Entropic Diagnostics for Asset Pricing SDFs:

A Critique

Michael Stutzer

JEL Classifications: C11, C14, G11, G12

Not For Circulation nor Citation

Abstract

A modern approach to asset pricing models characterizes them by their respective stochastic discount factors (SDFs), which determine risk-neutral pricing measures under which expected returns in excess of the riskfree rate are all zero. Relatively recent articles in the top journals and working papers have utilized *entropy bounds* on the SDFs that provide diagnostics of model failure. But these papers were written after econometric publications that provided alternative entropic diagnostics. We show how the earlier diagnostics are related to the recent ones. We show that the earlier diagnostics provide good insights, we provide further development of one of them heretofore neglected, and recommend more extensive use of it.
1 Introduction

In an article written for the Scandinavian Journal of Economics in honor of the 2013 Nobel Laureates in Economics (Lars Hansen, Eugene Fama, and Robert Schiller), Prof. John Campbell [4](2014, p. 601) wrote:

The idea of using asset-return data to restrict the properties of the SDF remains a fruitful one. More recent work by Stutzer (1995), Bansal and Lehmann (1997), Alvarez and Jermann (2005), and Backus et al. (2011), for example, shows how asset returns place lower bounds on the entropy of the SDF. Entropy, an alternative to variance as a measure of randomness, plays an increasingly important role in asset pricing theory...

The concept of entropy has long fascinated social scientists seeking something as usefully predictive as statistical mechanics is for the behavior of large particle systems (e.g. gases and glasses). The statistical concept of entropy and its behavior is the foundation of statistical mechanics. There have been some notable attempts to fuse conventional microeconomic theory with statistical mechanics, e.g. Brock and LeBaron [3] (1996) and subsequent developments. Being a student of Josiah Willard Gibbs’ last student (i.e. Edwin Bidwell Wilson), Paul Samuelson perhaps had the best intellectual capital for producing an econophysical entropic theory of this nature, but at the peak of his career he was not sanguine about that prospect.

During Samuelson’s reign, Claude Shannon, Solomon Kullback, and others developed a purely statistical information theory that became the foundation for modern communications
engineering. But the methods were far more applicable than that. Kullback’s [11] (1959) influential text developed entropic alternatives for the standard canon of Fisherian estimation and inference procedures. Some established econometricians picked up on this (e.g. Henri Theil and Arnold Zellner), and it has subsequently been pursued by a relatively small but dedicated group of scholars, who now have their own funded institute (see the website for the Info-Metrics Institute).

But entropic methods have not often been taught or used within business school finance departments. The recent upsurge in interest has its origins in the asset pricing model estimation and testing framework developed and refined by 2013 Nobel Laureate Lars Peter Hansen. The development started with the wherewithal to reduce the implications of asset pricing models to a set of expectation conditions often written as $E[R_i m(\theta)] = 1$, where $R_i$ is the gross return over some horizon from an asset the theory is intended to price, and $m$ is what Hansen dubbed a stochastic discount factor (SDF) that is produced by the theory (see the next section of this paper for elaboration). Hansen’s Generalized Method of Moments (GMM) was tailor-made to estimate model parameters $\theta$ subject to those moment conditions, and provides associated specification tests of those moment conditions. Hansen and Jagannathan (1991) then used related methods to develop a lower bound on the variance of any $m(\theta)$ that does satisfy the moment conditions for a pre-determined set of test assets, and showed that a believably parameterized $m$ often has too little variance to exceed the lower bound. John Cochrane’s [6](2001) influential textbook wisely made much use of the moment condition representation and variance bound, which have become well-known in
business school finance departments.

This section started with a quote from John Campbell, that referred to extensions of Hansen and Jagannathan’s work based on an information theoretic statistic called relative entropy. The following Section 2 reviews the stochastic discount factor representation of asset pricing models. Section 3 reviews the latest entropic diagnostics. An empirical application is developed to illustrate the required computations and their interpretation. Section 4 reviews the older entropic diagnostics. Section 5 compares and contrasts the newer diagnostics with the older ones, showing that they provide little or no value-added over the originals. Section 6 contains some new results that further develop a specification error entropy (Kitamura and Stutzer (2002), Diagnostic Criterion 3.2), which generalizes the Hansen and Jagannathan (1997) specification error diagnostic to reflect the impact (if any) of higher (than 2nd) order cumulants. Section 7 reviews very recent developments by Ghosh, et.al. (2014), and shows that the specification error entropy can be applied to provide similarly motivated results. Section 8 concludes.

2 Stochastic Discount Factors of Asset Pricing Models

Arbitrage-free asset pricing models imply the existence of a convex set of random variables, called stochastic discount factors (SDFs) Hansen and Jagannathan (1991). Specifically, an SDF $m$ is defined by the following conditions:

$$ E[X_{i+1} | I_t] = X_i^t \quad i = 1, \ldots, N $$

(1)
where $X_t^i$ denotes the price of risky asset $i$ at time $t$ and $X_{t+1}^i$ denotes its payoff the next period. Expectations are taken conditional on information $I_t$ at time $t$. While one does not strictly need to presume the existence of a riskfree asset, I will follow many other analysts in presuming its existence. The $N$ assets in (1) are risky assets of most interest to those constructing and testing models 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 satisfy (1). For example, in the venerable representative agent model of Lucas [13] (1978) with constant relative risk aversion equal to $\alpha$ and time discount factor equal to $\delta$, the agent chooses investments out of income and savings to maximize the consumption path objective function $E\left[\sum_{t=0}^{\infty} \delta^t C_{t+1}^{1-\alpha} / (1 - \alpha)\right]$. The (random) IMRS between time $t$ and $t+1$ is

$$m \equiv m(\theta) = \delta(C_{t+1}/C_t)^{-\alpha}$$

where $C_t$ is (aggregate) consumption at time $t$ and $\theta$ is a vector of the (two) parameters $\delta$ and $\alpha$. SDFs implied by different theories have different parameters. When necessary for clarity, the notation $m(\theta)$ will be used to emphasize the dependence on variable parameters.

More generally, the assumption of nonsatiation implies the existence of positive SDFs. More elaborate consumption-based asset pricing models imply more complex, additionally parameterized functions than (2). And while it isn’t presented this way to undergraduates, the CAPM and its multifactor extensions can be exposited by noting that they imply that $m$ is an affine or linear function of factors $m = \beta_0 + \sum_j \beta_j F_j$ (e.g. see Campbell and Cochrane
[5] (2000)).

Drop the time subscripts and divide the equations in (1) by their respective right hand sides to produce the most common representation:

\[ E[R_i m] = \int R_i m \, dP = 1 \quad i = 1, \ldots, N. \]  (3)

where \( R_i \) denotes the period’s gross real return from asset \( i \) and \( P \) is the data generating probability measure. The riskfree asset, presumed for convenience in this literature, has a known fixed gross return during the period, which may vary from period to period. Denote its gross return as \( R_f \) and substitute it for \( R_i \) in (3) to find

\[ E[m] = \frac{1}{R_f}. \]  (4)

Divide (3) by (4) to find

\[ E[R_i m E[m]] = R_f \quad i = 1, \ldots, N \]  which is more compactly written

\[ E^Q[R_{ie}] \equiv \int (R_i - R_f) \, dQ = 0 \quad i = 1, \ldots, N \]  (5)

by utilizing a risk-neutral change of measure

\[ dQ = \frac{m}{E[m]} \, dP. \]  (6)

The term ”risk-neutral” arises because the change of measure makes the expected excess gross return \( R_{ie} \) of each risky asset in (5) equal to zero.

The notation no longer indicates that expectations are conditional at time \( t \). But it can be re-interpreted as holding unconditionally, governed by a stationary measure \( P \). For example, applying (3) conditionally to the known riskfree rate between \( t \) and \( t+1 \) yielded conditional expectation (4), taking the unconditional expectation and ignoring the empirically
tiny Jensen’s Inequality bias $E[1/R_f] - E[1/E[R_f]]$ permits us to interpret the unconditional $E[m] \approx 1/E[R_f]$, in which case (5) holds unconditionally. This benign approximation (accurate to the sixth decimal place for widely-used 1 mo. T-Bill returns) explains slight differences between this exposition and those found in some papers cited later.

When we define a specific risk-neutral measure associated with a specific SDF $m$, we will use the notation $dQ^m$ rather than the generic $dQ$ notation in the constraints (5) that restrict the set of potential risk-neutral measures.

3 Jensen’s Inequality Entropic Diagnostics

Relative entropy is a nonnegative, asymmetric, bivariate function of two absolutely continuous probability measures, that is zero if and only if the two measures coincide. Denoting the data generating measure by $P$ and a possible “risk-neutral” measure (5) by $Q$, the entropy of $P$ relative to $Q$ is $D(P \mid Q) \equiv E^P[\log \frac{dP}{dQ}] = \sum_i P_i \log \frac{P_i}{Q_i}$, when the state space is discrete. This is not equal to $D(Q \mid P) \equiv E^Q[\log \frac{dQ}{dP}] = \sum_i Q_i \log \frac{Q_i}{P_i}$. Either way, the result is nonnegative, and zero if and only if the two measures coincide. Letting $Q^m$ be the risk-neutral measure determined by a specific stochastic discount factor $m$, (6) implies that the density $\frac{dP}{dQ^m} = \frac{E[m]}{m}$, so the entropy of $P$ relative to $Q^m$ is then $D(P \mid Q^m) = \log E[m] - E[\log m]$, while $D(Q^m \mid P) \equiv E[\frac{m}{E[m]} \log \frac{m}{E[m]}]$, where we delete the superscript $P$ when the expectation is taken with respect to it. Relatively recent papers in the Journal of Finance use the former, while the earlier papers in the Journal of Econometrics use the latter.

Following the exposition in Backus, Chernov, and Zin (2014), use (4) to compute:
\[ D(P \mid Q^m) \equiv E[\log \frac{dP}{dQ^m}] = E \log \left[ \frac{E[m]}{m} \right] = \log E[m] - E[\log m] = -\log E[R^f] - E[\log m] \] (7)

where (as noted earlier) we ignore the tiny Jensen’s Inequality bias that causes \( E[1/R^f] \) to deviate from \( 1/E[R^f] \).

Now follow them in taking both sides of the \( i \)th equation in (3) and apply Jensen’s Inequality to find:

\[ \log 1 = 0 = \log E[R_i m] \geq E[\log R_i] + E[\log m] \] (8)

Now add (7) to (8) to find

\[ D(P \mid Q^m) \equiv \log E[m(\theta)] - E[\log m(\theta)] \geq E[\log R_i - \log R_f] \quad i = 1,\ldots,N \] (9)

(9) shows that the entropy of \( P \) relative to \( Q^m \) can’t be lower than the log gross return risk premium achievable when pricing assets using the analyzed model’s SDF \( m \). A value for the maximum achievable maximum risk premium and inequality (9) thus provide a restriction on admissible values for the parameter vector \( \theta \) (see Bansal and Lehmann (1997)).

In addition, Backus, Chernov, and Martin (2011) evaluate the cumulant generating function \(^2\) of \( \log m \), i.e. \( CGF^{\log m}(s) \equiv \log E[e^{s \log m}] \) at \( s = 1 \) to obtain \( \log E[m] \), note that \( E[\log m] \) is the first such cumulant, and substitute into (19) to derive the equivalent expression:

\[ D(P \mid Q^m) = CGF^{\log m}(1) - \kappa_1^{\log m} \equiv \sum_{l=2}^{\infty} \frac{\kappa_l^{\log m(\theta)}}{l!} \geq E[\log R_j - \log R_f] \] (10)

which they use to decompose their \( D(P \mid Q^m) \) into an inverse factorially weighted sum of \( \log m \)'s cumulants. To illustrate this calculation, Backus, Chernov, and Martin (2011) use
(2) to write $\log m(\theta) = \log \delta - \alpha \log \frac{C_{t+1}}{C_t}$, so that $\log m(\theta)$ will be normally distributed when $\log \frac{C_{t+1}}{C_t}$ is also. Under normality, only the 2nd cumulant in the expansion (10) is nonzero, yielding $D(P \mid Q^m) = \alpha^2 VAR[\log \frac{C_{t+1}}{C_t}] / 2$, which does not depend on the discount factor $\delta$.

### 3.1 Permanent and Temporary Components

Alvarez and Jermann (2005) devote most effort to developing a bound on the permanent component $m^p$ in the multiplicative decomposition $m = m^p m^T$. As noted in Backus, Chernov, and Zin (2014, p. 60), Alvarez and Jermann (op.cit.) prove that

$$D(P \mid Q^{m^p}) = D(P \mid Q^m) - (E[\log R^{f\infty}] - E[\log R^f])$$

(11)

where $R^{f\infty}$ denotes the gross period return on a hypothetical infinite horizon zero coupon bond. (9) and (11) immediately yield the permanent component entropy bound:

$$D(P \mid Q^{m^p}) \geq E[\log R_i - \log R^{f\infty}] \approx E\left[\frac{R_i}{R^{f\infty}} - 1\right]$$

(12)

### 3.2 Illustrative Calculations

Here and throughout the rest of the paper, we use historical *monthly* data for US consumption (the earliest available month is January 1999), the S&P 500 and Fama-French HML (i.e. value factor portfolio) monthly total return series, and the Fama-French monthly riskfree return series between 1999-2014, and substitute time series averages for expectations where needed to provide estimates. The right hand side of (9) shows that the $D(P \mid Q^m)$ bound will be tightest when calculated using the asset with the highest (log gross return) risk
premium. With our data, that is the Fama-French HML portfolio, with a (log gross return) risk premium close to 33 bpm. Now let us expand the set of $N = 2$ test assets to estimate Bansal and Lehmann’s (1997) preferred choice: the “growth optimal” portfolio of the two, i.e. the portfolio of HML and the S&P 500 having the highest sample average log gross return. A portfolio with 19% of funds invested in the S&P 500 and the other 81% of funds invested in HML has the highest sample average log gross return of nearly 50.7 bpm. Subtracting the 17.3 bpm sample average log gross riskfree rate yields 33.4 bpm for the right hand side of (9).

The left hand side of (9) must be higher than 33.4 bpm, while remaining consistent with the unconditional expectation $E[m] = E[1/R_f]$ implied by (4). To calibrate parameters $\delta$ and $\alpha$ in (2) to be consistent with these two restrictions, we find values that solve the two nonlinear equations:

$$\frac{1}{T} \sum_{t=1}^{T} m_t(\delta, \alpha) = \frac{1}{T} \sum_{t=1}^{T} \frac{1}{R_f} = 0.9983$$

$$\log \left[ \frac{1}{T} \sum_{t=1}^{T} m_t(\delta, \alpha) \right] - \frac{1}{T} \sum_{t=1}^{T} \log m_t(\delta, \alpha) = 33.4 \text{ bpm} \quad (13)$$

The first equation requires that the historical time-series mean of $m(\delta, \alpha)$ in (2) equals the historical time-series mean of the inverse monthly riskfree return (which in turn equals the inverse of the mean monthly riskfree rate up to the sixth decimal point). The second equation requires that the historical time-series entropy of $m(\delta, \alpha)$ equals the aforementioned 33.4 bpm for the right hand side of (9).

The procedure described above finds values $\delta = 1.01$ and $\alpha = 3.45$. Increasing $\alpha$ beyond
that increases the entropy above 33.4 bpm (and thus still in accord with the entropy bound (9)), and necessitates increasing δ above 1.01 to restore consistency with the first equation in (13). In this sense, the admissible parameter range implied by the Jensen’s Inequality entropy bound is \( \delta \geq 1.01 \) and \( \alpha \geq 3.45 \).

Again using values \( \delta = 1.01 \) and \( \alpha = 3.45 \), the cumulant expansion (10) can be used to decompose the 33.4 bpm entropy. Only 2.5 bpm arises from half the second cumulant (i.e. variance) of \( \log m \), so the rest must come from the sum of the higher than 2nd order cumulants. Were \( \log m \) normally distributed, its higher order cumulants would all be zero. Hence the empirical finding that the sum of higher order cumulants accounts for much of the entropy is due to significant non-normalities in \( \log m \).

Finally, the decomposition into permanent and temporary components is governed by equation (11). One needs unobservable monthly returns from infinite horizon zero coupon bonds, denoted \( R_{\infty}^{f} \). To approximate the monthly yields for these hypothetical zero coupon bonds, we approximate the monthly yields of 30 year zero coupon bonds by use of the reported monthly series of 30yr. constant maturity yields, and then use that series to estimate the needed (albeit hypothetical) monthly returns \( R_{\infty}^{f} \). Using the bound-constrained admissible values \( \delta = 1.01 \) and \( \alpha = 3.45 \) described above, \( D(P \mid Q^{m}) = 33.4 \) bpm, while \( E[\log R_{\infty}^{f} - \log R^{f}] = 10.4 \) bpm, so (11) implies that the permanent component entropy \( D(P \mid Q^{m^{p}}) \approx 23 \) bpm, i.e. the permanent component accounts for about 2/3 of the entropy.
4 Optimized Entropic Diagnostics

Equation system (5) presents a finite number of linear equation constraints that generally underdetermine a risk-neutral probability measure $Q$ that solves it. This is yet another type of linear inverse problem solvable by entropic optimization. Stutzer (1995) derived the risk-neutral measure with minimal entropy relative to the data generating measure $P$. That is the solution of the convex optimization problem

$$Q^* = \arg \min_Q \left[ D(Q \mid P) \equiv E^Q[\log \frac{dQ}{dP}] \right] \quad \text{subject to} \quad (5). \quad (14)$$

As shown in Stutzer (1995), the solution $Q^*$ of problem (14) has the following exponentially tilted, a.k.a. Gibbs canonical density:

$$\frac{dQ^*}{dP} = \frac{e^{\sum_{i=1}^N w_i^* R_i^*}}{E[e^{\sum_{i=1}^N w_i^* R_i^*}]} \quad (15)$$

which is positive by construction, as a possible SDF density must be, i.e. it is a proper benchmark density. This density’s parameters $w_1^*, \ldots, w_N^*$ solve the following unconstrained, convex minimization

$$w^* = \arg \min_w \left( CGF^{R^*}(w) \equiv CGF \sum_i w_i R_i^* (1) \equiv \log E[e^{\sum_i w_i R_i^*}] \right) \quad (16)$$

as is seen by noting that the first order condition for (16) satisfies (5) under the measure $Q$ with density (15).

Finally, it can be verified by substitution that

$$D(Q^* \mid P) = -CGF^{R^*}(w^*) = -CGF \sum_i w_i^* R_i^* (1) \quad (17)$$
That is, the entropy of $Q^*$ relative to $P$ is just the absolute value of the objective function (i.e. the CGF) needed to find the minimizing $w^*$.

A model’s implied SDF $m(\theta)$ satisfying (5) defines a measure $Q^m$ with density $dQ^m = \frac{m(\theta)}{E[m(\theta)]} dP$. Its entropy relative to the data generating measure $P$ is $D(Q^m \mid P) \equiv E^{Q^m}[\log \frac{dQ^m}{dP}] = E \left[ \frac{m(\theta)}{E[m(\theta)]} \log \frac{m(\theta)}{E[m(\theta)]} \right]$. Because $Q^*$ in (15) achieves the minimum relative entropy among those in (5), a model’s implied $Q^m$ must satisfy the optimized entropy bound:

$$D(Q^m \mid P) \equiv E \left[ \frac{m(\theta)}{E[m(\theta)]} \log \frac{m(\theta)}{E[m(\theta)]} \right] \geq D(Q^* \mid P) \tag{18}$$

To illustrate a parametric calculation of the bound, suppose the excess return vector is multivariate normal. Then only its first two cumulants are nonzero. Problem (16) becomes:

$$w^* \equiv \arg \min_{w_1, \ldots, w_N} CGF^{R^e}(w)^{normal} = E[R^e]'w + \frac{1}{2}w'Cov[R^e]w \tag{19}$$

Set the first derivative equal to zero to find:

$$w^*^{normal} = -Cov[R^e]^{-1}E[R^e] \tag{20}$$

Substitute (20) into (19) and that into (17). A little algebra then yields the optimized entropy bound when excess returns are multivariate normal:

$$D(Q^* \mid P) = -CGF^{R^e}(w^*)^{normal} = \frac{1}{2}E[R^e]'Cov[R^e]^{-1}E[R^e] \tag{21}$$

which is proportional to the celebrated Hansen and Jagannathan (1991) variance bound on the analyzed model’s SDF $m$ with density $dQ^m$. In the special case of $N = 1$ test asset with excess return $R^e_1$, (21) becomes:

$$D(Q^* \mid P) = -CGF^{R^e}(w)^{uninormal} = \frac{1}{2} \left( \frac{E[R^e_1]}{\sqrt{Var[R^e_1]}} \right)^2 \tag{22}$$
which is half the squared Sharpe Ratio of the single test asset’s return. When excess returns aren’t multivariate normal, (21) and (22) are 2nd order approximations to the general bound (17). We thus see that popular variance-based diagnostics arise as a special case of the more general optimized entropy bound, with differences arising from the presence of higher than 2nd order cumulants in the multivariate excess returns of the test asset(s).

### 4.1 Permanent and Temporary Components of $m$

A decade after these $D(Q^m \mid P)$-based results were published in Stutzer (1995), Alvarez and Jermann (2005) used $D(P \mid Q^m)$ to derive the additional implications (11) and (12) arising from their decomposition of $m = m^P m^T$ into multiplicative permanent and temporary components. We now derive some analogous, new results for the optimized bound defined in the previous section.

Alvarez and Jermann (2005, Proposition 2) showed that the temporary component $m^T \equiv 1/R_f^{\infty}$. Then each moment condition $E[R_i m^P m^T] = 1$ can be rewritten as $E\left[\frac{R_i}{R_f^{\infty}} m^p\right] = 1$. Now divide both sides by $E[m^p]$ and perform the change of measure $dQ^p = \frac{m^p}{E[m^p]} dP$ to rewrite this as $E^{Q^p} \left[\frac{R_i}{R_f^{\infty}}\right] = \frac{1}{E[m^p]}$. Alvarez and Jermann (2005, Proposition 1) also showed that $E[m^p] = 1$, so that can be rewritten $E^{Q^p} \left[\left(\frac{R_i}{R_f^{\infty}} - 1\right)\right] = 0$. Comparing this to (5), we see that equations (15) - (18) will characterize an entropy bound $D(Q^{*p} \mid P)$ on the permanent component $m^p$, once we replace test assets’ excess returns $R_i^e := R_i/R_f^{\infty} - 1 \approx \log R_i - \log R_f^{\infty}$. The latter expression uses the approximation $\log 1 + x \approx x$ that is good enough for short period lengths (e.g. 1 month), but does not have to be used. Alvarez and Jermann (2005,
Example 1, p. 1981) note that when $C_{t+1}/C_t$ in (2) is IID in (2), all shocks are permanent and interest rates of any maturity are all constant and equal to each other, so that $R^f_\infty = R^f$ is the same constant. In that case, the required excess return $R^*_t := R_t/R^f_\infty - 1 = (1/R_f)(R_t - R_f)$. When substituted into (16) and then (17), this produces the same solution as before, because the constant $1/R_f$ gets absorbed into the minimizer $w^*$, i.e. $w^* := w^*/R_f$. As a (new) result, when $C_{t+1}/C_t$ in (2) is IID, the optimized entropy bound equals the permanent component entropy bound just derived, as it should be when all shocks are permanent.

4.2 An Illustrative Empirical Application

Using the same illustrative dataset employed in Section 3.2, the numerical solution of the historical time-series analog of (16) is $w^*_1 = -1.754$ and $w^*_2 = -3.51$, corresponding to the excess returns on the S&P 500 and Fama-French HML portfolios, respectively. The absolute value of the objective function in (16) is the estimated relative entropy (17), equal to about 97.3 bpm. To estimate the bound-admissible range for $\delta$ and $\alpha$ satisfying (18), we follow the same procedure used in Section 3.2, i.e. find values for $\delta$ and $\alpha$ satisfying the two equations:

\[
\frac{1}{T} \sum_{t=1}^{T} m_t(\delta, \alpha) = \frac{1}{T} \sum_{t=1}^{T} \frac{1}{R_t^f} \equiv .9983
\]

\[
\frac{1}{T} \sum_{t=1}^{T} \log \frac{m_t(\delta, \alpha)}{.9983} = 97.3 \text{ bpm}
\]

A numerical nonlinear equation solver found $\delta = 1.104$ and $\alpha = 30.347$. As in Section 3.2, increasing $\alpha$ increases the relative entropy above its lower bound, necessitating an increase in
$\delta$ to the first equality. *In this sense, the admissible parameter range implied by the optimized entropy bound is $\delta \geq 1.104$ and $\alpha \geq 30.347$.*

Again using values $\delta = 1.104$ and $\alpha = 30.347$, the cumulant expansion (17) can be used to decompose the 97.3 bpm entropy. Only 19.3 bpm arises from the first cumulant in the expansion (i.e. the mean of $w_1^*R_1^e + w_2^*R_2^e$) in the expansion, with only about -1 bpm arising from the the second cumulant (i.e. its variance), so most of the entropy must come from the sum of the higher than 2nd order cumulants. Were the (nonnormalized) portfolio excess return $w_1^*R_1^e + w_2^*R_2^e$ normally distributed, its higher order cumulants would all be zero. Hence the empirical finding that the sum of higher order cumulants accounts for much of the entropy is due to significant non-normalities in that excess return.

Finally, to estimate an entropy bound $D(Q^p \mid P)$ on the permanent component $m^p$ in the decomposition $m = M^p m^T$, the procedure in Section 4.1 requires substitution of both the S&P 500 and Fama and French HML excess returns by the their respective returns relative to $R_f^\infty$. After proxying the latter (unobservable) series as described in Section 3.1, we re-solve (16) and substitute into (17) to find the permanent component entropy bound $D(Q^p \mid P)$ is 23.6 bpm. A valid time series model for the permanent component $m^p$ should have an entropy larger than that. Note that this accounts for $23.6/97.3 \approx 24\%$ of the overall entropy bound $D(Q^* \mid P)$. 

16
5 Interpretation and Comparison of the Two Bounds

Both entropy bounds imply that the monthly “discount” factor $\delta$ is in excess of one, and hence would be better termed a “premium” factor, i.e. consumers prefer to delay gratification. Moreover, it is doubtful that the integral defining the expected discounted intertemporal utility will even converge when (the monthly) $\delta$ is much larger than one. But the range $\delta \geq 1.10$ implied by the optimized Entropy Bound is much more restrictive and damning for the theory underlying (2) than the range $\delta \geq 1.01$ implied by the Jensen’s Inequality Bound. Similarly, while the Jensen’s Inequality Bound only implies that the coefficient of risk aversion must exceed a plausible value $\alpha \geq 3.45$, the optimized Bound implied that $\alpha \geq 30.35$, a value far higher than that commonly thought sensible in other applications of power utility. Hence the optimized Bound illustrates the *Equity Premium Puzzle* while the Jensen’s Inequality Bound does not. The same quick calculations can be applied to any arbitrage-free asset pricing model, as each produces its own candidate for $m$.  

So for the purpose of quickly estimating a range of parameter values for which candidate SDF $m$ is consistent with returns from a set of test assets, the optimized bound produced a more restrictive range. This is because the Jensen’s Inequality Bound employed by Alvarez and Jermann (op.cit.), the Backus, et. al. and related papers does *not* compare the relative entropy of the analyzed model’s SDF density to its corresponding risk neutral constrained minimum achievable relative entropy, but instead compare it to a bound derived by the relatively weak force exerted by Jensen’s Inequality bias, which does not achieve this minimum minimorum.
Both entropies are evaluated using cumulant generating functions. The Jensen’s Inequality entropy is the candidate SDF log $m$ evaluated at $s = 1$, while the optimized entropy is the CGF of an optimized benchmark portfolio’s excess return evaluated at $s = 1$. In our illustrative application, both were largely determined by higher than 2nd order cumulants. Finally, permanent component entropy bounds exist for both entropies. In our illustrative application, the Jensen’s Inequality permanent component entropy was a much larger fraction of its total entropy than the optimized permanent component entropy was of its total entropy.

6 A Better Tool

Note from (14) that even if $Q^m$ satisfies (5), $Q^*$ in (14) will not equal $Q^m$ unless the latter is relative entropy-closest to the data generating measure $P$. There is no reason to expect that. Moreover, note from (18) that the associated entropy bound will be zero if and only if $Q^* \equiv P$, i.e. the real world is risk-neutral. As shown in Section 4, the optimized entropy bound is a generalization of the Hansen-Jagannathan (1991) variance bound, so the latter has analogous drawbacks. This led Hansen and Jagannathan to derive a different diagnostic by minimizing the “least squares distances between the stochastic discount factor proxy and families of stochastic discount factors that price correctly the vector of securities used in an econometric analysis.”(Hansen and Jagannathan (1997, p.558)). As a result, subsequent empirical work used this Hansen-Jagannathan (1997) specification error diagnostic, e.g. Campbell and Cochrane (2000) [5]. A relative entropy analog of the H-J
specification error diagnostic was published by the Journal of Econometrics in Kitamura and Stutzer (2002) [10]. We will exposit and develop uses for this tool below, which will show that it is more useful than the entropy bounds described earlier in this paper, much like the Hansen-Jagannathan (1997) specification error diagnostic has supplanted their variance bound diagnostic.

Replace the data generating measure $P$ in (14) by an analyzed model’s measure $Q^m$. Formally, we seek:

$$Q^{*m} = \arg \min_Q D(Q \mid Q^m) \text{ subject to (5).} \quad (24)$$

Note that the minimized value $D(Q^{*m} \mid Q^m)$ from (24) will be zero if and only if the analyzed model’s SDF measure $Q^m$ does satisfy (5), like the Hansen and Jagannathan (1997) specification error diagnostic. $D(Q^{*m} \mid Q^m) > 0$ is diagnostic of specification error, so let’s call it the specification error entropy. The solution to (24) is found by substituting $Q^m$ for $P$ in (15) - (17), requiring that expectations operators are defined over $Q^m$ rather than the data generating measure $P$. The results are:

$$\frac{dQ^{*m}}{dP} = \frac{e^{\sum_{i=1}^N w^{*m}_i R^e_i m}}{E[e^{\sum_{i=1}^N w^{*m}_i R^e_i m}]} = \frac{e^{\sum_{i=1}^N w^{*m}_i R^e_i + \log m}}{E[e^{\sum_{i=1}^N w^{*m}_i R^e_i + \log m}]} \quad (25)$$

$$w^{*m} \equiv \arg \min_{w_1, \ldots, w_N} \log E^{Q^m}[e^{\sum_{i=1}^N w_i R^e_i}]$$

$$= \arg \min_{w_1, \ldots, w_N} \log E[e^{\sum_{i=1}^N w_i R^e_i m}/E[m]]$$

$$= \arg \min_{w_1, \ldots, w_N} \log E[e^{\sum_{i=1}^N w_i R^e_i + \log m}] - \log E[m]$$
\[
= \arg \min_{w_1, \ldots, w_N} \log E[e^{\sum_{i=1}^{N} w_i R_i^e + \log m}] + \log E[R_f]
\] (26)

and the specification error entropy equals the absolute value of the minimized objective function, i.e.

\[
D(Q^*m \mid Q^m) = -CGF \sum_{i=1}^{N} w_i^* R_i^e + \log m (1) - \log E[R_f] \equiv -\sum_{j=1}^{\infty} \frac{\kappa_j \sum_{i=1}^{N} w_i^* R_i^e + \log m}{j!} - \log E[R_f]
\] (27)

where \(\kappa_j\) again denotes the \(j\) th cumulant of the superscripted random variable. While the Jensen’s Inequality entropy (10) is determined solely by the cumulants of \(\log m\), and the optimized entropy (17) is determined solely by the cumulants of an optimally chosen portfolio’s excess return, note from (27) that the specification error entropy is determined by the cumulants of (the sum of) both. A specification error diagnostic must depend on both in order to capture the inability of an analyzed model’s SDF \(m\) to satisfy (3), which depends on the test assets’ returns and the analyzed model’s SDF \(m\).

Ignoring 3rd and higher order cumulants (i.e. assuming normality) results in an index similar to the Hansen and Jagannathan (1997) specification error diagnostic. This 2nd order approximation is:

\[
D(Q^*m \mid Q^m) \overset{normal}{=} \frac{1}{2} E^{Q^m}[R^e]'Cov^{Q^m}[R^e]^{-1} E^{Q^m}[R^e]
\]

\[
= \frac{1}{2} E[R^e m/E[m]]'Cov^{Q^m}[R^e]^{-1} E[R^e m/E[m]]
\] (28)

If one replaces the covariance matrix in (28) by the matrix \(E[R^e R^e]\), this 2nd order approximation would be proportional to the Hansen and Jagannathan (1997) specification error diagnostic.

20
Because $Q^*m$ is the measure satisfying (5) that is relative entropy minimal from $Q^m$, 
\[
\frac{dQ^*m}{dQ^m} \sim e^{\sum_{i=1}^{N} w_i^m R_i^f} \text{ will deviate from unity in ways that characterize the analyzed model } m \text{'s inability to satisfy (3). Use of this diagnostic tool is illustrated in Section 6.2 below.}
\]

### 6.1 Permanent and Temporary Components of $m$

New results in section 4.1 explained the simple modification needed to formulate an optimized entropy bound on the permanent component of $m := m^p m^T$. The same technique works here to derive a new bound on the specification error of the permanent component. Define the density of a proposed permanent component $dQ^m = m^p dP$. The permanent component specification error index is then the minimized value in the following problem:

\[
\min_{Q^p} D(Q^p \mid Q^m) \quad \text{subject to } E[Q^p \left( \frac{R_i}{R_f} - 1 \right)] = 0 \quad i = 1, \ldots, N \quad (29)
\]

The solution is found by substitution of the redefined excess returns and density $dQ^m$ in (25) - (27):

\[
\frac{dQ^m}{dP} = \frac{e^{\sum_{i=1}^{N} w_i^m (R_i/R_f^\infty - 1) + \log m^p}}{E[e^{\sum_{i=1}^{N} w_i^m (R_i/R_f^\infty - 1) + \log m^p}]} \quad (30)
\]

where

\[
w_{*m^p} = \arg \min_{w_1, \ldots, w_N} \log E[e^{\sum_{i=1}^{N} w_i (R_i/R_f^\infty - 1) + \log m^p}] - (\log E[m^p] = 0) \quad (31)
\]

The permanent component specification error entropy is then the absolute value of the objective function in (31)

\[
D(Q^m \mid Q^m) = -CGF \sum_{i=1}^{N} w_i^m (R_i/R_f^\infty - 1) + \log m^p (1) \equiv - \sum_{j=1}^{\infty} \frac{\sum_{i=1}^{N} w_i^m (R_i/R_f^\infty - 1) + \log m^p}{j!} \quad (32)
\]
6.2 An Illustrative Empirical Application

Using our illustrative data set, we first calculate the specification error entropy at the boundary of the parameter space admissible under the entropy bounds developed in Sections 3 and 4. In Section 3.2, we found that $\delta = 1.01$ and $\alpha = 3.45$ satisfied the Jensen’s Inequality entropy bound. Substituting those values into the historical time series analog of (26) and numerically minimizing results in a 94 bpm specification error entropy, i.e. the absolute value of the minimized objective function in (26). In section 4.2, we found that $\delta = 1.10$ and $\alpha = 30.35$ satisfied the optimized entropy bound, with a specification error entropy of 70 bpm. Searching over the parameter space subject to the riskfree asset pricing constraint $E[m(\delta, \alpha)] = E[1/R_f]$, we find that the least specification error entropy is 59 bpm, attained by $\delta = 1.18$ and $\alpha = 55$. Theory constrains these two parameters to satisfy the $N = 2$ moment conditions (3) (for pricing the S&P 500 and the Fama-French HML) as well as the riskfree asset pricing constraint $E[m(\delta, \alpha)] = 1/E[R_f]$, so there is one over-identifying restriction. As such, it isn’t surprising that no parameter vectors reduce the sample specification error entropy to zero.

Because $Q^m$ is the measure satisfying (5) that has minimal entropy relative to $Q^m$, comparing the density $dQ^m$ from (25) to the density of the analyzed model’s SDF $m$ reveals the distributional deficiencies that prevent the latter from simultaneously pricing both the S&P 500 and Fama and French HML portfolios. Figure 1 contrasts histograms of the two densities. Note that the failed density of $Q^m$ is more concentrated than the density of $Q^m$. While both have similarly positive monthly skewness around 2.3 and excess kurtosis around
9.4, \(Q^m\) has a variance over 160 bpm higher than \(Q^m\), which due to the high value \(\alpha = 55\), itself has a variance of 732 bpm. These variances are more than three orders of magnitude higher than the 0.2 bpm variance of the consumption growth rate \(\frac{C_{t+1}}{C_t}\). Given the prior belief that \(\alpha\) should be much smaller, the main deficiency of (2) should be attributed to the lack of variance in the consumption growth rate, and not a mismatch of higher order cumulants.

The historical time series comparison of the two densities’ values is seen in Figure 2, which shows \(\frac{dQ^m}{dQ^*} \sim e^{\sum_i w^m_i R_t^i}\) for each month \(t\) in the 1999-2014 sample. The months in which the plotted value deviates most from unity are those most problematic for the analyzed model’s \(m\) to satisfy (3) in-sample. Because \(w^m_{S&P} \approx 0\) while \(w^m_{HML} = -3.12\), months with unusually high HML returns correspond to months where the plot is unusually low, and vice-versa. This is another facet of the Value Stock Anomaly: even if we parameterize (2) to satisfy the S&P 500 (i.e. growth portfolio) return constraint in (3), \(m(\theta)\) won’t satisfy the value portfolio return constraint in (3).

### 6.3 Summary

1. The specification error entropy is positive if and only if parameters \(\theta\) can’t be found to make a model-determined SDF \(m(\theta)\) satisfy the moment constraints (3), in which case it is zero.

2. The specification error entropy depends on the cumulants of the sum of the model-determined log \(m\) plus the excess return from a portfolio of the test assets. The Jensen’s
Inequality entropy of Section 3 depends only on the former, while the optimized entropy of Section 4 depends only on the latter.

3. The specification error entropy is the relative entropy generalization of the Hansen and Jagannathan (1997) specification error diagnostic, incorporating 3rd and all higher order cumulants.

4. Calculation of the specification error entropy requires determination of the measure satisfying (5) with minimal entropy relative to the analyzed model SDF $m$. Comparing the density of this measure to the density of the analyzed model’s SDF $m$ reveals statistics and observations that are most consequential for the failure of the SDF $m$ to satisfy (3).

7 Repairing an SDF: Ghosh, Juilliard and Taylor [9](2014)

Ghosh, Juilliard, and Taylor (2014) examine models in which a rejected consumption-growth based SDF $m^{o}$ is multiplied it by another component $\psi$, i.e.

$$ m = m^{o}\psi $$

(33)

where we now use the notation $m^{o}$ to denote the original component of the SDF that needs fixing.

Substitute (33) into (3) and fix $m^{o}$ to define a feasible set of all such $\psi$ that enable $m \equiv m^{o}\psi$ to satisfy (3). Keeping $m^{o}$ fixed, Ghosh, et.al.(2014) work with the change of
measure

\[ d\Psi = \frac{\psi}{E[\psi]} dP \] (34)

and rewrite the moment conditions as

\[ E^\Psi[R_i^o m^o] = 0 \quad i = 1, \ldots, N \] (35)

They devote considerable attention to the following measure and associated bound:

\[ \Psi^* = \arg \min_\Psi \left[ D(\Psi \mid P) \equiv E^\Psi[\log \frac{d\Psi}{dP}] \right] \quad \text{subject to } (35). \] (36)

The solution method corresponds to (15) - (17), specifically:

\[ \frac{d\Psi^*}{dP} = \frac{e^{\sum_{i=1}^N w_i^o R_i^o m^o}}{E[e^{\sum_{i=1}^N w_i^o R_i^o m^o}]} \] (37)

This density’s parameter vector \( w^* \) solves the following unconstrained, convex minimization of the CGF for \( R^o m^o \):

\[ w^* = \arg \min_w \left( \text{CGF}_{R^o m^o}(w) \equiv \log E[e^{\sum_i w_i R_i^o m^o}] \right) \] (38)

and

\[ D(\Psi^* \mid P) = -\text{CGF}\sum_i w_i^o R_i^o m^o \] (1) (39)

Note from (35) that if \( m^o \) itself satisfies (3), then \( \Psi^* \equiv P \) so the relative entropy (39) will be zero. This will be realized by finding \( w^* = 0 \) in (38), so the density (37) will collapse
to unity. Otherwise the density (37) identifies the “missing component” \( \psi^* = e^{\sum_{i=1}^N w_i^* R_i^m} \) that completes the benchmark

\[
m^\psi^* \equiv m^\psi^*.
\] (40)

The benchmark \( m^\psi^* \) from (40), which they dub the “\( \psi^* \)-filtered SDF”, is decomposed into the analyst-designated \( m^\psi \) and a minimal relative entropy multiplicative component needed to make (40) satisfy (3). The benchmark density derived by normalizing (40) serves a purpose analogous to that of \( \frac{dQ^* m^\psi}{dP} \) from Section 6 above, while the density (37) of the missing component serves a purpose analogous to that of \( \frac{dQ^* m^\psi}{dQ_m} \) developed in Section 6.

7.1 Comparison to Specification Error Entropy Diagnostics

In our illustrative example, suppose \( m^\psi \) is given by (2). We continue to use the \( N = 2 \) risky S&P 500 and HML portfolios, numerically minimizing the sample analog of (38) and substituting the minimizer into (37) to produce the density \( \frac{d\Psi^*}{dP} \). Its numerator is the multiplicative missing component \( \psi^* \) in (40) needed to make that satisfy (3).

In our illustrative example, there is very little – if any – information in these Ghosh, et. al. (2014) constructs that can not be provided by the aforementioned specification error entropy diagnostics developed in Section 6. To wit, there is a a 99% correlation between the time series of the density formed from the “\( \psi^* \)-filtered SDF” (40) and the density (25) for \( Q^* m^\psi \). Figure 3 overlays the two time series, which are indistinguishable. Not surprisingly, there is a 96% correlation between the time plots of the missing component \( \Psi^* \) density (37) and the analogous \( \frac{dQ^* m^\psi}{dQ_m} \sim e^{\sum_i w_i^* m^\psi R_i^m} \) from Section 6. Figure 4 shows that the two illustrate
the same phenomena. The results are not identical, because \( m^o \) enters the exponent in (37) and (40) multiplicatively, while \( \log m^o \) would enter the exponent in (25) and (26). But the two methods use different optimized values of the \( w \) vectors, which in our example, adjust for most differences.

In other applications, the two methods may provide different insights. Our application of the specification error entropy did not require that the SDF \( m \) has a multiplicative form (33). This would be a strength when the analyst either does not want to focus on the multiplicative decomposition, or when there is no naturally useful multiplicative decomposition. But this could be viewed as a weakness otherwise. Finally, note that \( m^o \) and \( \psi^* \) in (40) cannot represent permanent and temporary components as in Alvarez and Jermann (op.cit.), because Ghosh, et. al. (2014) do not impose the theoretical constraint \( E^{Q^o} \left[ \left( \frac{R_i}{R_{\infty}} - 1 \right) \right] = 0 \) in their relative entropy minimization, as we derived in Section 4.1.

7.2 Summary of Ghosh, et.al.(2014) \( \Psi \)-density

1. The missing component density \( \frac{d\Psi^*}{dP} \) plays a role analogous to \( \frac{dQ^{m^o}}{dQ^{m_{\infty}}} \) in the Kitamura and Stutzer (2003) specification error entropy further developed in Section 6, but the latter does not require the multiplicative decomposition \( m = m^o \psi \).

2. In our illustrative example, the \( \psi^* \)-filtered density is nearly identical to \( \frac{dQ^{m^o}}{dP} \) developed in Section 6.

3. As a result, associated diagnostics of the analyzed SDF \( m \)'s shortcomings add little value to the specification error entropy diagnostics in Section 6.
8 Conclusions

Relative entropy is a nonnegative, asymmetric, bivariate function of two absolutely continuous probability measures, that is zero if and only if the two measures coincide. Denote the data generating measure by \( P \), use \( m \) to denote a stochastic discount factor (SDF), and its corresponding risk-neutral density \( dQ^m \equiv \frac{m}{E[m]} dP \). The entropy of \( Q^m \) relative to \( P \) is

\[
D(Q^m \mid P) \equiv E^{Q^m}[\log \frac{dQ^m}{dP}] = E\left[\frac{m}{E[m]} \log \frac{m}{E[m]}\right],
\]

which simplifies to \( \sum_i Q^m_i \log \frac{Q^m_i}{P_i} \) when there only a finite number of states \( i \). This notion of entropy as an asset pricing diagnostic preceded the appearance of the argument-reversed \( D(P \mid Q^m) \equiv E[\log \frac{dP}{dQ^m}] = \log E[m] - E[\log m] \) as a different asset pricing diagnostic. Either way, the nonnegative bivariate function is zero if and only if the two measures coincide. Computation of \( D(Q^m \mid P) \) is equivalent to evaluating the cumulant generating function of an optimally chosen portfolio of the test assets, while computation of \( D(P \mid Q^m) \) is equivalent to evaluating the cumulant generating function of \( \log m \).

Using \( \mathbf{R} \) to denote a vector of test assets’ gross returns, the familiar SDF moment condition system \( E[\mathbf{R}m] = \mathbf{1} \) has been used to derive separate lower bounds on the two entropies. Violation of those bounds is diagnostic of model failure, i.e. failure of an asset pricing model’s SDF \( m \) to satisfy the moment system. Because an asset pricing model’s SDF \( m \) often depends on a vector of parameters \( \theta \), the bounds restrict the range of parameters to values consistent with them. Bounds determined by optimization using \( D(Q^m \mid P) \) (i.e. those in Stutzer (1995), Kitamura and Stutzer (2002) and Ghosh, et.al. (2014)) are more restrictive and hence more useful than bounds determined by Jensen’s Inequality using \( D(P \mid Q^m) \), e.g.
in Bansal and Lehmann (1997), Alvarez and Jermann (2005), Backus, Chernov, and Martin (2011), and Backus, Chernov, and Zin (2014).\textsuperscript{7}

But while these entropy bound diagnostics can quickly reveal problematic features, they are not an end in themselves. In assessing the value of the closely-related and now venerable Hansen-Jagannathan (1991) variance bound, Cochrane and Hansen (1992, p.129) concluded that

...the bounds tests “are not a substitute for directly testing the pricing implications of a model.”

The Cochrane and Hansen (op.cit.) assessment of the Hansen-Jagannathan variance bound is an equally astute observation about the entropy bounds. The literature shows that once enough test assets are used to form moment conditions, all asset pricing models fail the standard specification tests, e.g. the Hansen GMM J-Statistic. So Hansen and Jagannathan (1997) developed a separate variance-based specification error diagnostic to facilitate comparison of failed models. That specification error diagnostic is also variance-based, depending on both the test assets’ returns $R$ and the analyzed model’s SDF $m$. An entropic generalization of that index was defined in Kitamura and Stutzer (2002, Diagnostic Criterion 3.2), and further developed and applied in Section 6 herein. Like the Hansen-Jagannathan (1997) specification error diagnostic, the specification error entropy is zero if and only if the analyzed model’s SDF $m$ satisfies the $E[Rm] = 1$, and is positive otherwise, indicative of model failure. Unlike the Hansen-Jagannathan (1997) specification error diagnostic, it also depends on higher order moments in addition to the first two, and in fact is computed as
a cumulant generating function (CGF) of an optimally-chosen portfolio of the test assets
\[+ \log m, \text{ as it must to capture all the variables present in } E[R_m] = 1. \]
Unfortunately, the two entropy bound diagnostics mentioned above depend solely on one or the other, so they do not have the useful specification error index properties of the H-J (1997) diagnostic nor the specification error entropy in Kitamura and Stutzer (2002, Diagnostic Criterion 3.2) further developed in Section 6 herein. The Ghosh, et.al.(2014) \(\Psi_2\)-bound and its associated diagnostics serve a similar function, but require an additional assumption that the SDF \(m\) can be decomposed into the product of two random variables that may or not be components of the structural model being analyzed.

Other entropic bounds could be derived from permutations not described above, e.g. by constrained minimization of \(D(P \mid Q)\) over moment-constrained \(Q\), rather than the same minimization of \(D(Q \mid P)\) used in Section 4 herein. This and related permutations are examined within Ghosh, et.al. (2014), but they found that there was little value-added from them. Ghosh, et.al.(2014,p.11) conclude that:

“Since we have proposed two different relative entropy minimization approaches, we get two different estimates of the most likely SDF given the data. Asymptotically, the two should be identical given the MLE property of these procedures, nevertheless in any finite sample they could potentially be very different. As shown in our empirical analysis, the two estimates are very close to each other, suggesting that their asymptotic behaviour is well approximated in our sample.”

Finally, while we have seen that the various entropy diagnostics are informative about
some shortcomings of an analyzed model’s SDF \( m \), they will not revolutionize the econometrics of asset pricing in the way that the entropic methods of Boltzmann and Gibbs revolutionized statistical physics. Arnold Zellner [1991, p.21] , an early supporter of entropic econometrics, wrote that:

“...much more empirical and theoretical work needs to be done to get a ‘maximum entropy, thermodynamic model’ that performs well in explaining and predicting the behavior of economic systems.”

We hope that readers will join us in accepting Prof. Zellner’s challenge.

Notes

\(^0\) Correspondence to: Prof. Michael Stutzer, Department of Finance, University of Colorado, Boulder, Colorado michael.stutzer@colorado.edu.

\(^1\) In our illustrative data set, the two expressions for the expected inverse riskfree rate agree up to the sixth decimal place. Backus, Chernov, and Zin (2014, p. 59) agree that the difference is “usually small”, after showing that it accounts for the sole difference between their entropy bound and that implied years earlier by Alvarez and Jermann (2005), and by Bansal and Lehmann \(^2\) (1997) years before that. The latter paper does not use the term “entropy” when describing the concept.

\(^2\) The Cumulant Generating Function CGF of \( X \) evaluated at \( s \), a.k.a. the log moment
generating function, is defined by $CGF_X(s) \equiv \log E[e^{sX}]$. When $X$ is a vector, $s$ is a comformable vector producing the inner product $sX$.

3Alvarez and Jermann (2005) note that the long end of the yield curve is relatively flat, justifying their own use of 13 and 29 year yields as approximations.

4Much as in the case of the minimum norm linear projection producing the Hansen-Jagannathan (1991) variance bound, the solution to (14) exists when the set of measures satisfying (5) is closed (in the variation metric)(see Csiszar [8](1975, Thm. 2.1)).

5This and related diagnostics are not intended to substitute for formal econometric parameter estimation and tests of the moment constraints or other theoretical implications. For example, Cochrane and Hansen [7] (1992, p.129) stated that the variance bound diagnostics “are not a substitute for directly testing the pricing implications of a model”. As such, there isn’t much reason to derive interval estimates to be used for formal testing of the hypothesis that the bounds are violated, although that can be done either analytically or by employing a bootstrap, e.g. see the discussion in Stutzer (1995, Sec.3.3).

6This insight would not be revealed by the cumulant decomposition of Jensen’s Inequality entropy (10) that Backus, Chernov, and Martin (2011) extensively employ; see their Figures 1, 2, and 4 and their Table 3. At the (a-priori implausible) but best-fitting parameter values $\delta = 1.18$ and $\alpha = 55$ found above, the Jensen’s Inequality entropy (9) is 81 bpm. The first term in its cumulant expansion (10) is half the variance of $\log m$, which accounts for only 38 of the 81 bpm. While we would concur with their analysis that most of the Jensen’s Inequality
entropy results from cumulants higher than 2nd order, we fail to see how that provides more
insight than our analysis above. In fact, skewness and kurtosis are *standardized* cumulants
that finance researchers have much experience interpreting. Kurtosis is the 4th cumulant
divided by the square of the 2nd cumulant (i.e. divided by the square of the variance), while
skewness is the 3rd cumulant divided by variance raised to the $3/2$ power). Statisticians
haven’t been remiss in using standardized cumulants as summary statistics characterizing
deviation from normality. Indeed, Backus, Chernov, and Martin (op.cit.) do make use of
conventional skewness and kurtosis statistics in addition to their cumulant decompositions,
and despite repeated use of (9) and (10) in their text, the word “entropy” appears only
once in their 2 1/2 pages of concluding remarks. Nonetheless, readers who like cumulant
decompositions of entropies are invited to apply and compare our alternatives (17) and (27)
to their use of (10).

$^7$The $D(P \mid Q^m)$ entropy is also utilized in Bakshi and Fousseni [1](2014) and in Liu [12]
(2013) to provide bounds on powers of $m$.  

33
Figure 1: Comparison of Densities $dQ^m$ to $dQ^{*m}$

The analyzed model $m$ is too sharply peaked to serve as an SDF.
Figure 2: Time Series of $dQ^*/dQ^m$
Unusually low values on dates with unusually high HML returns illustrate the "Value Stock Anomaly"
Figure 3: Time Plot of $dQ^m/dP$ Overlayed with Ghosh, et.al. (2014) Psi-Filtered SDF Density: The two plots are nearly identical
Figure 4: Time Plots of $dQ^*/dQ^m$ vs. $d\Psi^*/dP$:

They convey the same information.
References


