3.2.1. Minimum distance interpretations

First, recall that an asset pricing model which defines a candidate SDF \( m_c \) with mean \( E[m_c] = c \) satisfies the variance bound inequality (8). When the candidate SDF is nonnegative, \( m_c/E[m_c] \) satisfying (11) must, by substituting (13) into (14), satisfy the following information bound inequality:

\[
I(v', \mu) \leq E \left[ \frac{m_c}{E[m_c]} \log \left( \frac{m_c}{E[m_c]} \right) \right] = I(v, \mu),
\]  

(20)

where the left-hand side of (20) is the information bound (19) attained by the Gibbs benchmark SPD (16). This is analogous to the variance bound inequality (8).

Also like the variance bound, the information bound has minimum distance interpretations. It is well-known that \( Z(v, p) \geq 0 \), with equality only when \( v = p \). From (15), the information bound \( I(v', \mu) = 0 \) when \( \mu \) is in the constraint set (12), implying that all mean excess returns are zero, i.e., the vector of risk premia \( E[R] - 1/c = 0 \). Thus \( I(v', \mu) > 0 \) if and only if risk premia (positive or negative) are present. Also, \( \sqrt{2I(v, \mu)} \geq \| v - \mu \| \), where the latter is the widely used \textit{variation metric} distance between \( v \) and \( \mu \) (Csiszar, 1975, p. 148). So the larger the variation distance between \( v \) and \( \mu \), the larger \( I(v, \mu) \) must be. In other words,

---

8 One might guess that the numerator in (16) is an SDF with mean equal to \( c \). This is not generally true. But this is of no consequence, because we have just shown that a candidate SDF \( m_c \) satisfying (2) and (5), which produces an SPD \( m_c/E[m_c] \), must satisfy (12) and hence satisfy (20). A Gibbs benchmark SDF is easily produced, though, by multiplying (16) by the constant \( E[m_c] = c \).

9 The KLIC distance \( I(v, \mu) \) is not a metric. But this is not very important in our application. Metrics are particularly appropriate as measures of distance in applications requiring a symmetric distance between two \textit{arbitrary} measures. But here the desired distance is that from a measure \( v \) satisfying (12) to a \textit{fixed} measure \( \mu \). Similarly, in goodness-of-fit applications, Pearson's \( \chi^2 \) goodness-of-fit statistic is a useful, nonmetric distance between a sample frequency distribution and the fixed theoretical distribution. Furthermore, the KLIC projection (15) producing the information bound does share an important property of the minimum norm linear projection which produces the variance bound: when \( v \) satisfies (12), the 'orthogonal' decomposition \( I(v, \mu) = I(v, v') + I(v', \mu) \) of the distance still holds (Csiszar, 1975, Eq. 3.1).
the information bound is larger when a risk-neutral change of measure \( \nu \) deviates more from the actual state probability measure \( \mu \). Moreover, there is an additional frequentist quantification of this interpretation, valid in the special case of IID returns. Denote the (random) vector of mean returns, from a time series of length \( T \), by \( \bar{R} \). In the Appendix, precise statements are proven having the following interpretation: when the information bound is positive, it is the asymptotic rate at which \( \text{Prob}[\bar{R} \approx 1/c] \) goes to zero, as \( T \to \infty \), and the approximation tolerance goes to zero. That is, when the information bound is large, it is less likely that one would see a large sample vector of risk premia in a small neighborhood of zero. Thus, a positive information bound does not merely signal that there are nonzero risk premia; the 'size' of the risk premia (in the above sense) is also indicated by the size of the information bound.

Second, the affine benchmark weight vector (7) determines the particular affine combination (6) which is closest (in mean square distance) to any SDF satisfying (2) and (5). Similarly, the Gibbs benchmark weight vector (17) determines the particular density in the Gibbs class \( d\nu(w)/d\mu \) of generalized exponential densities with the form (16), which is KLIC-closest to any SPD satisfying (12). More formally, the KLIC projection is one of the few nonlinear projections satisfying the following directed orthogonality property (Jones, 1989, pp. 652–653):

\[
\nu' \equiv \arg \min_{\nu} I(\nu, \mu) \quad \text{subject to} \quad (12) = \arg \min_{\nu(w)} I(\nu, \nu(w)). \tag{21}
\]

### 3.2.2. A quasi-maximum likelihood interpretation

The directed orthogonality property (21) may also be used to give a quasi-maximum likelihood interpretation of the Gibbs benchmark SPD (16). To see this, use the definition (14) to compute

\[
I(\nu, \nu(w)) = \int \log \left( \frac{d\nu}{d\nu(w)} \right) d\nu
= \int \log \left( \frac{d\nu/d\mu}{d\nu(w)/d\mu} \right) d\nu
= \int \log \left( \frac{d\nu}{d\mu} \right) d\nu - \frac{d\nu}{d\mu} d\mu - \int \log \left( \frac{d\nu(w)}{d\mu} \right) d\nu
= \mathbb{E} \left[ \frac{d\nu}{d\mu} \log \left( \frac{d\nu}{d\mu} \right) \right] - \mathbb{E}_\nu \left[ \log \left( \frac{d\nu(w)}{d\mu} \right) \right]. \tag{22}
\]
Because the first term in (22) does not depend on the parameter vector \( w \), the parameter vector which maximizes the second term in (22) minimizes \( I(v, v(w)) \). In a different context, White (1982, p. 3) describes and utilizes the notion of Akaike (1973) that maximization of the second term’s sample counterpart is a quasi-maximum likelihood procedure. To apply their reasoning in our context, suppose there is an unknown SPD \( dv/d\mu \) satisfying (12), which is misspecified to be in the Gibbs class \( dv(w)/d\mu \). The second term in (22) is the population counterpart of the quasi-log-likelihood of a sample of risk-adjusted returns from the unknown density. The directed orthogonality property (21) implies that the parameter vector maximizing the second term in (22) is the Gibbs benchmark parameter vector \( w^j \) in (16), determining the Gibbs benchmark SPD (16) KLIC-closest to this unknown SPD \( dv/d\mu \).

### 3.2.3. An optimal portfolio choice interpretation

The third interpretation of the affine benchmark (7) is that it determines a mean–variance efficient portfolio, with risky asset weights given by (9). We will now see that the Gibbs benchmark SPD parameter vector (17) determines the optimal portfolio for expected-utility-maximizing investors who have constant absolute risk aversion (CARA) utility. Consider the standard model of optimal portfolio choice (Huang and Litzenberger, 1988, pp. 18–19), in which an investor with initial wealth \( W_0 \) invests \( w_i \) units of wealth in risky asset \( i \), \( i = 1, \ldots, N \), while the rest \( (W_0 - \sum_i w_i) \) is invested in a riskless asset with gross return \( 1/c \). Terminal wealth is then \( W = W_0(1/c) + \sum_{i=1}^{N} w_i(R_i - 1/c) \) (Huang and Litzenberger, 1988, Eq. 1.18.1). The CARA utility with absolute risk aversion coefficient equal to \( a \) is \( U(W) = -e^{-aw}/a \). The expected-utility-maximizing portfolio solves

\[
V = \max_{w_1, \ldots, w_N} E[U(W)]
\]

\[
= - \frac{e^{-aw_0/c}}{a} E \left[ e^{\sum_{i=1}^{N} w_i(R_i - 1/c)} \right]
\]

\[
= U \left( W_0 \frac{1}{c} \right) E \left[ e^{\sum_{i=1}^{N} -aw_i(R_i - 1/c)} \right].
\]

Note that for CARA utility, \( U(W) < 0 \), although marginal utility is positive. So the maximum of (23) will be attained at the minimum of its second term. Comparison of the second term and (17) immediately shows that the optimal portfolio \( w^* \) solving (23) is just:

\[
w^* = - \frac{w^j}{a}.
\]
Thus, the Gibbs SPD weights \( w^j \) solving (17) determine the optimal portfolio with CARA utility. The vector of risky asset investments is always proportional to \( w^j \). The proportionality constant is \(-1/a\), where \( a \) is the coefficient of absolute risk aversion.

This finding provides the basis for a type of ‘utility-based metric’ (Rossi and McCulloch, 1990) which helps to assess the economic consequences of changes in the information bound. To see this, first note that \( U(W_0/c) \) is the utility attainable when there are no risky assets, i.e., if all initial wealth must be invested in the riskless asset with gross return \( 1/c \). Now divide both sides of (23) by \( U(W_0/c) \) to produce \( V/U(W_0/c) \). CARA utility has \( \lim_{W \to -\infty} U(W) = 0 \), so \( V/U(W_0/c) > 0 \). Furthermore, because \( 0 > V \geq U(W_0/c) \), \( V/U(W_0/c) \leq 1 \). So \( \log(V/U) \leq 0 \). Substituting (24) into (23), we derive \(-\log(V/U) = (19) \geq 0\). Taking absolute values yields the following alternative characterization of the information bound:

\[
\left| \log\left( \frac{V}{U(W_0/c)} \right) \right| = I(v^j, \mu).
\]

From (25), we see that the information bound \( I(v^j, \mu) \) is a measure of the growth in CARA utility achievable by investing in the risky assets used in (12). Note that despite the fact that the optimal CARA portfolio's risky asset investments (24) depend on the coefficient of absolute risk aversion, the benefit measure (25) does not, and hence is valid for all CARA investors regardless of their specific coefficient of absolute risk aversion or initial wealth!

### 3.2.4. A Bayesian interpretation

Finally, suppose a Bayesian wishes to choose a risk-neutral measure \( \nu \), incorporating prior knowledge of the measure \( \mu \) and the moment conditions (12) defining risk neutrality. There is always information gained by knowledge of \( \mu \) and (12), which is reflected in the risk-neutral change of measure \( \nu \) satisfying (12). In order that the choice not be influenced by unspecified restrictions, it is reasonable to select a measure which reflects no additional information other than (12). It is shown below that the Gibbs benchmark SPD (16) is axiomatically rationalized as that density which minimizes the information gained by a change of measure satisfying (12), and thus does not embody extraneous information. It is this axiomatic rationalization of the KLIC, coupled with the flexibility of measures produced by simple problems like (15), that have motivated many econometric applications of information theory. In particular, when \( \mu \) is uniformly distributed (i.e., a flat prior), solution of (15) is equivalent to constrained
maximization of (Shannon) entropy—the well-known Maximum Entropy (MaxEnt) Principle of Jaynes (1979).\textsuperscript{10}

The axiomatic rationalization of KLIC, as a measure of information gained, starts by considering an arbitrary function \( I(\nu, \mu) \) of two discrete probability measures, \( \nu_j \) and \( \mu_j, j = 1, \ldots, K \). Hobson (1971, Sect. 2.3) postulates the following axioms. First, \( I(\nu, \mu) \) should be a continuous function of \( \mu \) and \( \nu \). Second, a mere relabelling of the discrete outcomes (i.e., permuting the indices on the probabilities) should not change the value of \( I \). Third, \( I(\mu, \mu) = 0 \). That is, no information is gained unless there is a change of measure. Fourth, suppose that \( \mu \) is uniformly distributed on a subset of \( m \) outcomes (zero elsewhere), while \( \nu \) is also uniformly distributed, but on only \( n \) of those outcomes, \( n \leq m \). Then, \( I \) should be increasing in \( m \), because more information is gained when \( \nu \) rules out more of the outcomes possible under \( \mu \). Furthermore, \( I \) should be decreasing in \( n \), as less information is gained when \( \nu \) is more diffuse. Fifth and finally, a complicated 'composition rule' is postulated, which is also not unreasonable, but too complicated to compactly summarize here. Then, Hobson (1971, App. A) shows that the only functions satisfying these axioms are proportional to \( I(\nu, \mu) = \sum \lambda_j v_j \log(v_j/\mu_j) \), which is the discrete case of the Lebesgue integral (14). An insightful combinatorial interpretation of this discrete case is given by Snickars and Weibull (1977).

Alternative, more complicated axiomatic rationalizations for applying the KLIC in our context are given in Shore and Johnson (1980) and Csiszar (1991). Our context is a problem of selecting one of many possible benchmarks satisfying the linear constraints (2) or (12). These are dubbed linear inverse problems in the scientific and engineering literatures (Jones, 1989). These papers advance desiderata for a benchmark selection rule. Desiderata are formulated whose satisfaction requires that selection be made by KLIC projection.\textsuperscript{11}

Finally, a large deviations theorem of Csiszar (1985) shows that if one conditions on the (thin) set of realizations for which the large-sample mean returns are all close to 1/c, the empirical distribution of the state will be close to the Gibbs benchmark measure \( \nu' \). Csiszar's interpretation of this result provides an argument for considering \( \nu' \) to be a Bayesian posterior update of the prior risk-neutral density \( \mu \), in light of sample information consistent with the constraints (12).


\textsuperscript{11} Because the work is quite complicated and not central to this paper, the reader is referred to those papers for more information.
3.3. Estimation

Given a time series of asset returns, one estimates the solution to (17) by substituting a time average for the expectation operator, obtaining the following estimator $\hat{w}^l$ of $w^l$:

$$\hat{w}^l \equiv \arg \min_w \hat{M}(w) = \frac{1}{T} \sum_{t=1}^{T} e^{\sum_{i=1}^{N} w_i (R_t - 1)^i}. \tag{26}$$

Because both the gradient and Hessian matrix of (26) are easy to compute analytically, the Newton-Raphson method is used to solve (26). For reasons shown in Section 4.1.1, minus one times the sample counterpart of (9) provides the good initial guess required for Newton-Raphson. In practice, convergence is rapid to a numerical solution, which is substituted into (19) to produce the estimated information bound. The estimator (26) is an extremum estimator. Theorems 4.1.2 and 4.1.3 in Amemiya (1985, pp. 110–111) characterize regularity conditions under which $\hat{w}^l$ will be a consistent and asymptotically normal estimate of $w^l$. When these regularity conditions hold, $(1/\sqrt{T})(\hat{w}^l - w^l)$ approaches a normal distribution with mean 0 and covariance matrix $H(w^l)^{-1} B(w^l) H(w^l)^{-1}$, where $H$ denotes the Hessian matrix of the function $M$ in (17) and $B$ denotes the covariance matrix of the gradient of $e^{\sum_i w_i (R - 1)^i}$, taken with respect to $w$. With the data set used in this paper, the sample counterparts of these matrices evaluated at $\hat{w}^l$ were found to be positive definite, and were used to produce standard errors of quantities dependent on $\hat{w}^l$.

3.4. An illustrative example

This example is mainly intended to illustrate the interpretations of Section 3.2.1, using the same data base used to produce Fig. 1. The portfolio choice interpretation of Section 3.2.3 will be used in Section 4.2.1.

To ensure consistency of the estimator, we assume, as most researchers do, that the search for $\hat{w}^l$ is conducted over a large enough set to contain the unique, global minimizer $w^l$. This satisfies condition (A) of Amemiya’s Theorem 4.1.2. Its condition (B) is satisfied, and we assume a form of ergodicity strong enough to satisfy its condition (C). $\hat{w}^l$ then consistently estimates $w^l$. To ensure asymptotic normality of $\hat{w}^l$, note that condition (A) of Theorem 4.1.3 is obviously satisfied, and we additionally assume returns processes which permit the satisfaction of its conditions (B) and (C), which respectively ensure convergence of the sample Hessian matrix of $M$ to its population counterpart, and the ability to apply a central limit theorem to $(1/\sqrt{T}) (\partial \sum_{t=1}^{T} e^{w_i (R - 1)^i} / \partial w)$. As a check on the reasonableness of standard errors computed using this asymptotic theory, alternative standard errors were computed, under an IID assumption, using 100 bootstrap replications of the data. The bootstrapped standard errors were within 10 percent of the asymptotic standard errors on coefficients significantly different from zero.
Convexity and the envelope theorem applied to (17) imply that the information bound should be a convex function of $1/c$, with slope equal to $\sum_i w_i$ at the point $1/c$. This geometry is present in the estimated frontier of Fig. 2. Hansen and Jagannathan (1991b) produced a candidate SDF $m_c$ by finding the best-fitting affine function of the NYSE real excess return ($NYSEr$). This produced a candidate $m_c = -2.44 + 1.01 NYSEr$. In Fig. 2, we see that the information gained by the estimated candidate risk-neutral change of measure $dv_c = m_c/\tilde{m}_c d\mu$ (i.e., $KLIC = 0.0054$ on the right-hand side of (20)) is far less than the estimated information bound (i.e., $KLIC = 0.0321$ on the left-hand side of (20)) at $1/c = 1/\tilde{m}_c = 0.9953$, despite the fact that it is produced by a reasonable linear factor model, i.e., a best-fit affine function of a broad-based market excess return. That is, the candidate change of measure fails to perform the required degree of risk adjustment. It is just too close to the actual state probability measure $\mu$ to do so. This is further illustrated in Fig. 3, which depicts the histogram based on the estimate of the Gibbs benchmark SPD $dv/\mu$ and the histogram of the candidate density $dv_c/\mu$. There, we see that the latter is too sharply peaked relative to the former. That is, the candidate density is relatively closer to the density $dv/\mu \equiv 1$ that would prevail were there no risk premia. As a result, the
Candidate Density vs. Gibbs Density

\[ m = -2.44 + 1.01 \text{ NYSEx} \]

Fig. 3. The candidate risk-neutral measure is too close to the actual state probability measure (with degenerate density) to perform the degree of risk adjustment.

candidate change of measure will not produce risk-adjusted risk premia equal to zero. The IID, large deviations interpretation of the information bound (see the appendix) implies that the probability of observing near-zero sample risk premia goes to zero at an estimated asymptotic rate of \( \hat{I} = 0.0321 \) per month. If the risk premia were much smaller, this rate would have been much less. Had it been less than the candidate KLIC value (0.0054), the candidate possibly could have produced risk-adjusted risk premia equal to zero. But the estimated rate was
actually around six times higher than this, so the candidate measure can not perform the degree of risk adjustment that valid SPDs must do.13

While the above only used point estimates, interval estimates derived in Section 3.3 will be needed and used in the empirical applications developed in Section 4.2.1.

4. Comparison of the bounds

In addition to their differing statistical interpretations, the two variational characterizations differ in the way they impose a positivity constraint, and in the optimal portfolios they determine.

4.1. The positivity constraint

There is a problem associated with the affine benchmark SDF $m^a$, found by substituting (7) into (6), which is alleviated by the Gibbs benchmark SPD (16). First, the absence of arbitrage implies that there is a nonempty, convex set of positive $m$ satisfying (2). Yet there is no guarantee that the affine benchmark SDF $m^a$ will be positive. This is no problem if the sole use of the benchmark is to 'price' assets with returns replicated by portfolios of the $N$ payoffs in (1), for they will satisfy (2) regardless of the sign of $m^a$.14 But suppose one wishes to test whether a benchmark will price assets which are not static portfolios of the $N$ assets in (2). These might be derivatives, such as options on one or more of the $N$ assets in (2), a managed portfolio not included in (2), or just some other asset. Insisting on a nonnegative $m$, as Bansal and Viswanathan (1993) do, insures that it will never predict a negative price for assets with nonnegative payoffs.15

Accordingly, Hansen and Jagannathan (1991a) devote several pages of their article and its appendix investigating the addition of both nonnegativity and strict positivity constraints in the variance bound problem. The nonnegative benchmark SDF that they construct is no longer a simple affine combination of the returns, because as noted by Cochrane and Hansen (1992, p. 129): 'Unfortunately, we can no longer appeal to least-squares regression theory to derive these bounds.' Consequently, it no longer determines the mean–variance efficient portfolio weights (9) (Snow, 1991, fn. 10). Defining $R_0 = 1/c$, some papers do attempt to estimate the weight vector $w$ solving the following optimality

---

13 As in the case of the variance bound application of Section 2.3, the amount by which a candidate SPD fails to exceed the information bound is not a 'utility-based metric' (Rossi and McCulloch, 1990) for evaluating the economic significance of the candidate's mispricing.

14 I am indebted to John Cochrane for this argument.

15 I am indebted to Ravi Bansal for this argument.
Table 1
Positive analysis: Unconstrained affine SDF

<table>
<thead>
<tr>
<th>Rate</th>
<th>Var.</th>
<th>Max.</th>
<th>Min</th>
<th>% &lt; 0</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.99</td>
<td>0.207</td>
<td>3.69</td>
<td>-1.03</td>
<td>1.8</td>
</tr>
<tr>
<td>0.994</td>
<td>0.078</td>
<td>2.33</td>
<td>-0.63</td>
<td>0.9</td>
</tr>
<tr>
<td>0.998</td>
<td>0.048</td>
<td>1.66</td>
<td>-0.23</td>
<td>0.6</td>
</tr>
<tr>
<td>1.0</td>
<td>0.070</td>
<td>1.86</td>
<td>-0.06</td>
<td>0.9</td>
</tr>
<tr>
<td>1.004</td>
<td>0.185</td>
<td>2.66</td>
<td>-1.04</td>
<td>3.0</td>
</tr>
</tbody>
</table>

The last column shows that the monthly data affine benchmark SDF has negative values in only a small percentage of months. As expected, there are more negative values when the variance bound is higher. Maximum SDF elements are larger in absolute value than minimum elements. As a result, the nonnegativity constraint does little to change the variance bound (see Fig. 4).

4.1.1. Approximation when the variance bound is low

Moreover, when the variance bound is relatively low, use of a quadratic approximation permits the Gibbs benchmark SPD's weights (17) and the information bound (19) to be approximated by simple transformations of the affine benchmark SDF weights (7) and variance bound (8). To see this, expand the exponential term in (17) to second order about the vector of gross mean returns $E[R]$, and take the unconditional expectation to obtain

$$e^{E[R] - 1/c}w(1 + 1/2 w' \text{cov}[R]w).$$
Minimizing (28) is equivalent to minimizing its logarithm. When $w' \text{cov}[R]w$ is relatively small, one can substitute the approximation $x \approx \log(1 + x)$ and solve the resulting quadratic minimization problem

$$\min_w (E[R] - 1/c)'w + \frac{1}{2} w' \text{cov}[R]w,$$

(29)

to yield the following quadratic, affine benchmark-based approximation to (17), denoted $\tilde{w}$:

$$\tilde{w}' = - \text{cov}[R]^{-1}(E[R] - 1/c) = w^*/c,$$

(30)

Fig. 4. Nonnegativity constraint, though binding, has little affect in this range of $1/c$. 

Effect of Nonnegativity Constraint

Variance Bound: Monthly Returns
where the latter equality follows from (7). From (9), we see that (30) is just \(-1\) times the vector of risky asset weights in a mean-variance efficient portfolio. Substituting (30) into (29) and simplifying yields the following quadratic, variance-bound-based approximation of the information bound

\[
\bar{I}(\mu, \mu) = \text{var}[m^\alpha]/2c^2.
\]

(31)

Because the approximation \(\log(1 + x) = x\) is more accurate the smaller \(w^\alpha \text{cov}[R]w^\alpha = \text{var}[m^\alpha]\) is, one would expect the approximation to be better when the variance bound \(\text{var}[m^\alpha]\) is smaller. To investigate this issue empirically, Fig. 5 illustrates the approximation error using the monthly data. The approximation error is less than 13.5 percent, with the worst approximation near the bottom of the frontier. Substituting the nonnegatively constrained variance bound (see Fig. 4) into (31) only slightly improves the approximation, with a maximum error of 12.7 percent at \(1/c = 0.99\) and an error near the bottom of the frontier of 12.4 percent.

4.1.2. When the variance bound is high

But there are plenty of applications in which the variance bound is not low. For example, Hansen and Jagannathan (1991a, Fig. 6) show that imposition of
Information Bound vs. Approximations
Quarterly Returns

Fig. 6. Approximation error is worse with quarterly data. The maximum error is 22 percent when using the variance bound with nonnegativity imposed.

nonnegativity makes a big difference when considering quarterly holding period returns from different Treasury bills.\(^{16}\) Fig. 6 in the present paper shows that the quadratic approximation (31) is worse with quarterly data, where the variance bounds are higher. Using (31), the maximum approximation error is now 25 percent, at the upper end of the interval (corresponding to an approximate mean real riskless rate of 1.4 percent per quarter), and is generally over 15 percent. Even after substituting the higher, nonnegatively constrained variance bound from (27) into (31), the maximum error is still 22.5 percent, although it performs better at the upper end of reasonable mean riskless rates. In general, with lower frequency data the approximation will be worse. This finding is important, for some have argued that important asset pricing models fit better at lower frequencies. For example, Brainard, Nelson, and Shapiro (1991) show that the consumption CAPM model explains data better at longer horizons (up to three years) than it does at much shorter intervals, and provide intuitive reasons for the failure of consumption-based models to fit monthly data.\(^{17}\)

\(^{16}\) I am indebted to Erzo Luttmer for pointing this out.

\(^{17}\) For other results showing that the measurement interval matters in tests of the CAPM, see Handa et al. (1993).
Moreover, unlike the information bound, the solution to (27) is nonnegative, but not necessarily positive. Hansen and Jagannathan (1991a, p. 240) summarize their approach to the positivity constraint as follows: 'The resulting volatility bounds for nonnegative random variables...also apply when the random variables are restricted to be strictly positive...However, in this case the lower bounds may only be approximated rather than attained.' In contrast, the Gibbs benchmark SPD (16) is always positive, producing a risk-neutral measure \( v' \) which puts positive probability wherever the state probability measure \( \mu \) does. The minimized value in problem (15) is the information bound \( I(v', \mu) \) attained by this measure. Yet (16) is still determined by a simple linear combination \( w' \) of the asset returns, which still determines the optimal portfolios for a frequently analyzed class of investors. Neither the variance bound nor the bounds created by Snow (1991) from the \( p \)th moment of the return vector and the \( q \)th moment of \( m \), where \( 1/p + 1/q = 1 \), have all these properties.

4.2. Optimal portfolio choice

As described in Section 2.2, the affine benchmark and variance bound determine a mean–variance efficient portfolio with risky asset weights (9) and the maximal Sharpe ratio among mean–variance efficient risky asset portfolios. This interpretation is the basis of some applications of the bound. For example, DeSantis (1993b) has examined changes in variance bounds induced by adding international asset returns to a set of domestic asset returns, with a view toward examining whether estimated gains from international diversification are statistically significant. He describes one of his tests as follows (DeSantis, 1993b, p. 7): 'The test can be used to establish whether the investment opportunities associated with two sets of assets (such that one set always contains the other) are tangent in the point with the highest Sharpe-ratio.' While there are econometric advantages to using the variance bound framework to examine questions like these, the questions answered are still about mean–variance efficiency frontiers. Is it conceivable that 40 years of mean–variance analysis has brought us to the point of diminishing returns on the research frontier? Furthermore, as noted by Huang and Litzenberger (1988, p. 61): 'For arbitrary distributions, the mean–variance model can be motivated by assuming quadratic utility .... Unfortunately, quadratic utility displays the undesirable properties of satiation and increasing absolute risk aversion.' Going past the point of satiation leads to negative IMRSs, causing the pricing problems associated with nonpositive SDFs discussed in Section 4.1.18

18 Mean–variance analysis may also be justified by the assumption that the joint returns distribution is multivariate normal. But the comparative advantage of the bounds approach is its avoidance of explicit distributional assumptions.
The small-firm weight \( w_i \) in the Gibbs benchmark SPD is generally significantly different from zero (standard errors are in parentheses). The last column is the slope of the information bound frontier. Changing its sign yields the net use of risky assets in the CARA portfolio with \( a = 1 \), which is positive for relatively low values of \( 1/c \) and negative for relatively high values, in accord with intuition.

The following empirical application will use the information bound to examine whether the small firm index plays a statistically significant role in CARA optimal portfolios, and whether the economic gain from this diversification is statistically or economically significant.

4.2.1. Small-firm effects

As previously noted, Snow (1991) has generalized the Hansen–Jagannathan approach to analyze bounds on certain other moments, in addition to the variance. He used these bounds to investigate small-firm effects. He described the core of his analysis in the following words (Snow, 1991, p. 973): 'A more meaningful analysis is to determine whether, after generating the frontiers with the returns of large firms and/or a market proxy, adding in the returns of small firms causes the frontier to shift upward and hence to become more restrictive.' Snow (1991, p. 981) provides the following interpretation of his findings over the data period covered by our data base: 'Examination of returns from 1959–1987 yielded no evidence of a small-firm effect.'

To illustrate a different way of testing for small-firm effects over the same period with a different data set, a test of small firms' impact on CARA optimal
Table 3
Gibbs benchmark SPD weights when small firms are excluded from the investment opportunity set

<table>
<thead>
<tr>
<th>Rate</th>
<th>Gibbs benchmark weights</th>
<th>Slope</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$w^i_1$</td>
<td>$w^i_2$</td>
</tr>
<tr>
<td>0.990</td>
<td>— 2.65</td>
<td>— 8.56</td>
</tr>
<tr>
<td></td>
<td>(4.13)</td>
<td>(2.76)</td>
</tr>
<tr>
<td>0.994</td>
<td>3.17</td>
<td>— 3.30</td>
</tr>
<tr>
<td></td>
<td>(3.80)</td>
<td>(2.34)</td>
</tr>
<tr>
<td>0.998</td>
<td>— 4.51</td>
<td>1.71</td>
</tr>
<tr>
<td></td>
<td>(3.57)</td>
<td>(2.27)</td>
</tr>
<tr>
<td>1.0</td>
<td>— 5.87</td>
<td>3.99</td>
</tr>
<tr>
<td></td>
<td>(3.54)</td>
<td>(2.36)</td>
</tr>
<tr>
<td>1.004</td>
<td>— 10.89</td>
<td>8.04</td>
</tr>
<tr>
<td></td>
<td>(3.8)</td>
<td>(2.81)</td>
</tr>
</tbody>
</table>

Comparing the last column of Tables 2 and 3 shows that the total use of risky assets does not change much when small firms are excluded. Standard errors are in parentheses.

benchmark portfolios is performed. Table 2 shows that the standard error of estimate on the small-firm return weight $w^i_1$ is small relative to the weight, for a variety of $1/c$ values. Thus, the small-firm index does play a statistically significant role in CARA optimal portfolios which also include a large-firm index. CARA optimal portfolios have long positions ($-w^i_1/\alpha > 0$) in the small-firm index.19 The last column in Table 2 shows that the portfolio for $\alpha = 1$ is long in risky assets ($-\sum_i w^i_l > 0$) when the riskless return $1/c$ is relatively low, and short in the index when the riskless return is relatively high, in accord with intuition.

Comparison of the last column of Table 2 with the last column of Table 3 shows that the total use of risky assets is almost the same when the small-firm index is excluded from the investment opportunity set. So, rather than drawing more funds into risky assets, investors reallocate risky asset funds when the small-firm index is available. But it is possible that the gain in expected utility from diversification into small firms may not be too large. To determine this, Fig. 7 contrasts the estimated information bound frontier without the small firm

---

19 The information in the table is sufficient to analyze the CARA optimal portfolio for any coefficient $\alpha$, because the mean of $-w^i_1/\alpha$ divided by its standard error is the same as the mean of $-w^i_1$ divided by its standard error.
Fig. 7. Small-firm effect on the information bound over a reasonable range of $1/c$. The increase in the bound is a measure of the increase in CARA utility attributable to the ability to invest in an index of small NYSE firms.
riskless rate range where the reliance on risky assets is the least, i.e., where the slope of the frontier $\sum_i w_i$ is small in absolute value.\(^{20}\)

In summary, it is fair to conclude that during this period and with this data set, CARA optimal portfolios would include a small-firm index. The relatively large upward shift in the information bound frontier, in the middle range of $1/c$ values, shows that there is information in small-firm returns not already contained in large-firm returns. Eq. (25) shows that this information gain is a measure of the incremental benefits to investors from diversification into small firms. And unlike the use of the word ‘information’ in Snow (1991) or in Cochrane and Hansen (1988, p. 116), this use of the word is axiomatically rationalized, as described in Section 3.2.4.

5. Conclusions

To evaluate the performance of arbitrage-free asset pricing models intended to price a subset of assets, it is useful to select a benchmark state price density (SPD) for a risk-neutral measure which correctly prices these assets. This may be compared to the SPD implied by a particular asset pricing model. This paper proposes a new way of selecting a benchmark SPD, by minimizing the Kullback–Leibler Information Criterion (KLIC) ‘distance’ between the set of risk-neutral measures and the actual probability measure over states of nature — a nonlinear projection problem utilized in information-theoretic econometrics and some of its uses by Bayesians. The solution to the projection problem is a Gibbs benchmark SPD, and the minimized KLIC value attained by the solution is an information bound which asset pricing models’ SPDs must exceed if they are to price the assets correctly. Statistical interpretations of the information bound support its use as a quantitative measure of the degree of risk adjustment required of any asset pricing model seeking to price the subset of returns under investigation.

In addition, the Gibbs benchmark SPD is determined by an easily computed linear combination of the asset returns. The weights in this index of returns are proportional to the optimal risky asset investments of expected-utility-maximizing investors with constant absolute risk aversion (CARA) utility. Using this result, the information bound has an economic interpretation as a utility-based measure of the benefit associated with changes in the investment opportunity.

\(^{20}\) A broad-based stock index ‘market portfolio’ is not in the set of assets, although the largest NYSE decile is. Furthermore, the small-firm index contains only the smallest decile of NYSE firms. So our findings are not directly comparable to Snow (1991, Table 2). Because the main intent of this Section is to illustrate a use of the information bound with this data set, and the main purpose of this paper is to introduce the information bound to a general audience, no further investigations were made.
set. These results were used to show that between 1959:7 and 1986:12 investment in an index of small NYSE firms would occur in CARA optimal portfolios, which would otherwise be restricted to investments in large NYSE firms, long-term government bonds, and three managed portfolios involving large firms and Treasury bills. The incremental benefits from small firm investment were relatively large in a range of riskless rates over which investors' overall use of risky assets would be relatively small.

The most frequently used alternative tool for these purposes is the Hansen–Jagannathan (1991a) variance bound on the set of stochastic discount factors (SDFs). In contrast to the variance bound, the information bound automatically incorporates a positivity constraint useful in some applications. And while the affine benchmark SDF which attains the variance bound also determines an optimal portfolio, this portfolio is on the assets' mean–variance frontier. When the assets' returns are not normally distributed, emphasis on discrete-time mean–variance analysis is justified by an assumption that investors possess quadratic utility. Quadratic utility, however, possesses the undesirable properties of satiation and increasing absolute risk aversion. CARA utility does not suffer from these defects.

Finally, there were a number of topics left for future research. One topic is the effect that transactions costs have on the information bound, analogous to the work of Luttmer (1991) and He and Modest (1992). Proportional transactions costs (e.g., bid–asked spreads) will change the constraints on SPDs from a system of equations (12) into a system of inequalities. As such, the information bound will be minimized over a larger set, and could be substantially lower. Another topic for investigation is the use of the Gibbs density in the estimation of asset pricing model parameters, as a possible alternative to the generalized method of moments.

Appendix

In Section 3.2.1, it was asserted that when the information bound $I(v', \mu) > 0$, it is the asymptotic rate at which $\text{Prob}[\bar{R} \approx 1/c]$ goes to zero as the sample length $T \to \infty$ and the approximation tolerance goes to zero. A powerful large deviations result of Ellis (1984), as described and applied in Bucklew (1990, pp. 20–22), is used below to precisely state and prove the sense in which this is true.

Let $Y_1, \ldots, Y_T$ be a sequence of random $N$-vectors. Define the extended real-valued function:

$$
\phi_T(w) \equiv 1/T \log E[e^{\sum_{t=1}^T w'} \sum_{t=1}^T y_t] = 1/T \log E\left[ \prod_{t=1}^T e^{w'y_t} \right].
$$

Under the mild regularity conditions that $\phi(w) \equiv \lim_{T \to \infty} \phi_T(w)$ exists or is unambiguously $+\infty$, and is a lower semicontinuous convex function of $w$,
Ellis' Theorem holds that
\[
\limsup_{T \to \infty} \frac{1}{T} \log \Pr[\bar{Y}/F] \leq - \inf_{w \in F} \sup_{x \in X} \left[ w'x - \varphi(w) \right],
\]
where $F$ is any compact set in $\mathcal{R}^N$, and $\varphi(w) = \lim_{T \to \infty} \Phi_T(w)$, so that for suitably large $T$

\[
\Pr[\bar{Y}/F] \leq e^{-T \inf_{w, \sup w} \left[ w'x - \varphi(w) \right]}.
\]  

Result (34) is now used to derive an upper bound for $\Pr[\bar{R} \approx 1/c]$ when gross returns are IID. Replace $Y$ by the returns vector $R$, and successively use the independence and identical distribution assumptions to write (32) as

\[
\phi_T(w) = \frac{1}{T} \log \prod_{i=1}^{T} \mathbb{E}[e^{w'R}] = \log \mathbb{E}[e^{w'R}] = \varphi(w).
\]

Thus in the IID case, $\varphi(w)$ is just the logarithmic moment generating function of the returns vector. Now let

\[
F_\varepsilon = \{x|\varepsilon - 1/c \leq x_i \leq \varepsilon + 1/c, i = 1, \ldots, N\}
\]

denote the compact rectangle of tolerance $\varepsilon$ about $1/c$. Using (35), note that

\[
\inf_{x \in F_\varepsilon} \sup_{w \in F, \sup w} \left[ w'x - \varphi(w) \right] = \inf_{x \in F_\varepsilon} \left[ \inf_{w \in F, \sup w} \left[ \log \mathbb{E}[e^{w'R}] - w'x \right] \right] = \inf_{x \in F_\varepsilon} \left[ \inf_{w_1, \ldots, w_N} \log \mathbb{E}[e^{\sum w(R_i - x_i)}] \right]
\]

\[
= I_\varepsilon.
\]

Substituting (37) into (34), $\Pr[\bar{R} \in F_\varepsilon] \leq e^{-T \cdot I_\varepsilon}$ for suitably large $T$.

Now consider the open rectangle $G_\varepsilon$ contained in $F_\varepsilon$. Under additional regularity conditions, Ellis' Theorem produces a corresponding lower bound for $\Pr[\bar{R} \in G_\varepsilon]$, via:

\[
\liminf_{T \to \infty} \frac{1}{T} \log \Pr[\bar{R} \in G_\varepsilon] \geq \inf_{x \in G_\varepsilon} \left[ - \inf_{w_1, \ldots, w_N} \log \mathbb{E}[e^{\sum w(R_i - x_i)}] \right] = I_\varepsilon.
\]

So $\Pr[\bar{R} \in G_\varepsilon] \geq e^{-T \cdot I_\varepsilon}$ for suitably large $T$.

Now note that as the approximation tolerance $\varepsilon \downarrow 0$, the right-hand sides of both (37) and (38) approach the information bound $I(v', \mu)$ defined by (17) and (19). This is the sense in which to interpret the claim that

\[
\Pr[\bar{R} \approx 1/c] \approx e^{-TT(v', \mu)}
\]

for suitably large $T$, so that the information bound is the asymptotic rate at which $\Pr[\bar{R} \approx 1/c]$ goes to zero.
References


Cecchetti, Stephen G., Pok-sang Lam, and Nelson C. Mark, 1992. Testing volatility restrictions on intertemporal marginal rates of substitution implied by Euler equations and asset returns (Department of Economics, Ohio State University, Columbus, OH).

Chen, Zhiwu and Peter J. Knez, 1992a. A measurement framework of arbitrage and market integration (Graduate School of Business, University of Wisconsin, Madison, WI).

Chen, Zhiwu and Peter J. Knez, 1992b. A pricing operator-based testing foundation for a class of factor pricing models (Graduate School of Business, University of Wisconsin, Madison, WI).


De Santis, Georgio, 1993a. Asset pricing and portfolio diversification: Evidence from emerging financial markets (Department of Finance and Business Economics, University of Southern California, Los Angeles, CA).

De Santis, Georgio, 1993b. Volatility bounds for stochastic discount factors: Tests and implications from international stock returns (Department of Finance and Business Economics, University of Southern California, Los Angeles, CA).


Epstein, Larry and Stanley Zin, 1991. The independence axiom and asset returns (Graduate School of Industrial Administration, Carnegie–Mellon University, Pittsburgh, PA).


Hansen, Lars Peter and Ravi Jagannathan, 1991b, Assessing specification errors in stochastic discount factor models (Department of Finance, Carlson School of Management, University of Minnesota, Minneapolis, MN).

He, Hua and David Modest, 1992, Market frictions and consumption-based asset pricing (Department of Finance, University of California, Berkeley, CA).


Maasoumi, Esfandiar, 1993, A compendium to information theory in economics and econometrics, Econometric Reviews 12, 137-181.


Theil, Henri and Denizil Fieberg, 1984, Exploiting continuity: Maximum entropy estimation of continuous distributions (Ballinger, Cambridge, MA).


