Economic Implications of Nonlinear Pricing Kernels

Caio Almeida *
FGV/EPGE

René Garcia †
Edhec Business School

November 13, 2015

Abstract

Based on a family of discrepancy functions, we derive nonparametric stochastic discount factor (SDFs) bounds that naturally generalize variance (Hansen and Jagannathan, 1991), entropy (Backus, Chernov and Martin, 2011), and higher-moment (Snow, 1991) bounds. These bounds are especially useful to identify how parameters affect pricing kernel dispersion in asset pricing models. In particular, they allow us to distinguish between models where dispersion comes mainly from skewness from models where kurtosis is the primary source of dispersion. We analyze the admissibility of disaster, disappointment aversion and long-run risk models with respect to these bounds.

Keywords: Stochastic Discount Factors, Information-Theoretic Bounds, Robustness, Minimum Contrast Estimators, Implicit Utility Maximizing Weights.

JEL Classification: C1,C5,G1

---

*Email: calmeida@fgv.br, Graduate School of Economics, FGV/EPGE- Escola Brasileira de Economia e Financas, Rio de Janeiro, Brazil.

†Email: rene.garcia@edhec.edu, Address for correspondence: Edhec Business School, 393, Promenade des Anglais, BP 3116, 06202 Nice Cedex 3. We are grateful to Alastair Hall, Raymond Kan, Marcel Rindisbacher, and Allan Timmermman for their insightful discussions at the 2008 Financial Econometrics Conference at Imperial College, NFA 2008 in Calgary, EFA 2008 in Athens, and AFA 2009 in San Francisco, respectively. We also thank Yacine Aït-Sahalia, Rafael Azevedo, Diego Brandão, Tolga Cenesizoglu, Mikhail Chernov, Magnus Dahlquist, Darrell Duffie, Robert Engle, Olesya Grishchenko, Lars Peter Hansen, Daniela Kubudi, Anthony Lynch, Nour Meddahi, Stefan Nagel, Stijn Van Nieuwerburgh, Liuren Wu and seminar participants at the 2007 CIREQ Conference on GMM, 2008 Inference and Test in Econometrics Conference in Marseille, SoFiE 2008 Conference in New York, 2008 ESEM in Milan, 2011 Measuring Risk Conference at Princeton, QFE seminar at NYU-Stern, 2013 NASM of the Econometric Society, EDHEC Business School, Getulio Vargas Foundation, Stanford Financial Mathematics Seminar, Stockholm School of Economics, Warwick Business School, and Baruch College for their useful comments. Kym Ardison provided excellent research assistance. A previous version of this paper circulated under the title "Empirical Likelihood Estimators for Stochastic Discount Factors". The first author thanks CNPq-Brazil for financial support. He is also thankful to Darrell Duffie and Stanford Graduate School of Business for their hospitality. The second author is a research Fellow of CIRANO and CIREQ.

1 Introduction

Observed asset returns provide information about how future cash flows are discounted. This is the fundamental insight of Hansen and Jagannathan (HJ, 1991), who derived a minimum variance stochastic discount factor (SDF) bound. The SDF is obtained by minimizing a quadratic norm involving the first two moments of observed payoffs, resulting in a linear projection on the space of observed payoffs.

While very useful, SDFs obtained by linear projections may not be informative enough to diagnose asset pricing models. This becomes especially true for models whose pricing kernel dispersion is generated by nonlinearities in the kernel or non-Gaussianity in returns. In such cases, higher moments of the kernel play an important role. Taking into account these more complex cases, Backus, Chernov and Martin (2011) suggest analyzing asset pricing models with a combination of entropy (as a measure of dispersion) and the cumulant-generating function, to assess how higher moments affect such dispersion.[1]

Entropy considers a specific combination of SDF moments that gives similar weights to pairs of odd and even moments in the space of SDFs.[2] Therefore, allowing for more distinct weights across these two sets of moments (odd versus even) might be helpful to better identify and separate the effects of skewness (odd moments) from kurtosis (even moments) on pricing kernel dispersion. Building on this point, our main contribution is to propose a new family of nonparametric SDF bounds that puts different sets of weights on higher moments of SDFs and therefore complements Backus, Chernov and Martin (2011). In particular, we show that the new bounds bring additional non-redundant information when analyzing asset pricing models.

Given a set of basis assets payoffs, we minimize general convex functions of SDFs called Minimum Discrepancy (MD) measures (Corcoran, 1998) in order to obtain a projected nonlinear SDF that prices exactly a set of selected basis assets. Our new SDF bounds naturally generalize the original HJ variance bounds, entropic bounds (Stutzer, 1995 ; Bansal and Lehmman, 1997) and the extended higher-moment bounds proposed by Snow (1991). Each MD information bound generates as a byproduct a strictly positive SDF that correctly prices the primitive assets and that incorporates information about moments of returns higher than the variance.

[1]Martin (2012) suggests cumulant-generating functions as a way to assess the importance of higher moments of consumption growth in consumption-based models with disaster risk. Backus, Chernov and Zin (2014) consider entropy and a new measure of horizon dependence that captures dynamics, when analyzing sources of dispersion in representative agent models.

[2]By a pair of moments we define two neighbors, like for instance, the third and fourth moments.

The solutions for these SDFs are obtained through dual problems that are easier to solve than the primal problems and offer a nice economic interpretation. Each primal minimum discrepancy problem corresponds to a dual optimal portfolio problem, with the maximization of a specific utility function in the Hyperbolic Absolute Risk Aversion (HARA) family. Therefore the duality results stressed in HJ (1991), where maximizing the Sharpe ratio in the space of excess returns corresponds to finding a minimum variance in the space of SDFs, naturally carry out for the whole family of MD bounds.[3] The first-order conditions for these HARA optimization problems imply the nonlinear and positive SDFs mentioned above.

We illustrate the usefulness of our approach by diagnosing several asset pricing models featured recently in the literature. We analyze the admissibility of disaster models, long-run risk models, and models with disappointment aversion preferences, which are now pervasive in the consumption-based asset pricing literature.[4] In disaster models (Barro, 2006), dispersion comes primarily from asymmetric negative jumps on consumption growth that introduce positive skewness on the pricing kernel. Long Run Risk models (Bansal and Yaron, 2004) depart from the basic CCAPM by introducing persistence in consumption growth and time-varying, persistent volatility. Nevertheless, their implied pricing kernel, which comes from an approximate solution to the equilibrium problem, is log-normal making kurtosis an important source of dispersion. We also include in our analysis the demand side model (Albuquerque, Eichenbaum and Rebello, 2012), an extension of the traditional long run risk model that adds preference shocks correlated with consumption and dividends, with the objective to reduce the magnitude of risk aversion compared with Bansal and Yaron (2004). Finally, we consider a prominent recursive utility function recently adopted in the context of long run risk models that features disappointment aversion (Routledge and Zin, 2010). By looking at these models we show how intrinsically different generating mechanisms for dispersion produce distinct diagnoses within our family of bounds. In particular, too much positive skewness in the pricing kernel will make it harder for it to pass some of our extreme bounds, while kurtosis will have the opposite effect. Therefore, our bounds impose a data-driven balance between the amount of skewness and kurtosis that any admissible pricing kernel should satisfy.

[3] Our approach encompasses the exponential tilting (ET) criterion of Stutzer (1995) and its corresponding optimum portfolio of a CARA investor, as well as the empirical likelihood (EL) criterion and its corresponding log utility maximizing portfolio, denominated growth portfolio by Bansal and Lehmann (1997).

[4] In this paper, we do not provide statistical tests based on the probability distribution of the information bounds. For the asymptotic distribution of the proposed information bounds see Almeida and Garcia (2012); for the asymptotic distribution of the HJ bounds see Hansen, Heaton and Luttmer (1995), and Kan and Robotti (2015) for finite sample properties.

Our implied nonlinear SDFs are related to a number of previous studies that feature nonlinear SDFs. Bansal and Viswanathan (1993) propose a neural network approach to construct a nonlinear stochastic discount factor that embeds specifications by Kraus and Litzenberger (1976) and Glosten and Jagannathan (1994). Our approach provides a family of SDFs given by different hyperbolic functions of basis assets returns implied by portfolio problems. In Dittmar (2002), who also analyzes nonlinear pricing kernels, preferences restrict the definition of the pricing kernel. Under the assumption of decreasing absolute risk aversion, he finds that a cubic pricing kernel is able to best describe a cross-section of industry portfolios. Our nonparametric approach embeds such cubic nonlinearities implicitly. Although not based on preferences, our pricing kernels are also consistent with dual HARA utility functions that can exhibit decreasing absolute risk aversion and decreasing absolute prudence.[5]

Our nonparametric information bounds are also related to a number of studies. Stutzer (1995) suggests a nonparametric bound to test asset pricing models based on the minimization of the Kullback Leibler Information Criterion (KLIC). Bansal and Lehman (1997) propose a related entropic bound that is obtained by maximizing the growth portfolio. This bound generates the measure of entropy adopted by Backus, Chernov and Martin (2011) in tests of disaster-based models. In a recent paper, Ghosh, Julliard and Taylor (2012) propose a class of asset pricing models whose SDFs can be factorized into an observable component (a parametric function of consumption) and an unobservable nonparametric one, and exploit this decomposition to derive new entropic bounds that are obtained based on either the ET or the EL criteria. The bounds proposed by Stutzer (1995) and Bansal and Lehman (1997) are particular elements of our proposed family of nonparametric bounds. Those obtained by Ghosh, Julliard and Taylor (2012), although based on two specific members of the Cressie Read family (ET and EL), take into account an observed component of the SDF combined with a nonparametric function of basis assets returns, while ours are a pure nonparametric function of the basis assets returns.

A significant literature aims at sharpening the variance bounds by conditioning on information available to economic agents. Gallant, Hansen and Tauchen (1990) derive an optimal variance bound when the first two conditional moments are known, while Bekaert and Liu (2004) propose

[5] Dittmar (2002) starts with an approximation of an unknown marginal utility function by a Taylor series expansion but restricts the polynomial terms in the expansion by imposing decreasing absolute prudence (Kimball, 1993). Therefore, the risk factor obtains endogenously from preference assumptions and is a sole function of aggregate wealth. Our SDFs come from solutions to dual optimal HARA portfolio problems that endogenously determine aggregate wealth as a linear combination of a predetermined set of basis assets. These solutions potentially satisfy the desirable properties of decreasing absolute risk aversion (Arditti, 1967) and decreasing absolute prudence.

an optimally-scaled bound which is valid even when the first and second conditional moments are misspecified. Chabi-Yo (2008) introduces higher moments of returns and conditional information in volatility bounds by finding the SDFs that are linear functions of payoffs and squared payoffs (volatility contracts).[6] In contrast, the discrepancy measures we propose in this paper put weights on all moments of the distribution of returns in the dual optimization problem. Moreover, by considering a family of discrepancy measures, we add robustness to our diagnosis since each discrepancy puts different weights on the various moments of returns.

The rest of the paper is organized as follows. In section 2, we describe how the minimum discrepancy SDFs are derived and how the corresponding bounds are constructed. In Section 3, we assess the disaster, long-run risks, and disappointment aversion models with our discrepancy-based information frontiers. Section 4 concludes.

2 Minimum Discrepancy Stochastic Discount Factors

Let $(\Omega, \mathcal{F}, P)$ be a probability space, and R denote a K-dimensional random vector on this space representing the returns of K primitive basis assets. In this static setting, an admissible SDF is a random variable m for which $E(mR)$ is finite and satisfies the Euler equation:

$$E(mR) = 1_K, \tag{1}$$

where 1_K represents a K-dimensional vector of ones.

As in Hansen and Jagannathan (1991), we are interested in the implications of Equation (1) for the set of existing SDFs. Imagining a sequence of (m_t, R_t) that satisfies Equation (1) for all t, and observing a time series $\{R_t\}_{t=1,\ldots,T}$ of basis assets returns, we assume that the composite process (m_t, R_t) is sufficiently regular such that a time series version of the law of large numbers applies.[7] Therefore, sample moments formed by finite records of measurable functions of (m_t, R_t) will converge to population counterparts as the sample size T becomes large.

In such context, Hansen and Jagannathan (1991) find a minimum variance SDF by minimizing

[6] Kan and Zhou (2006) tighten the HJ bound by assuming that the pricing kernel is a reduced-form function of a finite set of state variables.

[7] For instance, stationarity and ergodicity of the process (m_t, R_t) are sufficient (see Hansen and Richards, 1987). In addition, we further assume that all moments of returns R are finite in order to deal with general entropic measures of distance between pairs of stochastic discount factors.

a quadratic function in the space of nonnegative admissible SDFs with fixed mean a:[8]

$$m^*_{HJ} = arg \min_m E[m^2],$$
$$\text{subject to } E\left[m\left(R - \tfrac{1}{a}1_K\right)\right] = 0_K,\ E[m] = a, m \geq 0. \tag{2}$$

They showed that the SDF solving Equation (2) is a linear combination of the original returns, truncated at zero.[9] HJ (1991) also solve an unconstrained minimum variance problem where they search within the set of admissible SDFs possibly assuming negative values in some states. In this case, the solution is simply a linear combination of the original returns. It has been used in many papers in the financial literature to impose minimum variance restrictions to pricing kernels implied by asset pricing models.

In this paper, we propose alternative moment restrictions to pricing kernels by using a convex and homogeneous discrepancy function $\phi(m)$. Therefore, we search for a Minimum Discrepancy (MD) SDF that solves the following minimization problem in the space of nonnegative (or strictly positive) admissible SDFs that present well-defined $E[\phi(m)]$:[10]

$$m^*_{MD} = arg \min_m E[\phi(m)],$$
$$\text{subject to } E\left[m\left(R - \tfrac{1}{a}1_K\right)\right] = 0_K,\ E[m] = a, m \geq 0 \text{ (or } m >> 0). \tag{3}$$

The general discrepancy function ϕ will imply a bound that restricts a particular combination of moments of admissible SDFs. Such restrictions will allow us to diagnose asset pricing models by going beyond the minimum variance bound.

Note that while Hansen and Jagannathan have the nonnegative restriction $m \geq 0$ in (2) we have either $m \geq 0$ or $m >> 0$ in (3). The domain of the function $\phi, [0, \infty) \rightarrow \Re$, whether it includes 0 or not, will determine which condition to use. The distinction is crucial since it is related to the theoretical condition of no-arbitrage in the market. When zero is not included, the existence of at least one admissible strictly positive SDF relies on a condition of no-arbitrage among the original primitive returns. Therefore, we assume this absence of arbitrage.[11]

[8] The set of admissible SDFs will depend on the market structure. The usual case when dealing with the above-mentioned observed time series of vector R, is to have an incomplete market, i.e., the number of states of nature (T) larger than the number of basis assets K. In such a case, an infinity of admissible SDFs will exist.

[9] For a detailed analysis on the HJ bounds with nonnegativity constraints, see Kan and Robotti (2015).

[10] By well-defined we mean $E[\phi(m)] < \infty$.

[11] However, in any specific sample of returns, there might exist in-sample arbitrages (see Gospodinov, Kan and

The minimization problem in (3) is based on an infinite-dimensional space.[12] In the next theorem, we make use of results in Borwein and Lewis (1991) to prove that, in general, problem (3) can be solved in a simpler finite dimensional dual space.

Theorem 1. *Consider the primal problem:*

$$\min_{m} E[\phi(m)],$$
$$\text{subject to } E\left[m\left(R - \tfrac{1}{a}1_K\right)\right] = 0_K,\ E[m] = a, m \geq 0 \ (\text{or } m >> 0). \tag{4}$$

and the dual problem:

$$\sup_{\alpha\in\Re,\lambda\in\Re^K} a * \alpha - E\left[\phi^{*,+}\left(\alpha + \lambda'\left(R - \frac{1}{a}1_K\right)\right) + \delta([\alpha\ \lambda]|\Lambda(R))\right], \tag{5}$$

where $\Lambda(R) = \{\tilde{\alpha} \in \Re, \tilde{\lambda} \in \Re^K : (\tilde{\alpha} + \tilde{\lambda}'\left(R - \frac{1}{a}1_K\right)) \in dom\ \phi^{*,+}\}$,[13] $\delta(.|C)$ *represents a set indicator function in Rockafellar's (1970) sense,*[14] *and* $\phi^{*,+}$ *denotes the convex conjugate of* ϕ*:*

$$\phi^{*,+}(z) = \sup_{w\in[0,\infty)\cap domain\ \phi} zw - \phi(w). \tag{6}$$

Absence of arbitrage implies that the values of the primal and the dual problems coincide (with dual attainment). A sufficient condition allowing the Minimum Discrepancy SDF to be obtained from the solution of the dual optimization problem is that either $d = lim_{x\to\infty}\frac{\phi(x)}{x} = \infty$ *or* ($d < \infty$ *and* $c = lim_{x\to\infty}((d - \phi'(x))x) > 0$)*. In such cases, the implied SDF is obtained by:*

$$m^*_{MD} = \frac{\partial\phi^{*,+}(z)}{\partial z}_{z=\left(\alpha^*+\lambda^{*\prime}\left(R-\frac{1}{a}1_K\right)\right)} \tag{7}$$

with

$$[\alpha^*\ \lambda^*] = arg \sup_{\alpha\in\Re,\lambda\in\Re^K} a * \alpha - E\left[\phi^{*,+}\left(\alpha + \lambda'\left(R - \frac{1}{a}1_K\right)\right) + \delta([\alpha\ \lambda]|\Lambda(R))\right], \tag{8}$$

Proof: See Appendix.

In the above theorem, λ is a vector of K Lagrange Multipliers that comes from the Euler equations for the primitive basis assets. The Lagrange Multiplier α comes from the original

Robotti (2014)) that prevent the existence of a strictly positive admissible SDF for that sample. We will discuss this assumption further when we look at the sample versions of our MD problems.

[12] The space of strictly positive SDFs is equivalent to the space of risk-neutral measures, while the space of non-negative SDFs contains the space of strictly positive SDFs.

[13] We define as domain of $\phi^{*,+}(z)$, the values of z for which the function is finite ($\phi^{*,+}(z) < \infty$).

[14] That is, $\delta(x|C) = 0$, if $x \in C$, and ∞, otherwise.

restriction $E(m) = a$ and can be concentrated out of the optimization problem (see the proof of the theorem). The nonnegativity (or positivity) restriction $m \geq 0$ (or $m >> 0$) on the original primal problem restricts the convex conjugate to be calculated on the nonnegative (or positive) real line in Eq. (6). The delta function $\delta(.|\Lambda(R))$ restricts, for each vector of returns R in the probability space, the optimization problem to a subset $\Lambda(R)$ of $\Re^K$ where the convex conjugate assumes finite values. Most importantly, from the theorem above we see that no-arbitrage is a fundamental condition to make sure that the solutions of the primal and dual problems will coincide.[15]

To arrive at empirical estimates of minimum discrepancy SDFs, we choose the Cressie-Read (1984) family of discrepancies defined as:

$$\phi^{\gamma}(m) = \frac{(m)^{\gamma+1} - a^{\gamma+1}}{\gamma(\gamma+1)}, \gamma \in \Re. \tag{9}$$

This family embeds as particular cases restrictions on the space of SDFs derived by HJ (1991), Snow (1991), Stutzer (1995), Bansal and Lehmann (1997) and Cerny (2003).[16]

This family has several advantages. First, restrictions coming from ϕ^{γ}'s for large negative and small positive values of γ will imply rich combinations of SDF moments and will allow us to better put forward the strengths and weaknesses of complex asset pricing models (in particular disaster and long-run risk models). Second, it offers a nice economic motivation to our information theoretic minimization problems since solving the latter will be equivalent to solving dual HARA utility maximization problems (see section 2.1). Third, it has been recently adopted in the econometric literature to build one-step alternatives to Generalized Method of Moments (GMM) estimators with useful higher-order properties (see Newey and Smith, 2004 and Kitamura, 2006), in particular their consistency in their sample form (see subsection 2.2).

With this family of discrepancies, we characterize in the following corollary the dual problem to be solved.

Corollary 1. *Let the discrepancy in the minimization problem (3) belong to the Cressie Read family:* $\phi^{\gamma}(m) = \frac{m^{\gamma+1} - a^{\gamma+1}}{\gamma(\gamma+1)}$ *with* $\gamma \in \Re$*, and assume that there is no-arbitrage in the economy,*

[15] In fact, no-arbitrage guarantees the existence of an interior point in the space of admissible nonnegative SDFs. Moreover, any strictly positive admissible SDF will also be in the interior of the space of strictly positive admissible SDFs, since this is an open set by the continuity of the linear pricing operator $E(.)$.

[16] When dealing with a specific Cressie Read discrepancy ϕ^{γ}, condition $E(\phi^{\gamma}(m)) < \infty$ is equivalent to the existence of the moment $E(m^{\gamma+1})$. In addition, for $\gamma \leq -1$, condition $E(\phi^{\gamma}(m)) < \infty$ implies that the minimization is restricted to the space of strictly positive admissible SDFs.

such that there exists at least one strictly positive admissible SDF. Then, letting $\Lambda_{CR}(R) = \{\lambda \in \Re^K,\ s.t.\ \left(a^\gamma + \gamma\lambda'\left(R - \frac{1}{a}1_K\right)\right) > 0\}$*:*

i) if $\gamma > 0$*, (8) specializes to:*

$$\alpha^*_\gamma = \frac{a^\gamma}{\gamma},\ \lambda^*_\gamma = arg\ \sup_{\lambda \in \Re^K}\ E\left[\frac{a^{\gamma+1}}{\gamma+1} - \frac{1}{\gamma+1}\left(a^\gamma + \gamma\lambda'\left(R - \frac{1}{a}1_K\right)\right)^{\left(\frac{\gamma+1}{\gamma}\right)} I_{\Lambda_{CR}(R)}(\lambda)\right] \tag{10}$$

ii) if $\gamma < 0$*, it specializes to:*

$$\alpha^*_\gamma = \frac{a^\gamma}{\gamma}, \lambda^*_\gamma = arg\ \sup_{\lambda \in \Re^K}\ E\left[\frac{a^{\gamma+1}}{\gamma+1} - \frac{1}{\gamma+1}\left(a^\gamma + \gamma\lambda'\left(R - \frac{1}{a}1_K\right)\right)^{\left(\frac{\gamma+1}{\gamma}\right)} - \delta(\lambda|\Lambda_{CR}(R))\right] \tag{11}$$

iii) if $\gamma = 0$*, the maximization is unconstrained:*

$$\alpha^*_0 = 1 + ln(a),\ \lambda^*_0 = arg\ \sup_{\lambda \in \Re^K}\ a - aE\left[e^{\lambda'\left(R - \frac{1}{a}1_K\right)}\right], \tag{12}$$

where $I_A(.)$ *represents a set indicator function in the usual sense.*[17]

Proof: See Appendix.

Let us explain why there are three different conditions depending on the value of γ. In (6), in order to calculate the convex conjugate $\phi^{*,+}(z)$ at a certain point z, we need to solve for a w that satisfies the first order condition: $g^z(w) = z - \phi'(w) = 0$. When for a given z, there is no solution to $g^z(w) = 0$, for $\gamma > 0$, $g^z(w)$, as a function of w, will be strictly negative implying $\phi^{*,+}(z) = 0$, justifying the appearance of the indicator function $I_{\Lambda_{CR}(R)}(\lambda)$ in (10). For $\gamma < 0$, $g^z(w)$, as a function of w, will be strictly positive implying $\phi^{*,+}(z) = \infty$ justifying the appearance of the delta function $\delta(\lambda|\Lambda_{CR}(R))$ in (11). Finally, for $\gamma = 0$, there is always a solution for $g^z(w) = 0$, for any $z \in \Re$, implying an unconstrained maximization problem as shown in (12).

In the next corollary, we use Theorem 1 to identify the implied MD SDFs and to verify if the different members of the Cressie Read family of discrepancies satisfy the regularity sufficient conditions that would allow us to obtain those MD SDFs from the first derivative of the convex conjugate $\phi^{*,+}(.)$.

Corollary 2. *Assume that the discrepancy in the minimization problem (3) belongs to the Cressie Read family,* $\phi^\gamma(m) = \frac{m^{\gamma+1} - a^{\gamma+1}}{\gamma(\gamma+1)}$ *with* $\gamma \in \Re$*. For any* $\gamma \geq -1$*, at least one of the regularity sufficient conditions stated in Theorem 1 is satisfied by* ϕ^γ *and the corresponding MD implied*

[17] That is, $I_A(x) = 1,$ if $x \in A$, and 0 otherwise.

SDF will be given by:

$$m^*_{MD}(\gamma, R) = \left(a^\gamma + \gamma\lambda^{*\prime}_\gamma\left(R - \frac{1}{a}1_K\right)\right)^{\frac{1}{\gamma}} I_{\Lambda_{CR}(R)}(\lambda^*_\gamma), \gamma > 0 \tag{13}$$

$$m^*_{MD}(\gamma, R) = \left(a^\gamma + \gamma\lambda^{*\prime}_\gamma\left(R - \frac{1}{a}1_K\right)\right)^{\frac{1}{\gamma}}, -1 \leq \gamma < 0 \tag{14}$$

$$m^*_{MD}(0, R) = ae^{\lambda^{*\prime}_0\left(R - \frac{1}{a}1_K\right)}, \gamma = 0. \tag{15}$$

where for $\gamma > 0$, λ^*_γ solves (10), for $-1 \leq \gamma < 0$, λ^*_γ solves (11), and λ^*_0 solves (12). For any $\gamma \geq -1$, the λ^*_γs are such that $E\left[m^*_{MD}(\gamma, R)\right] = a$.

For $\gamma < -1$, both stated regularity conditions in Theorem 1 are not satisfied. In such a case, an alternative sufficient condition for the MD implied SDF to be given by the expression in (14), with λ^*_γ solving (11)[18]*, is that the expression raised to the power $\left(\frac{1}{\gamma}\right)$ remains strictly positive. This is guaranteed to be satisfied for any sample space with a finite number of states.*

Proof: See Appendix.

2.1 Interpretation as an Optimal Portfolio Problem

Problems (10), (11) and (12) have an interesting economic interpretation as optimal portfolio problems. The solution for the MD bound for each Cressie Read estimator will correspond to an optimal portfolio problem based on the following HARA-type utility function

$$u^\gamma(W) = -\frac{1}{\gamma+1}\left(a^\gamma - \gamma W\right)^{\left(\frac{\gamma+1}{\gamma}\right)}, \tag{16}$$

with $a > 0$ and W such that $a^\gamma - \gamma W > 0, \gamma < 0$ $(a^\gamma - \gamma W \geq 0, \gamma > 0)$, which guarantees that function u is well defined for an arbitrary γ, is concave, and strictly increasing.[19]

Based on a standard two-period model of optimal portfolio choices, we provide an interpretation to all Cressie-Read MD problems as optimal portfolio problems on the dual space.

Suppose an investor distributes his/her initial wealth W_0 putting λ_j units of wealth on the risky asset R_j and the remaining $W_0 - \sum_{j=1}^K \lambda_j$ in a risk-free asset paying $r_f = \frac{1}{a}$. Terminal wealth

[18] and such that: $E\left[m^*_{MD}(\gamma, R)\right] = a$.

[19] Specific values of γ will specialize the optimal portfolio problems to widely adopted utility functions. A value of -1 will correspond to a logarithmic utility function, 0 to the exponential, and 1 to quadratic utility. To obtain the logarithmic and exponential limiting cases we adopt the translated utility $\frac{a^{\gamma+1}}{\gamma+1} - \frac{1}{\gamma+1}\left(a^\gamma - \gamma W\right)^{\left(\frac{\gamma+1}{\gamma}\right)}$ exactly as it appears in Corollary 1, and make use of L'hopital's rule. The corresponding SDFs are easily obtained from the expressions in Corollary 2. In particular, for $\gamma = 1$, exactly as in the HJ case with nonnegativity constraint, the optimal SDF will be a nonnegative linear function of excess returns.

is then $W(\lambda) = W_0 * r_f + \sum_{j=1}^{K} \lambda_j * (R_j - r_f)$. Assume in addition that this investor maximizes the HARA utility function in (16), solving one of the following optimal portfolio problems:

$$\Omega = \underset{\lambda \in \Re^K}{sup}\, E\left[u^{\gamma}(W(\lambda)) I_{\Lambda}(\lambda)\right], \gamma > 0 \tag{17}$$

$$\Omega = \underset{\lambda \in \Re^K}{sup}\, E\left[u^{\gamma}(W(\lambda)) + \delta(\lambda|\Lambda)\right], \gamma < 0 \tag{18}$$

$$\Omega = \underset{\lambda \in \Re^K}{sup}\, E\left[u^{0}(W(\lambda))\right], \gamma = 0 \tag{19}$$

where $\Lambda = \{\lambda \in \Re^K : a^{\gamma} - \gamma W(\lambda) > 0\}$. Note that by scaling the original vector λ to be $\tilde{\lambda} = \frac{-\lambda}{\left(a^{\gamma} - \frac{\gamma W_0}{a}\right)}$, we can approximately decompose the utility function in $u(W) \approx u(W_0 * r_f) * \left(a^{\gamma} + \gamma \tilde{\lambda}\left(R - \frac{1}{a} 1_K\right)\right)^{\left(\frac{\gamma+1}{\gamma}\right)}$.[20] This decomposition essentially shows that solving the optimality problem in (10), (11), or (12) will measure the gain achieved when switching from a total allocation of wealth at the risk-free asset paying r_f to an optimal (in the utility u sense) diversified allocation that includes both risky assets and the risk-free asset.

2.2 The Sample Version of the MD Bounds

Let us consider the sample version of the population problem presented in (3):

$$\hat{m}_{MD} = \underset{\{m_1,\ldots m_T\}}{arg\ min}\ \frac{1}{T}\sum_{i=1}^{T} \phi(m_i),$$
$$\text{subject to } \frac{1}{T}\sum_{i=1}^{T} m_i\left(R_i - \frac{1}{a}1_K\right) = 0_K,\ \frac{1}{T}\sum_{i=1}^{T} m_i = a, m_i \geq 0 \text{ or } (m_i >> 0)\forall i. \tag{20}$$

This minimization is based on the space of nonnegative (or strictly positive) discrete SDFs with dimension T (sample dimension). As mentioned before, if there is no in-sample arbitrage (see Gospodinov, Kan and Robotti, 2014), there is at least one strictly positive admissible SDF for the observed sample, and Theorem 1 guarantees that the solution of (20) can be obtained by solving a dual portfolio problem in a space with dimension K (the number of primitive assets). In what follows, we provide the sample version of Theorem 1 that formalizes this argument:

[20] This decomposition is exact when $a = 1$.

Theorem 2. *Consider the primal problem (20), and the dual problem:*

$$\underset{\alpha\in\Re,\lambda\in\Re^K}{sup} \; a*\alpha - \sum_{i=1}^{T}\frac{1}{T}\left[\phi^{*,+}\left(\alpha+\lambda'\left(R_i-\frac{1}{a}1_K\right)\right)\right]. \tag{21}$$

where $\Lambda = \{\tilde{\alpha}\in\Re, \tilde{\lambda}\in\Re^K : \phi^{*,+}\left(\tilde{\alpha}+\tilde{\lambda}'\left(R_i-\frac{1}{a}1_K\right)\right) < \infty, \forall i=1,...T\}$*, and* $\phi^{*,+}(z)$ *is the same as in Theorem 1.*

If there is no-arbitrage in the observed sample, the values of the primal and the dual problem coincide (with dual attainment). A sufficient condition allowing the Minimum Discrepancy SDF to be obtained from the solution of the dual optimization problem is that either $d = lim_{x\to\infty}\frac{\phi(x)}{x} = \infty$ *or* $(d<\infty$ *and* $c = lim_{x\to\infty}((d-\phi'(x))x) > 0)$*. In such cases, the implied SDF is obtained by:*

$$m^*_{MD} = \frac{\partial\phi^{*,+}(z)}{\partial z}_{z=\left(\hat{\alpha}^*+\hat{\lambda}^{*\prime}\left(R-\frac{1}{a}1_K\right)\right)} \tag{22}$$

with $[\hat{\alpha}^* \; \hat{\lambda}^*]$ *the optimizing values of (21).*

Notice that the $\delta(.)$ function has been eliminated in the primal problem (21) compared with Theorem 1. When dealing with the sample problem, we can simplify the region where we search for the Lagrange Multipliers α and λ to depend on all observed returns at once (as in Λ above) instead of being a random region that depends on each realization of the vector of returns $R(\omega)$ as in the definition of $\Lambda(R)$ in Theorem 1. This fact tremendously simplifies the implementation of the MD bounds. Otherwise the proof follows directly from the proof of Theorem 1.

2.2.1 Finding the Admissible Minimum-Discrepancy (MD) SDF

In what follows, the sample version of Corollary 1 provides an algorithm to obtain in practice the MD SDF $\hat{m}_{MD}$ when the discrepancy belongs to the Cressie Read family.

Corollary 3. *Assume that the discrepancy function belongs to* $\phi(m) = \frac{m^{\gamma+1}-a^{\gamma+1}}{\gamma(\gamma+1)}$ *with* $\gamma\in\Re$*, and that there is no in-sample arbitrage. In this case, solving (20) is equivalent to solving (**??**). And letting* $\Lambda_{CR} = \{\lambda\in\Re^K,$ *s.t. for* $i=1,...,T : \left(a^\gamma+\gamma\lambda'\left(R_i-\frac{1}{a}1_K\right)\right) > 0\}$*, the Lagrange multipliers that solve (**??**) specialize to:*

i) if $\gamma > 0$*:*

$$\hat{\alpha_\gamma} = \frac{a^\gamma}{\gamma},\; \hat{\lambda_\gamma} = \underset{\lambda\in\Re^K}{arg\ sup}\frac{1}{T}\sum_{i=1}^{T}\left(\frac{a^{\gamma+1}}{\gamma+1}-\frac{1}{\gamma+1}\left(a^\gamma+\gamma\lambda'\left(R_i-\frac{1}{a}1_K\right)\right)^{\left(\frac{\gamma+1}{\gamma}\right)} I_{\Lambda_{CR}}(\lambda)\right), \tag{23}$$

ii) if $\gamma < 0$:

$$\hat{\alpha_\gamma} = \frac{a^\gamma}{\gamma},\ \hat{\lambda_\gamma} = arg \underset{\lambda \in \Lambda_{CR}}{sup} \frac{1}{T} \sum_{i=1}^{T} \left(\frac{a^{\gamma+1}}{\gamma+1} - \frac{1}{\gamma+1} \left(a^\gamma + \gamma \lambda' \left(R_i - \frac{1}{a} 1_K \right) \right)^{\left(\frac{\gamma+1}{\gamma}\right)} \right), \tag{24}$$

iii) if $\gamma = 0$, the maximization is unconstrained:

$$\hat{\alpha_0} = 1 + ln(a),\ \hat{\lambda}_0 = arg \underset{\lambda \in \Re^K}{sup}\ a - a\frac{1}{T} \sum_{i=1}^{T} e^{\lambda' \left(R_i - \frac{1}{a} 1_K \right)}. \tag{25}$$

According to Corollary 2, the MD SDF $\hat{m}^\gamma_{MD}$ can be recovered via the first derivative of the convex conjugate $\phi^{*,+}$:[21]

$$\hat{m}^i_{MD}(\gamma, R) = \left(a^\gamma + \gamma \hat{\lambda}'_\gamma \left(R_i - \frac{1}{a} 1_K \right) \right)^{\frac{1}{\gamma}} I_{\Lambda_{CR}}(\hat{\lambda}_\gamma), i = 1, ..., T; \gamma > 0. \tag{26}$$

$$\hat{m}^i_{MD}(\gamma, R) = \left(a^\gamma + \gamma \hat{\lambda}'_\gamma \left(R_i - \frac{1}{a} 1_K \right) \right)^{\frac{1}{\gamma}}, i = 1, ..., T; -1 \leq \gamma < 0. \tag{27}$$

$$\hat{m}^i_{MD}(0, R) = a e^{\hat{\lambda}_0' \left(R_i - \frac{1}{a} 1_K \right)}, i = 1, ..., T; \gamma = 0. \tag{28}$$

Based on the regularity conditions assumed for (m_t, R_t), Almeida and Garcia (2012) obtain consistency results for the SDF $\hat{m}_{MD}$, the Minimum discrepancy sample functions $\phi(\hat{m}_{MD})$, and the Lagrange Multipliers $\hat{\lambda_\gamma}$, for all members of the Cressie Read family. In other words, they show that $\hat{m}_{MD}$, $\phi(\hat{m}_{MD})$, and $\hat{\lambda_\gamma}$ converge in probability to their population counterparts, m^*_{MD}, $\phi(m^*_{MD})$, and λ^*_γ, as T becomes large.

2.2.2 Minimum Discrepancy SDF Frontier

To complete our characterization of MD SDFs, we provide an operational algorithm to obtain such variables when there is no risk-free asset in the space of returns. Similarly to HJ, the idea is to propose a grid of possible meaningful values for the SDF mean, say fixing a set $A = \{a_1, a_2, ..., a_J\}$, and to solve the optimization problem in (23), (24) or (25) for each $a_l \in A$, obtaining the corresponding optimal weight vector $\hat{\lambda}_\gamma(a_l)$ for each SDF mean. The SDF frontier is given by the following expression:

[21] Or equivalently, from the first-order conditions of (23), (24) or (25) with respect to λ, evaluated at $\hat{\lambda}_\gamma$: $\frac{1}{T} \sum_{i=1}^{T} \left(a^\gamma + \gamma \hat{\lambda}'_\gamma \left(R_i - \frac{1}{a} 1_K \right) \right)^{\frac{1}{\gamma}} \left(R_i - \frac{1}{a} 1_K \right) = 0_K$.

$$I_P(a_l,\gamma) = \frac{a_l^{\gamma+1}}{1+\gamma} + \frac{1}{T}\sum_{i=1}^{T} -\frac{1}{\gamma+1}\left(a_l^{\gamma} + \gamma\hat{\lambda}_{\gamma}(a_l)'\left(R_i - \frac{1_K}{a_l}\right)\right)^{\left(\frac{\gamma+1}{\gamma}\right)}, l = 1,2,...,J. \tag{29}$$

Alternatively, we can go back to the basic definition of the bound as a minimum discrepancy problem, and write the solution by first obtaining the implied MD SDFs appearing in (26), (31) or (32), $\hat{m}^i_{MD,a_l}$, and substituting it in the sample divergence function ϕ, obtaining the MD SDF frontier:

$$I_{MD}(a_l,\gamma) = \frac{1}{T}\sum_{i=1}^{T}\frac{(\hat{m}^i_{MD,a_l})^{\gamma+1} - a_l^{\gamma+1}}{\gamma(\gamma+1)}, l = 1,2,...,J. \tag{30}$$

2.2.3 HJ with Positivity Constraint as a Particular Case

When we choose $\gamma = 1$ on the Cressie Read family, the discrepancy function becomes $\phi(m) = \frac{m^2-a^2}{2}$, and we are solving the following MD SDF bound:

$$\begin{aligned} \hat{m}_{MD(\gamma=1)} &= \underset{\{m_1,...m_T\}}{arg\ min}\, \tfrac{1}{T}\textstyle\sum_{i=1}^{T} \tfrac{m_i^2-a^2}{2}, \\ \text{subject to} &\tfrac{1}{T}\textstyle\sum_{i=1}^{T} m_i\left(R_i - \tfrac{1}{a}1_K\right) = 0,\ \tfrac{1}{T}\textstyle\sum_{i=1}^{T} m_i = a, m_i \geq 0\ \forall i. \end{aligned} \tag{31}$$

for SDFs with a fixed mean value equal to a.

This equation represents, apart from a normalization factor of $\frac{1}{2}$, the HJ (1991) variance bound with nonnegativity constraint.[22]. By looking at the sample version of Corollary 1, obtained when we substitute $\gamma = 1$ in Eq. (10), we note that the dual optimization problem is a quadratic problem truncated at zero by $I_{\Lambda_{CR}}$. This is equivalent to HJ (1991), with non-negativity constraint (see online appendix).

2.2.4 Snow (1991) Moment Specific Approach as a Particular Case

Snow (1991) solved moment specific problems of the type $\underset{m>0}{inf} E[m^{\delta}]^{\frac{1}{\delta}}$, for $\delta > 1$, where m is an admissible SDF. They correspond, apart from an affine transformation, to Cressie Read discrepancies where $\gamma > 0$. It excludes important cases like ET ($\gamma = 0$) and EL ($\gamma = -1$) whose discrepancies are respectively $\phi^{ET}(m) = mln(m) - aln(a)$ and $\phi^{EL}(m) = ln(a) - ln(m)$.[23] It

[22]Adopting geometric arguments based on inner product properties of a Hilbert Space, HJ (1991) showed that the variance bound obtained with nonnegative admissible SDFs is the tightest possible, meaning that restricting the minimization to strictly positive admissible SDFs doesn't improve the bound.

[23]To obtain the EL discrepancy as a limit of the Cressie Read family, letting $g^{\gamma}(m) = \frac{m^{\gamma+1}-a^{\gamma+1}}{\gamma+1}$, we apply L'hopital's rule on $lim_{\gamma\to-1} g^{\gamma}(m)$ noticing that $m^{\gamma+1} = e^{(\gamma+1)ln(m)}$ and $a^{\gamma+1} = e^{(\gamma+1)ln(a)}$. Since $\phi^{\gamma}(m) = \frac{g^{\gamma}(m)}{\gamma}$, this directly implies $lim_{\gamma\to-1}\phi^{\gamma}(m) = ln(a) - ln(m)$. The ET discrepancy is a bit trickier to obtain since the

also excludes discrepancies with negative powers of SDFs. As we will see in Section 2.3), Taylor expansions reveal the distinctive roles of odd and even moments for $\gamma \leq 0$ on pricing kernel variability. Therefore our approach provides a more complete and robust treatment to diagnosis of models and analysis of trading strategies.

2.3 Taylor Expansion of the Cressie Read Discrepancy and Higher Moment Weights

Backus, Chernov and Martin (2011) highlight the importance of analyzing the entropy of a pricing kernel (SDF) through the cumulant-generating function. Cumulants help to identify how much of the pricing kernel entropy comes from its variance, skewness, kurtosis and other higher moments. Similarly, for each member of the Cressie Read family, given an asset pricing model and its corresponding pricing kernel, we can measure the contribution of each moment of the pricing kernel to the overall model discrepancy. Each member will give different weights to the moments of a pricing kernel, potentially providing a way to better distinguish between different asset pricing models.

Given the Cressie Read discrepancy family, to be able to see how much weight is given to each SDF moment, let us fix the Cressie Read parameter γ and the SDF mean at a obtaining the following function: .

We are interested in Taylor expanding the expected value of $\phi(m) = \frac{m^{\gamma+1} - a^{\gamma+1}}{(\gamma(\gamma+1))}$ around the SDF mean a. Noting that $\phi(a) = 0$, $\phi'(m) = \frac{m^\gamma}{\gamma}$, $\phi''(m) = m^{\gamma-1}$, $\phi'''(m) = (\gamma-1)m^{\gamma-2}$, $\phi''''(m) = (\gamma-1)(\gamma-2)m^{\gamma-3}$...,Taylor expanding ϕ and taking expectations on both sides we obtain:[24]

$$E(\phi(m)) = \frac{a^{\gamma-1}}{2}E(m-a)^2 + \frac{(\gamma-1)a^{\gamma-2}}{3!}E(m-a)^3 + \frac{(\gamma-1)(\gamma-2)a^{\gamma-3}}{4!}E(m-a)^4 + ... \quad (32)$$

From this Taylor expansion we see that the weights given to skewness and kurtosis are respectively $\frac{(\gamma-1)a^{\gamma-2}}{3!}$ and $\frac{(\gamma-1)(\gamma-2)a^{\gamma-3}}{4!}$.[25] Now, considering values of the SDF mean a that are close

limit should be taken on the expectation of $\phi^\gamma(m)$. Here we should use the Dominated Convergence Theorem that guarantees that $lim_{\gamma \to 0} E(m^{\gamma+1}) = E(m)$, since $E(m^\delta) < E(m^{\gamma+1}) < \infty$ for any $\delta < \gamma + 1$. Equipped with this result, we replicate the EL proof by using L'hopital's rule on $lim_{\gamma \to 0} \frac{E(m^{\gamma+1} - a^{\gamma+1})}{\gamma}$. This implies $lim_{\gamma \to 0} E(\phi^\gamma(m)) = E(mln(m) - aln(a))$.

[24] Note that all functions in the Cressie Read family are analytic, that is, their derivative of any order exists. For this reason, the only condition that is needed for the Taylor expansion to be valid is the existence of the first four moments of the MD SDF. Note, however, that we only make use of the Taylor expansion to better clarify our results although it is not really necessary to validate our bounds, whose existence only depend on the existence of the moment $E[\phi(m)]$ for at least one admissible SDF.

[25] Note here that skewness is represented by the third central SDF moment and kurtosis by the fourth central moment.

to one, we have:

$$skew_{wei}(\gamma) = \frac{(\gamma - 1)}{6} \tag{33}$$

$$kurt_{wei}(\gamma) = \frac{(\gamma - 2)}{4} * skew_{wei}(\gamma) \tag{34}$$

There are two important effects to understand regarding the weights given to skewness and kurtosis in the discrepancy function. First, for values of γ close to one, both skewness and kurtosis have small weights when compared to the variance that has a weight equal to one half in the expansion. This implies that discrepancies with values of γ close to one do not capture much of the higher moment activity of pricing kernels. Once we move to more negative values of γ both skewness and kurtosis receive considerable weights in the expansion. The second important aspect to be observed refers to the relative weights that are given to skewness and kurtosis by different Cressie Read functions. In this sense, for $-2 < \gamma < 1$ note that the absolute weight given to kurtosis is smaller than the corresponding weight given to skewness. Nevertheless, once we look at values of $\gamma < -2$, kurtosis receives more weight than skewness. In fact all even higher-moments receive more absolute weight than their corresponding odd higher-moments in this region of γ.[26] We will come back to these observations when analyzing the disaster and long-run risks models.

Figure 1 is based on a Taylor expansion of the HARA utility functions that appear on the dual solutions of the MD bounds (see Appendix B in the online appendix). It shows the weights given to expected returns by the different discrepancies within the family in the dual portfolio problems. Observe that the signs are switched with respect to weights given in the primal problems. On the dual portfolio problems where $\gamma < 1$, positive weights are given to skewness while negative weights are given to kurtosis. Note however that similarly to the primal problems, Cressie-Read discrepancies with more negative values of γ put higher weights at both skewness and kurtosis, and in general more relative weight to kurtosis when γ is very negative.

3 Diagnosing Asset Pricing Models with the Minimum-Discrepancy Approach

In this section, we use the information bounds to verify admissibility of asset pricing models that are currently popular in the consumption-based asset pricing literature and that are particularly fitted to illustrate the importance of integrating higher moments in the construction of model diagnosing bounds. We chose a disaster risk model, two long-run risk models with different

[26] More precisely, the $k_{th}-$moment receives higher absolute weight than the $(k-1)_{th}-$moment, when k is even.

preferences, and a generalized disappointment aversion model. All four models create nonlinearities in the stochastic discount factor through expected consumption, consumption volatility or preferences and should therefore be evaluated with appropriate bounds that incorporate higher moments of returns of basis assets since all of them will pass easily the minimum variance bound of Hansen and Jagannathan (1991).[27]

3.1 The Disaster Model

In the model of Barro (2006), a disaster-like drop in aggregate consumption growth produces a large equity premium and captures other non-normal features of asset returns. Calibration of the left tail of the probability distribution of consumption growth is based on international evidence of such large drops in consumption growth. The distribution of consumption growth combines a gaussian component with a jump component and translates into non-normal asset returns. Such a model offers an ideal testing ground for our extended bounds since they precisely capture non-Gaussian returns in basis assets. In this model the logarithm of consumption growth is given by:

$$g_{t+1} = \eta_{t+1} + J_{t+1} \tag{35}$$

where η_{t+1} is the normal component $\aleph(\mu, \sigma^2)$ and J_{t+1} is a Poisson mixture of normals. The number-of-jumps variable j takes integer values with probabilities $e^{-\tau}\frac{\tau^j}{j!}$, where τ is the jump intensity. Conditionally on the number of jumps, J_t is normal:

$$J_t | j \sim \aleph(j\alpha, j\lambda^2). \tag{36}$$

In this model, the logarithm of the stochastic discount factor with power utility is:

$$\log m_{t+1} = \log \beta - \zeta g_{t+1} \tag{37}$$

where ζ is the coefficient of relative risk aversion. Therefore, the mean of the SDF is:

$$a = \exp\left\{\log \beta - \zeta\mu + \frac{1}{2}(\zeta\sigma^2) + \tau(e^{-\zeta\alpha + 0.5(\zeta\lambda)^2} - 1)\right\}. \tag{38}$$

The discrepancy bound for the Cressie-Read family is the expectation of $\phi^{\gamma}_{CR}(m)$ defined in

[27] Here, we refer to the variance bound with the T-bill and a market return.

(9). It can easily be obtained by taking the expectation of $\exp(\gamma+1)\log m$. These bounds have a direct link with the measure of entropy used in Backus, Chernov and Martin (2011). When $\gamma=-1$ our discrepancy function is $log(a)-E(log(m))$ as derived in section 2, which corresponds precisely to the entropy of the pricing kernel reported in their Equation (13). In the dual space, it corresponds to the maximum excess log return of the growth-optimal portfolio of Bansal and Lehman (1997). Another entropy measure (Shannon entropy) has been put forward by Stutzer (1995). It is obtained by minimizing the KLIC and corresponds to our discrepancy function $E(mlog(m))-alog(a)$ when $\gamma=0$. In fact, we can similarly define a whole family of entropy measures indexed by γ for our Cressie-Read discrepancy function.

To diagnose the disaster model, we compute the Cressie-Read bounds with the returns on the S&P 500 index and equity options strategies on this index.[28] Since the left tail of the option return distribution should be directly affected by large drops in consumption, a disaster model should price these derivatives portfolios. We construct frontiers based on our Cressie-Read discrepancy function for different values of γ. We calibrate the disaster model with $\tau=0.01$, $\alpha=-0.3$ and $\lambda=0.15$. It means that there is a 1% probability of a 30% drop (on average) in consumption growth relative to its mean. The overall mean of consumption growth is set at 0.02 and its variance at 0.035^2. Given that the theoretical mean for the Poisson is $\mu+\tau\alpha$ and the variance $\tau(\alpha^2+\lambda^2)$, we set $\mu=0.023$ and $\sigma=0.01$.

We report in the upper panel of Figure 2 a set of graphs where we diagnose the Poisson disaster model with the entropic bounds obtained for γ equal to 1, 0 and -1.[29] We set all the consumption parameters at the values indicated above and the risk aversion parameter to 6.8.[30] We vary the magnitude of the disaster, which is a key parameter in the model, from -0.30 to -0.10. The disaster model is admissible for the quadratic bound ($\gamma=1$) as all the model mean-entropy pairs are within the frontier. This is not the case however for the entropy bounds corresponding to $\gamma=0$ and $\gamma=-1$. For all disaster values, the mean-entropy points lie below the frontier.[31]

A key explanation for these results lies in the high values of skewness and kurtosis of the

[28]We use four options portfolios that consist of highly liquid at-the-money (ATM) and out-of-the-money (OTM) European call and put options on the S&P 500 composite index trading on the Chicago Mercantile Exchange. These have been constructed by Agarwal and Naik (2004) to study performance of hedge funds.

[29]The case of $\gamma=1$ corresponds to the minimum variance frontier of Hansen and Jagannathan (1991) with positivity constraints, multiplied by a one-half factor.

[30]Backus, Chernov and Martin (2011) use 5.19 for the risk aversion parameter when they evaluate the model with equities only. We increase slightly this value to serve also when we include the options portfolios in the basis assets.

[31]To compare with the findings of Backus, Chernov and Martin (2011) we have also computed frontiers with only the market returns as a basis asset. All the model points are inside the bounds for all values of γ. The model passes the bound even for lower disaster magnitudes then the one considered in Barro (2006).

option portfolio returns. Since the discrepancies for γ equal to 0 and -1 weight relatively more the higher moments of basis asset returns, the discrepancy bound is heightened enough with respect to the one for $\gamma = 1$ to make the model non-admissible irrespective of the size of the disaster. It is also informative to analyze the results from a model perspective. By looking at the Taylor expansions of Section 2.3, it is easy to rationalize why it becomes more difficult for the model to pass the bounds when we get to more negative values of γ in the Cressie Read family. The more negative the γ, the higher (in absolute value) are the negative weights given to skewness of the model implied pricing kernel. Having negative values of mean size of disaster jump risk (α) generates more positive skewness on the pricing kernel making it harder for the model, for a fixed value of α, to pass the frontiers for more negative values of γ.

In the lower panel of Figure 2, we now keep the size of the disaster constant at the original value -0.30 set by Barro (2006) and vary instead the risk aversion parameter of the representative investor in the spirit of Hansen and Jagannathan (1991) for the canonical CCAPM. According to the quadratic bound the model ($\gamma = 1$) is admissible for any value of the risk aversion parameter above 5. As we lower the value of γ, making the bound more restrictive, we naturally increase the value of the risk aversion parameter at which the model becomes admissible. Looking at (37) defining the disaster pricing kernel and again at the Taylor expansions of Section 2.3, we can see that increasing the risk-aversion coefficients strongly contributes, through the negative jump component, to increasing the skewness of the pricing kernel. Since skewness receives negative weights at the Cressie Read discrepancy functions analyzed here ($\gamma = -1, 0$), it becomes harder for the model, for a fixed value of the risk-aversion coefficient ζ, to pass the bounds for more negative values of γ.

This analysis of the disaster Poisson model has shown that a large drop in consumption will make a simple consumption-based asset pricing model easily admissible with respect to the usual minimum variance bound of Hansen and Jagannathan (1991), but that tighter bounds capturing higher moments of the basis asset returns with some non-normalities in returns impose more stringent conditions on the admissibility of the model. For an alternative analysis of the disaster model see Liu (2012), who makes use of generalized entropic bounds (derived by Holder's inequality) to estimate disaster's distribution in Barro's (2006) model, based on index option returns.

3.2 The Long-Run Risks Models

The previous model was very close to the benchmark CCAPM model. It had the same power utility preferences and differed only by the addition of a Poisson variable to the consumption growth process. In this section we will depart from the benchmark model by adding a small long-run predictable component in consumption growth and a fluctuating consumption volatility to capture economic uncertainty, as followed:

$$\begin{aligned} g_{t+1} &= \mu + x_t + \sigma_t \eta_{t+1} \\ x_{t+1} &= \rho x_t + \varphi_e \sigma_t e_{t+1} \\ \sigma^2_{t+1} &= \sigma^2 + \nu_1(\sigma^2_t - \sigma^2) + \sigma_w w_{t+1} \end{aligned} \tag{39}$$

where g_t is the logarithm of real consumption growth. All innovations are $\aleph, i.i.d.(0,1)$.

The logarithm of the intertemporal marginal rate of substitution (IMRS) is:

$$m_{t+1} = \theta \log \delta - \frac{\theta}{\psi} g_{t+1} + (\theta - 1) r_{a,t+1} \tag{40}$$

where $r_{a,t+1}$ is the return on the wealth portfolio.

3.2.1 The Bansal-Yaron Model

To derive solutions, Bansal and Yaron (2004) use the standard approximations of the return formula from Campbell and Shiller (1988):

$$r_{a,t+1} = \kappa_0 + \kappa_1 z_{t+1} - z_t + g_{t+1} \tag{41}$$

where z is the logarithm of the wealth-consumption ratio and κ_0 and κ_1 are approximating constants that depend only on the average level of z. The relevant state variables in solving for the equilibrium wealth-consumption ratio are x_t and σ^2_t. The approximate solution for z_t is conjectured to be:

$$z_t = A_0 + A_1 x_t + A_2 \sigma^2_t \tag{42}$$

where the expressions for A_0, A_1 and A_2 are given in section C of the online appendix.

The discrepancy function is given by:

$$E\left[\frac{M^s - [E(M)]^s}{s(s-1)}\right] \tag{43}$$

To find the expression for the discrepancy, it suffices to compute $E[M^s]$, that is:

$$\begin{aligned} E[M^s] &= E[\exp\{s(\log M_{t+1})\}] \\ &= E\left[\exp\left\{s[\theta\log\delta - \frac{\theta}{\psi}g_{t+1} + (\theta-1)r_{a,t+1}]\right\}\right] \\ &= exp\left\{s\left[\theta\log\delta - \frac{\theta}{\psi}E(g_{t+1}) + (\theta-1)E(r_{a,t+1})\right] + \frac{1}{2}s^2Var\left[\theta\log\delta - \frac{\theta}{\psi}g_{t+1} + (\theta-1)r_{a,t+1}\right]\right\} \end{aligned} \tag{44}$$

The expression in terms of fundamentals and preference parameters are given in section C of the online appendix. We calibrate the long-run risks model with the values used in Bansal and Yaron (2004).[32]

In order to diagnose the long-run risks model, we assess whether the model is within the mean-discrepancy frontier estimated with a set of value and size portfolios[33], therefore whether the model is compatible with the cross-section of equity returns.[34] In Figure 3 we can see how the model fares with respect to the discrepancy bound estimated with the six portfolios as well as with only a market portfolio. In the upper panel we compute the mean-discrepancy pairs of the model for various values of the persistence parameters in expected consumption ρ, from 0.983 to 0.995. The right-hand side graph corresponds to a value of 1 for the parameter γ indexing the discrepancy function. While the model passes the market frontier for all values of ρ, it stays outside of the six-portfolio frontiers for all these values. When we lower the discrepancy parameter to $\gamma = -3$, the model becomes compatible with the discrepancy bound for lower values of the persistence parameter ρ.

Since the model implies a log-normal SDF, changing ρ will affect all the moments of the log-normal distribution. Table 1 presents how variance, skewness and kurtosis of the SDF vary as a function of ρ. As expected, increasing ρ increases all these quantities that are linked in the

[32]For the fundamentals, $\mu = 0.0015$, $\sigma = 0.0078$, $\varphi_e = 0.044$, $\nu_1 = 0.987$, $\rho = 0.979$, and $\sigma_w = 0.23 \times 10^{-5}$. For the preference parameters, $\gamma = 10$, $\psi = 1.5$ and $\delta = 0.9989$. The values of κ_1 and κ_0 are determined endogenously through (42).

[33]They correspond to the six benchmark portfolios available on Kenneth French Data Library, Small Value, Small Neutral, Small Growth, and Big Value, Big Neutral, and Big Growth.

[34]The idea is that in equilibrium the differences in the risk premium across assets reflect the differences in their long-run risks betas (cash-flow betas). See in particular Bansal, Dittmar and Lunblad (2005).

log-normal distribution but the net effect is that kurtosis increases more relative to skewness. This can be seen by comparing the first and the last lines of this table. In the first line, when $\rho = 0.983$, skewness and kurtosis are very close to each other (skewness equals 0.0047 and kurtosis 0.0056), while in the last line, when $\rho = 0.995$, kurtosis is more than double the skewness value (0.2225 compared to 0.1088). According to our previous analysis of the Taylor expansion of the Cressie Read functions, higher values of kurtosis when compared to skewness will induce higher values of discrepancy when γ is large and negative. This is exactly what can be observed in Table 2. When we look at any line of this table from the right to the left (from $\gamma = 1$ to $\gamma = -3$), we see that the discrepancy function starts at half the variance of the pricing kernel and then decreases because skewness gets more weight than kurtosis for intermediate values of γ. It finally increases again for $\gamma = -3$ where kurtosis receives more weight than skewness.

In the lower panel of Figure 3 we show the Cressie Read bounds for γ equal to -1, 0, and 1 and when the persistence of consumption growth ρ varies in $\{0.979, 0.987, 0.995\}$ and the persistence of volatility ν_1 varies in $\{0.987, 0.993, 0.999\}$. The highest model discrepancy values are obtained when $\rho = 0.995$ and $\nu_1 \in \{0.987, 0.993, 0.999\}$. Increasing any of the two persistence parameters increases discrepancies and the effect is slightly stronger when we increase ρ.[35] Only for one pair of persistence parameters ($\rho = 0.995, \nu_1 = 0.999$) the model passes in both Cressie Read quadratic bounds ($\gamma = 1$) with the six FF portfolios. Note that even though this pair represents extremely high persistence values for consumption growth and volatility, the good model performance under the bound with $\gamma = -1$ suggests that there is room for introducing more skewness in the pricing kernel. For instance, introducing more skewness by introducing jumps in the state vector might help the model to pass more easily in the quadratic bounds, while jumps are a realistic feature of the markets. Therefore, an alternative version of the model that could potentially pass all discrepancy bounds with less tension on persistence parameters is the LRR model of Drechsler and Yaron (2011) with compound-Poisson jumps in the state vector of the economy.

3.2.2 The Demand Side Model

Albuquerque, Eichenbaum and Rebelo (2012) proposed an extension to the previous long-run risks model by adding preference shocks that are correlated with consumption and dividends. With this addition they need a much smaller risk aversion coefficient to match the historical risk

[35] Note that there is an upper bound for increasing the volatility persistence parameter if we want to keep the auto-regressive process stationary.

premium.

The intertemporal marginal rate of substitution is given by:

$$m_{t+1} = \theta log(\delta) + \theta log(\lambda_{t+1}/\lambda_t) - \frac{\theta}{\psi} g_{t+1} + (\theta - 1) r_{c,t+1} \tag{45}$$

where:

$$\begin{aligned} g_{t+1} &= \mu + \alpha_c(\sigma^2_{t+1} - \sigma^2) + \pi_{c\lambda}\epsilon^{\lambda}_{t+1} + \sigma_c \epsilon^c_{t+1} \\ log(\lambda_{t+1}/\lambda_t) &= x_t + \sigma_\eta \eta_{t+1} \\ x_{t+1} &= \rho x_t + \sigma_\lambda \epsilon^{\lambda}_{t+1} \end{aligned} \tag{46}$$

The dynamics of the volatility is the same as in the Bansal and Yaron (2004) model and the same i.i.d. assumptions are kept for the innovations. The solution of the model, which follows Albuquerque, Eichenbaum and Rebelo (2012), is detailed in section D of the online appendix.

The formula for the discrepancies is then given by:

$$\begin{aligned} E[M^s] &= E(exp(s \times log(M_{t+1}))] \\ &= E\left[exp\left(s\left[\theta log(\delta) + \theta log(\lambda_{t+1}/\lambda_t) - \frac{\theta}{\psi} g_{t+1} + (\theta - 1) r_{c,t+1}\right]\right)\right] \\ &= exp\left(s\left[\theta log(\delta) - \theta E[log(\lambda_{t+1}/\lambda_t)] - \frac{\theta}{\psi} E[g_{t+1}] + (\theta - 1) E[r_{c,t+1}]\right]\right. \\ &\left. + \frac{s^2}{2} Var\left[\theta log(\delta) - \theta log(\lambda_{t+1}/\lambda_t) - \frac{\theta}{\psi} g_{t+1} + (\theta - 1) r_{c,t+1}\right]\right) \end{aligned} \tag{47}$$

where:

$$r_{c,t+1} = \kappa_{c0} + \kappa_{c1} z_{c,t+1} - z_{c,t} + g_{t+1} \tag{48}$$

The expression in terms of the model parameters is presented in section D of the online appendix. To diagnose the model, we keep the basis assets used for the basic LRR model and vary the persistence parameter ρ of the preference shock persistence. The original value in Albuquerque, Eichenbaum and Rebelo (2012) for the persistence parameter is equal to 0.99891.[36] Thus we will vary ρ from 0.99 to 0.9995. In Table 3, we report the implied variance, skewness and kurtosis of the SDF when we vary ρ. As expected, increasing ρ increases all moments but the effect is stronger for kurtosis than for skewness, which helps model performance for more negative values of γ. Indeed, the values for the discrepancies for different $\gamma's$ and ρ are reported in Table 4. The

[36] We also keep the values used by Albuquerque, Eichenbaum and Rebelo (2012) to calibrate their general extended model. See Table 1 in our online appendix for the exact values.

effect of ρ on the discrepancy is not very marked except for $\gamma = -3$.

The higher panel of Figure 4 confirms these results since the model performs very well when evaluated with the S &P 500 implied bound but not when we add the Fama-French portfolios to build the frontiers. Even when $\gamma = -3$ the model does not generate sufficient discrepancy to pass the bound except for the highest extreme persistence value. In the lower panel of Figure 4 we vary the risk aversion parameter used in the model in the set $\{2, 2.5, 3, 3.5, 4\}$.[37] As expected from the previous section on the Bansal-Yaron model, a higher risk aversion makes it much easier for the model to pass the discrepancy bounds. Therefore, with a reasonable combination of the preference shock persistence and of a slightly higher value of the risk aversion coefficient, the demand side model performs well.

3.2.3 Generalized Disappointment Aversion

Another prominent utility function used recently in the context of long-run risk features disappointment aversion. First proposed by Gul (1991), it has been generalized by Routledge and Zin (2010) and applied to asset pricing with long-run risk by Bonomo, Garcia, Meddahi, and Tedongap (2011). Compared with expected utility, generalized disappointment aversion (GDA) overweights outcomes below a threshold set at a fraction of the certainty equivalent of future utility.[38] Given the nonlinear nature of this model, with a kink in the utility function, it provides a good testing ground for assessing its performance with our discrepancy bounds. We focus on a simple version of this model.

The functional form of GDA preferences is given by:

$$u(\mu(p)) = \sum_{x_i \in X} p(x_i)u(x_i) - \theta \sum_{x_i \leq \delta\mu(p)} p(x_i)(u(\delta\mu(p)) - u(x_i)), \tag{49}$$

with the usual CRRA utility function $u(.)$:

$$\begin{aligned} u(x) &= \frac{x^\alpha}{\alpha} \text{For } \alpha \leq 1,\ \alpha \neq 0 \\ u(x) &= log(x) \text{ For } \alpha = 0. \end{aligned} \tag{50}$$

In the utility function (49), p represents a generic lottery with outcomes $x_i \in X$ and $\mu(p)$ the lottery certainty equivalent implicitly defined. The key parameter θ represents the strength of

[37]Here we keep the preference shock persistence parameter, ρ, fixed equal to 0.99891.

[38]Disappointment aversion in Gul (1991) sets the threshold at the certainty equivalent.

the disappointment aversion. We can restate our preferences in terms of the certainty equivalent, as follows:

$$\mu(p)^\alpha = \frac{E[x^\alpha(1+\theta I(x \leq \delta\mu))]}{1+\theta\delta^\alpha E[I(x \leq \delta\mu)]} \tag{51}$$

When set in the usual recursive utility as in the previous sections, we obtain the following SDF:

$$M_{t+1} = \beta(x_{t+1})^{\rho-1}\left(\frac{U_{t+1}}{\mu(U_{t+1})}\right)^{\alpha-\rho}\left(\frac{1+\theta I(U_{t+1} < \delta\mu(U_{t+1}))}{1+\theta\delta^\alpha I(U_{t+1} < \delta\mu(U_{t+1}))}\right), \tag{52}$$

where x_t now represents consumption growth. To keep the model simple, we follow Routledge and Zin (2010) and assume that consumption growth follows a Markov chain with two possible states $x^L = 0.932$ and $x^H = 1.054$ and a transition matrix given by:

$$\begin{bmatrix} \pi_{LL} & \pi_{LH} \\ \pi_{HL} & \pi_{HH} \end{bmatrix} = \begin{bmatrix} 0.43 & 0.57 \\ 0.57 & 0.43 \end{bmatrix} \tag{53}$$

For the other parameters, we follow Bonomo, Garcia, Meddahi, and Tedongap (2011).[39] As noted above, the certainty equivalent is defined implicitly. Therefore, we need to use a numerical method to solve for the certainty equivalent and U_{t+1} for all possible state combinations, a total of four of them.[40] The SDF is then obtained using the formula given in (52). Finally, we compute the SDF conditional expected value for each of the two possible states and used the Marvok chain invariant distribution to compute the discrepancies values.

We first report in Table 5 the variance, skewness and kurtosis of the SDF as a function of the disappointment parameter θ. It is clear that for higher values of θ (stronger disappointment) all moments increase. This is intuitive since the disappointing outcomes (tail events) are discounted more heavily. This translates in a straightforward manner to the discrepancies reported in Table 6. They all increase strongly with θ. Again, recalling the discrepancy Taylor expansion discussed in section 2.3, we observe that the higher values for the discrepancies occur for more negative values of the Cressie-Read parameter γ. For the value of θ of 2.8, the discrepancy ratio between $\gamma = -3$ and $\gamma = 1$ is greater than 2. The effect is much stronger than what we observed in the demand side model.

When the model is confronted to the Cressie-Read bounds with the market returns and the

[39] All parameter values are reported in Table 2 in the online appendix.

[40] Routledge and Zin (2010) proved in their paper that the implicit certainty equivalent in (49) is a contraction. See also Dolmas (2013) for a solution algorithm.

Fama-French portfolios in Figure 5, it is then not surprising to see that it performs well for all values of γ and θ. It is only for very low disappointment that the model fails to pass the bound, that is when the model gets close to the the traditional Epstein-Zin utility without disappointment. A major difference between the disaster model and the GDA model lies in the effect on the moments of the SDF. While the disaster model generates more kurtosis in the SDF distribution it also adds skewness to it, reducing the overall effect on the discrepancy. On the contrary the GDA model puts a higher weight on the left tail of the distribution without increasing skewness substantially since the SDF does not change for outcomes above the disappointment threshold.[41]

4 Conclusion

We provide a new family of Minimum Discrepancy bounds for stochastic discount factors that help to determine sources of dispersion in pricing kernels of asset pricing models. It works as a complementary tool to the entropic methods (Backus, Chernov and Martin (2011), and Backus, Chernov and Zin (2014)) recently adopted in the asset pricing literature. We show how to solve for our bounds and give a portfolio interpretation to them by looking at optimization problems in the dual space of SDFs, that is the space of portfolios of returns of primitive assets.

In order to put forward the usefulness of these bounds, we use the new SDF frontiers to bring a novel perspective on diagnosing popular asset pricing models such as the disaster model, long-run risks models, and disappointment aversion preferences. Our extension of Hansen and Jagannathan (1991) makes clear how the nonlinearities and non-normalities built in these models affect the higher moments of the SDF and how the new frontiers are more discriminating than the mean-variance frontiers to assess the performance of these models.

In this paper, we have voluntarily left aside the important issue of estimating the parameters of the asset pricing models under scrutiny and limited ourselves to a diagnosis as in the original paper of Hansen and Jagannathan (1991). In Almeida and Garcia (2012), we are assessing specification errors in stochastic discount factor models with our new metrics to generalize the quadratic-norm evaluation methodology developed in Hansen and Jagannathan (1997). Given the general formulation of the discrepancy problem presented in that paper, where the moment conditions are a function of a vector of model parameters, such a generalization opens the door for a thorough statistical comparison of the intertemporal asset pricing models we have reviewed in this paper.

[41] Dolmas (2013) associates the disaster model with the GDA model, not surprisingly reducing the disaster severity required to match the data.

References

[1] Agarwal V., and N.Y. Naik (2004). Risk and Portfolio Decisions Involving hedge Funds. *Review of Financial Studies*, **17**, 1, 63-98.

[2] Albuquerque R., M. Eichenbaum and S. Rebelo (2012). Valuation Risk and Asset Pricing. NBER working paper 18617.

[3] Almeida C. and R. Garcia (2012). Assessing Misspecified Asset Pricing Models with Empirical Likelihood Estimators. *Journal of Econometrics*, **170**, 2, 519-537.

[4] Arditti F.D. (1967). Risk and Required Return on Equity. *Journal of Finance*, **22**, 1, 19-36.

[5] Backus D., M. Chernov, and I. Martin (2011). Disaster Implied by Equity Index Options. *Journal of Finance*, **66**, 6, 1969-2012.

[6] Backus D., M. Chernov, and S. Zin (2014). Sources of Entropy in Representative Agent Models. *Journal of Finance*, **69**, 1, 51-99.

[7] Bansal R., R. Dittmar and C. Lunblad (2005). Consumption, Dividends, and the Cross-Section of Equity Returns. *Journal of Finance*, **60**, 1639-1672.

[8] Bansal R. and B. Lehman (1997). Growth-Optimal Portfolio Restrictions on Asset Pricing Models. *Macroeconomic Dynamics*, **1**, 333-354.

[9] Bansal R. and S. Viswanathan (1993). No Arbitrage and Arbitrage Pricing: A New Approach. *Journal of Finance*, **48**, 1231-1262.

[10] Bansal R. and A. Yaron (2004). Risks for the Long Run: A Potential Resolution of Asset Pricing Puzzles. *Journal of Finance*, **54**, 4, 1481-1509.

[11] Barro R. (2006). Rare Disaster and Asset Markets in the Twentieth Century. *Quarterly Journal of Economics*, **121**, 823-866.

[12] Bekaert G. and J. Liu (2004). Conditioning Information and Variance Bounds on Pricing Kernels. *Review of Financial Studies*, **17**, 2, 339-378.

[13] Bonomo M., R. Garcia, N. Meddahi, and R. Tedongap (2011). Generalized Disappointment Aversion, Long-run Volatility Risk, and Asset Prices. *Review of Financial Studies*, textbf24, 1, 82-122.

[14] Borwein J.M. and A.S. Lewis (1991). Duality Relationships for Entropy-Like Minimization Problems. *SIAM Journal of Control and Optimization*, **29**, 2, 325-338.

[15] Campbell J. and R. Shiller (1988). The Dividend-Price Ratio and Expectations of Future Dividends and Discount factors. *Review of Financial Studies*, **1**, 3, 195-228.

[16] Cerny A. (2003). Generalized Sharpe Ratios and Asset Pricing in Incomplete Markets. *European Finance Review*, **7** , 191-233.

[17] Chabi-Yo F. (2008). Conditioning Information and Variance Bounds on Pricing Kernels with Higher-Order Moments: Theory and Evidence. *Review of Financial Studies*, **21**, 1, 181-231.

[18] Corcoran S.A. (1998). Bartlett Adjustment of Empirical Discrepancy Statistics. *Biometrika*, **85**, 967-972.

[19] Cressie N., and T.R.C. Read (1984). Multinomial Goodness-of-fit Tests. *Journal of the Royal Statistics Society Series B*, **46**, 440-464.

[20] Dittmar R. (2002). Nonlinear Pricing Kernels, Kurtosis Preference, and the Cross-Section of Assets Returns. *Journal of Finance*, **57**, 1, 369-403.

[21] Dolmas J. (2013). Disastrous Disappointments: Asset-Pricing with Disaster Risk and Disappointment Aversion. Working Paper, Federal Reserve Bank of Dallas.

[22] Drechsler I. and A. Yaron (2011). What's Vol Got to Do With It. *Review of Financial Studies*, **24**, 1, 1-45.

[23] Gallant R., L. Hansen, and G. Tauchen (1990). Using Condit. Moments of Assets Payoffs to Infer the Volatility of Intertemporal Marginal Rates of Substitution. *Journal of Econometrics*, **45**, 141-179.

[24] Ghosh A., C. Julliard, and A. Taylor (2012). What is the Consumption-CAPM Missing? An Information-Theoretic Framework for the Analysis of Asset Pricing Models. Working Paper, Carnegie Mellon University.

[25] Glosten L.R. and R. Jagannathan (1994). A Contingent Claim Approach to Performance Evaluation. *Journal of Empirical Finance*, **1**, 133-160.

[26] Gospodinov N., R. Kan, and C. Robotti (2014). On the Hansen-Jagannathan Distance with a No-arbitrage Constraint. Working Paper, University of Toronto.

[27] Gul F. (1991). A Theory of Disappointment Aversion. *Econometrica*, **59**, 3, 667-686.

[28] Hansen L.P., J. Heaton, and E.G.J. Luttmer (1995). Econometric Evaluation of Asset Pricing Models. *Review of Financial Studies*, **8**, 2, 237-274.

[29] Hansen L.P., and R. Jagannathan (1991). Implications of Security Market Data for Models of Dynamic Economies. *Journal of Political Economy*, **99**, 2, 225-262.

[30] Hansen L.P., and R. Jagannathan (1997). Assessing Specification Errors in Stochastic Discount Factor Models. *Journal of Finance*, **52**, 2, 557-589.

[31] Hansen L.P., and S. Richards (1987). The Role of Conditioning Information in Deducing Testable Restrictions Implied by Dynamic Asset Pricing Models. *Econometrica*, **55**, 587-613.

[32] Kan R., and C. Robotti (2015). The Exact Distribution of the Hansen-Jagannathan Bound. *Management Science*, forthcoming.

[33] Kan R., and G. Zhou (2006). A New Variance Bound on the Stochastic Discount Factor. *Journal of Business*, **79**, 2, 941-961.

[34] Kimball M.S. (1993). Standard Risk Aversion. *Econometrica*, **61**, 3, 589-611.

[35] Kitamura Y. (2006). Empirical Likelihood Methods in Econometrics: Theory and Practice. Working Paper, Yale University.

[36] Kraus A. and R.H. Litzenberger (1976). Skewness Preference and teh Valuation of Risk Assets. *Journal of Finance*, **31**, 4, 1085-1100.

[37] Lewellen J., S. Nagel, and J. Shanken (2010). A Skeptical Appraisal of Asset-Pricing Models. *Journal of Financial Economics*, **96**, 175-194.

[38] Liu Y. (2012). Index Option Returns and Generalized Entropy Bounds. *Working Paper*, Duke University.

[39] Martin I. (2012). Consumption-Based Asset Pricing with Higher Cumulants. *Review of Economic Studies*, **80**, 745-773.

[40] Newey W. and R. Smith (2004). Higher Order Properties of GMM and Generalized Empirical Likelihood Estimators. *Econometrica*, **72**, 219-255.

[41] Rockafellar R.T. (1970). Convex Analysis. Princeton University Press.

[42] Routledge B. and S. Zin (2010). Generalized Disappointment Aversion and Asset Prices. *Journal of Finance*, **65**, 4, 1303-1332.

[43] Snow K. (1991). Diagnosing Asset Pricing Models Using the Distribution of Asset Returns. *Journal of Finance*, **46**, 955-983.

[44] Stutzer M. (1995). A Bayesian Approach to Diagnosis of Asset Pricing Models. *Journal of Econometrics*, **68**, 367-397.

Appendix

In order to prove Theorem 1, we make use of Theorem 2.4 page 326 in Borwein and Lewis (1991) that we present here, for completeness.

Theorem 2.4 Borwein Lewis, 1991

Let X be a locally convex vector space, $f : X \to (-\infty, \infty]$, convex, $A : X \to \Re^n$ continuous and linear, $b \in \Re^n$, $C \subset X$ convex, and $P \subset \Re^n$ a polyhedral cone. Consider the following dual problems:

$$\underset{x \in X}{inf}\; f(x), \qquad \text{subject to } x \in C, Ax \in b + P. \tag{54}$$

and the dual problem:

$$\underset{\tilde{\lambda} \in P^+}{sup}\; b'\tilde{\lambda} - g^*(A'\tilde{\lambda}) \tag{55}$$

with $g = f + \delta(.|C)$, g^* being the convex conjugate of function g, P^+ the dual cone of P, and $\delta(.|C)$ the indicator function of set C in the sense of Rockafellar (1970). If there exists a feasible point $\hat{x}$ in the quasi-relative interior of $((\text{dom} f) \cap C)$for the primal problem, then the values of the primal and the dual problem coincide (with attainment in the dual problem).

Proof of Theorem 1

In Theorem 2.4 of Borwein and Lewis, set $X =$ the space of admissible SDFs m with $E(\phi(m)) < \infty$, $f(m) = E(\phi(m))$, $C = X^+$,the space of nonnegative admissible SDFs m with $E(\phi(m)) < \infty$, $A(m) = E(m[(R - \frac{1}{a}1_K)'\ 1]')$, $b = [0_K\ a]'$ and $P = 0$. Theorem 2.5 at page 327 in Borwein and Lewis (1991) allows us to conjugate $\phi(.)$ within the expectation to obtain $g^* = E(\phi^{*,+})$. In addition, we obtain $A'\tilde{\lambda} = \tilde{\lambda}'[(R - \frac{1}{a}1_K)'\ 1]'$, and $P^+ = \Re^K$. No-arbitrage guarantees the existence of at least one feasible point (a strictly positive admissible SDF) in the quasi-relative interior of X^+. Therefore, rewriting $\tilde{\lambda} = [\alpha\ \lambda]$ the primal problem in Theorem 1 has a solution that coincides with that of the following dual problem:

$$\underset{\tilde{\lambda} \in \Re^K}{sup}\; a * \alpha - E\left[\phi^{*,+}(\lambda'(R - \frac{1}{a}1_K) + \alpha) + \delta(\tilde{\lambda}|\Lambda(R))\right] \tag{56}$$

where $\Lambda(R) = \{\tilde{\lambda} \in \Re^{K+1} : (\alpha + \lambda'(R - \frac{1}{a}1_K)) \in \text{dom } \phi^{*,+}\}$.

Finally, if either $d = lim_{x\to\infty} \frac{\phi(x)}{x} = \infty$ or ($d < \infty$ and $c = lim_{x\to\infty}((d - \phi'(x))x) > 0$), Theorem 5.5 at page 335 of Borwein and Lewis (1991) guarantees that the unique primal optimal stochastic discount factor solution is obtained by differentiating the convex conjugate $\phi^{*,+}(z)$ at the optimal (Lagrange Multipliers) dual solution: $m^* = \frac{\partial \phi^{*,+}(z)}{\partial z}_{z=\left(\alpha^* + \lambda^{*\prime}\left(R - \frac{1}{a}1_K\right)\right)}$.

Proof Corollary 1

We need to obtain the convex conjugate ϕ^* of ϕ to substitute in (8), when ϕ is in the Cressie Read family: $\phi^\gamma(m) = \frac{m^{\gamma+1} - a^{\gamma+1}}{\gamma(\gamma+1)}$. By looking at the equation that defines the convex conjugate (Equation 6), we define an auxiliary function $h^z_\gamma(w) = zw - \frac{w^{\gamma+1} - a^{\gamma+1}}{\gamma(\gamma+1)}$, whose domain, according to Theorem 1, is $dom(h^z_\gamma) = [0, \infty) \cap \text{dom}(\phi^\gamma)$. Note further that $\text{dom}(\phi^\gamma)$ is a function of γ: for $\gamma > -1$ and $\gamma \neq 0$, $\text{dom}(\phi^\gamma) = [0, \infty)$, and for $\gamma \leq -1$ or $\gamma = 0$, $\text{dom}(\phi^\gamma) = (0, \infty)$. In order to obtain the supremum in $\phi^{*+}_\gamma(z) = sup_{w\in \text{dom}(h^z_\gamma)} h^z_\gamma(w)$, we differentiate $h^z_\gamma(w)$ with respect to w leading to: $\frac{dh^z_\gamma(\tilde{w})}{dw} = z - \frac{\tilde{w}^\gamma}{\gamma}$. Now we split our analysis in three cases: $\gamma > 0$, $\gamma < 0$, and $\gamma = 0$.

i) $\gamma > 0$ case. In this case, $dom(h^z_\gamma) = [0, \infty)$. If $z < 0$, h^z_γ is a decreasing function of w and achieves its maximum at $\tilde{w} = 0$. If $z \geq 0$, $\tilde{w} = (\gamma z)^{\frac{1}{\gamma}}$ will be the unique critical point where the function achieves its maximum. By combining these two solutions, we note that $dom(\phi^{*,+}_\gamma) = \Re$, and, for an arbitrary z, we obtain $\tilde{w} = (\gamma z)^{\frac{1}{\gamma}} * 1_{\{\gamma z \geq 0\}}$. Substituting $\tilde{w}$ in $\phi^{*+}_\gamma(z) = h^z(\tilde{w})$, the convex conjugate becomes:

$$\phi^{*+}_\gamma(z) = \frac{(\gamma z)^{\frac{\gamma+1}{\gamma}}}{\gamma + 1} * 1_{\{\gamma z \geq 0\}} + \frac{a^{\gamma+1}}{\gamma(\gamma+1)} \tag{57}$$

and the optimization problem becomes:

$$\lambda^*_\gamma = \underset{\alpha\in\Re, \lambda\in\Lambda}{arg\ sup}\ a * \alpha - E\left[\frac{\left(\gamma\left(\alpha + \lambda'\left(R - \frac{1}{a}1_K\right)\right)\right)^{\frac{\gamma+1}{\gamma}}}{\gamma+1} 1_{\{\gamma\left(\alpha+\lambda'\left(R - \frac{1}{a}1_K\right)\right)\geq 0\}}\right] - \frac{a^{\gamma+1}}{\gamma(\gamma+1)} \tag{58}$$

By the fact that ϕ^* is homogeneous when we discard the constant $\frac{a^{\gamma+1}}{\gamma(\gamma+1)}$, it is possible to concentrate out the restriction on the SDF mean (α) from the optimization problem in Equation (58). To that end, we follow Kitamura (2006, page 12) and make use of the homogeneity of ϕ^* to eliminate the Lagrange multipliers λ when obtaining α^*. Letting $\Gamma(\alpha) = a * \alpha - \frac{(\gamma\alpha)^{\frac{\gamma+1}{\gamma}}}{\gamma+1} 1_{\{\gamma\alpha\geq 0\}}$, the optimal concentrated α is obtained by maximizing Γ. From its first order condition we get: $\frac{d\Gamma(\alpha)}{d\alpha} = 0 \Rightarrow \alpha^*_\gamma = \frac{a^\gamma}{\gamma}$. Substituting α^*_γ in Equation (58) gives the desired result.

ii) $\gamma < 0$ case. In this case, $dom(h_\gamma^z) = [0, \infty)$ if $-1 < \gamma < 0$ or $dom(h_\gamma^z) = (0, \infty)$ if $\gamma \leq -1$. If $z \geq 0$, h_γ^z is an increasing function of w and achieves its maximum at $\tilde{w} = \infty$. If $z < 0$, $\tilde{w} = (\gamma z)^{\frac{1}{\gamma}}$ will be the unique critical point where the function achieves its maximum. The fact that $\phi_\gamma^{*,+}$ is ∞ for $z \geq 0$ and is finite otherwise directly implies that $dom(\phi_\gamma^{*,+}) = (-\infty, 0)$. Combining these two solutions, the convex conjugate becomes:

$$\phi_\gamma^{*+}(z) = \frac{(\gamma z)^{\frac{\gamma+1}{\gamma}}}{\gamma+1} + \delta(z|\{\tilde{z} \in \Re : \gamma \tilde{z} > 0\}) + \frac{a^{\gamma+1}}{\gamma(\gamma+1)} \tag{59}$$

and the optimization problem becomes:

$$\lambda_\gamma^* = \underset{\alpha \in \Re, \lambda \in \Lambda}{arg\ sup}\ a * \alpha - E\left[\frac{\left(\gamma\left(\alpha + \lambda'\left(R - \frac{1}{a}1_K\right)\right)\right)^{\frac{\gamma+1}{\gamma}}}{\gamma+1} + \delta([\alpha\ \lambda])|D)\right] - \frac{a^{\gamma+1}}{\gamma(\gamma+1)} \tag{60}$$

where $D = \{\tilde{\alpha} \in \Re, \tilde{\lambda} \in \Re^k : \gamma\left(\tilde{\alpha} + \tilde{\lambda}'\left(R - \frac{1}{a}1_K\right)\right) > 0\}$. Following the steps in the previous case to concentrate α out we obtain the desired result.

iii) $\gamma = 0$ case. Since the limit $lim_{\gamma \to 0} E(\phi^\gamma(m)) = E(mln(m) - aln(a))$, which coincides with the Kullback-Leibler Information Criterion (KLIC) (See Stutzer, 1995, page 375). Therefore we go back to the problem of obtaining the convex conjugate $\phi^{*,+}$ with $\phi^0(m) = mln(m) - aln(a)$, whose domain is dom$(\phi^0) = (0, \infty)$. Note that the auxiliary function defined above is now $h_0^z(w) = zw - (wln(w) - aln(a))$, and its first derivative is $\frac{dh_0^z(\tilde{w})}{dw} = z - 1 - ln(\tilde{w})$. Since the domain of $h_0^z(.)$ is $(0, \infty)$ and in this range $ln(w)$ covers the whole real line, for any value of z the only critical point will be $\tilde{w} = e^{z-1}$, implying that $dom(\phi_0^{*,+}) = \Re$. Substituting $\tilde{w}$ in $h_0^z(\tilde{w})$, the convex conjugate becomes:

$$\phi_0^{*+}(z) = e^{z-1} + aln(a) \tag{61}$$

and the optimization problem becomes:

$$\lambda_0^* = \underset{\alpha \in \Re, \lambda \in \Lambda}{arg\ sup}\ a * \alpha - E\left[e^{(\alpha-1)+\lambda'\left(R - \frac{1}{a}1_K\right)}\right] - aln(a) \tag{62}$$

To concentrate α out of (62), we define $\Gamma(\alpha) = a * \alpha - e^{\alpha-1}$, and obtain its first order condition, to see that it is maximized at $\alpha_0^* = 1 + ln(a)$. Substituting α_0^* in (62) we obtain the desired result.[42]

[42] An alternative (simpler) way to obtain the dual optimization problem for the CR estimator with $\gamma = 0$ is to calculate the limit of the dual expression in Equation (11) $E\left[\left(\frac{a^{\gamma+1}}{\gamma+1} - \frac{1}{\gamma+1}\left(a^\gamma + \gamma\lambda'\left(R - \frac{1}{a}1_K\right)\right)^{\left(\frac{\gamma+1}{\gamma}\right)}\right)\right]$ when $\gamma \to 0$.

Proof Corollary 2

Case I: $\gamma > 0$. It is easy to see that $d = lim_{x\to\infty} \frac{\phi^{\gamma}(x)}{x} = lim_{x\to\infty} \frac{x^{\gamma}-\frac{1}{x}}{\gamma(\gamma+1)} = \infty$, implying that the first sufficient condition stated at Theorem 1, for the primal problem to be solved by the optimal implied SDF from the dual problem, is satisfied.

Case II: $-1 < \gamma < 0$. Since in this case, $d = lim_{x\to\infty} \frac{x^{\gamma}-\frac{1}{x}}{\gamma(\gamma+1)} = 0$, we proceed to verify if $c = lim_{x\to\infty} ((d - \phi'(x))x) > 0$. As $\phi'(x) = \frac{x^{\gamma}}{\gamma}$, we obtain $c = lim_{x\to\infty} - \frac{x^{\gamma+1}}{\gamma} = \infty > 0$, implying that the second sufficient condition at Theorem 1 is satisfied.

Case III: $\gamma = -1$, since in this case $\phi^{-1}(x) = -ln(x)$, we have $\phi^{-1\prime}(x) = -\frac{1}{x}$ and $c = lim_{x\to\infty} - (-\frac{1}{x}) * x = 1 > 0$, implying again that the second sufficient condition at Theorem 1 is satisfied.

Case IV: $\gamma < -1$. In this case, $d = 0$ and $c = lim_{x\to\infty} - \frac{x^{\gamma+1}}{\gamma} = 0$, implying that both sufficient conditions appearing in Theorem 1 are not satisfied. Therefore, we invoke Theorem 4.5 at page 334 of Borwein and Lewis (1991), which states that if the optimal Lagrange multipliers in the dual problem of Theorem 1 satisfy $sup_{\omega\in\Omega} \ \left(a^{\gamma} + \gamma\lambda_{\gamma}^{*\prime}\left(R(\omega) - \frac{1}{a}1_K\right)\right) > 0$, this is an alternative sufficient condition guaranteeing that the unique primal optimal solution can be obtained by differentiating the convex conjugate.[43]

Now, we have to prove that condition (**??**) is satisfied, for any sample space with a finite number of states. To that end, we argue that any solution λ_{γ}^{*} of (11) would have to satisfy the following similar condition: $\forall\omega \in \Omega : \left(a^{\gamma} + \gamma\lambda_{\gamma}^{*\prime}\left(R(\omega) - \frac{1}{a}1_K\right)\right) > 0$. This should be true, otherwise if there is at least one $\tilde{\omega} \in \Omega$ such that $\left(a^{\gamma} + \gamma\lambda_{\gamma}^{*\prime}\left(R(\tilde{\omega}) - \frac{1}{a}1_K\right)\right) \leq 0$, the delta function $\delta(.|\Lambda_{CR}(R))$ in (11) would assume $\delta(\lambda_{\gamma}^{*}|\Lambda_{CR}(R(\tilde{w}))) = \infty$. This would guarantee that the expectation in (11) would explode to $-\infty$. On its turn, this would imply a contradiction since λ_{γ}^{*} could not be the solution of a maximization problem with (11) assuming $-\infty$, while any feasible point would imply a finite expectation in (11). Finally, the condition $\forall\omega \in \Omega$: $\left(a^{\gamma} + \gamma\lambda_{\gamma}^{*\prime}\left(R(\omega) - \frac{1}{a}1_K\right)\right) > 0$ directly implies condition (**??**), if the sample space has a finite number of states.[44]

In the last part of the proof, we show that, whenever (7) is valid, differentiating the convex

[43]The original condition appearing at Theorem 4.5 of Borwein and Lewis would be replicated in our problem by $sup_{\omega\in\Omega} \ \alpha^{*} + \lambda_{\gamma}^{*\prime}\left(R(\omega) - \frac{1}{a}1_K\right) < 0$. We adapt it here to take into account the fact that we concentrate the Lagrange multiplier α^{*}, which comes from the restriction on the risk-free rate.

[44]Note, however, that for sample spaces with an infinite number of states there could still happen that $\forall\omega \in \Omega$: $\left(a^{\gamma} + \gamma\lambda_{\gamma}^{*\prime}\left(R(\omega) - \frac{1}{a}1_K\right)\right) > 0$, but $sup_{\omega\in\Omega} \ \left(a^{\gamma} + \gamma\lambda_{\gamma}^{*\prime}\left(R(\omega) - \frac{1}{a}1_K\right)\right) = 0$, violating condition (**??**).

conjugate that appears in (57), (59) or (61), with respect to z gives:

$$\frac{\partial\phi^{*,+}(z)}{\partial z} = (\gamma z)^{\frac{1}{\gamma}} * 1_{\{\gamma z \geq 0\}}, \gamma > 0 \tag{63}$$

$$\frac{\partial\phi^{*,+}(z)}{\partial z} = (\gamma z)^{\frac{1}{\gamma}}, \gamma < 0 \tag{64}$$

$$\frac{\partial\phi^{*,+}(z)}{\partial z} = e^{z-1}, \gamma = 0 \tag{65}$$

And substituting $z = \left(\alpha_\gamma^* + \lambda_\gamma^{*\prime}\left(R - \frac{1}{a}1_K\right)\right)$, with $\alpha_\gamma^* = \frac{a^\gamma}{\gamma}$, for $\gamma \neq 0$ or $\alpha_0^* = 1 + ln(a)$, for $\gamma = 0$, in (63), (64), and (65) we obtain expressions (13), (14), (15), which are valid for $\gamma \in [-1, \infty)$. Expression (**??**), which is valid for $\gamma < -1$, is also obtained from (64).

Table 1: Consumption Growth Persistence and Higher Moments of the SDF in the Long Run Risk Model.
This table presents the central moments $E(m - E(m))^n$, $n = 2, 3, 4$ of the Long Run Risk SDF derived by Bansal and Yaron (2004). The moments are computed for different values of the persistence parameter ρ of the consumption growth process. Except for ρ, all the other parameters in the model calibration for both consumption growth and volatility follow exactly Bansal and Yaron (2004). Parameters are chosen to match returns on a monthly frequency. The persistence parameter ρ varies between 0.983 and 0.995.

Persistence Consump. Growth (ρ)	**Variance**	**Skewness**	**Kurtosis**
0.983	0.0393	0.0047	0.0056
0.985	0.0498	0.0076	0.0095
0.988	0.0656	0.0132	0.0177
0.990	0.0900	0.0250	0.0369
0.993	0.1277	0.0510	0.0865
0.995	0.1850	0.1088	0.2225

Table 2: Diagnosing the Long Run Risk SDF with Different Cressie-Read Discrepancy Functions.
This table presents discrepancy values for the Long Run Risk SDF derived by Bansal and Yaron (2004), for different Cressie-Read discrepancies and values of the persistence of consumption growth. The parameter γ that indexes the Cressie-Read family assumes four values: -3,-1, 0, and 1. $\gamma = 1$ is the quadratic Hansen and Jagannathan case. $\gamma = 0$ covers Stutzer's KLIC entropic case. $\gamma = -1$ is the Bansal and Lehman (1997) / Backus, Chernov and Martin entropic case. The value $\gamma = -3$ is chosen to represent an example where higher moments (skewness and kurtosis) receive high weights in the discrepancy. Except for ρ, all the other parameters in the model calibration for both consumption growth and volatility follow exactly Bansal and Yaron (2004). Parameters are chosen to match returns on a monthly frequency. The persistence parameter ρ varies between 0.983 and 0.995.

Persistence ρ	$\gamma = -3$	$\gamma = -1$	$\gamma = -0.5$	$\gamma = 0$	$\gamma = 1$	**ratio** $\frac{E(\phi(\gamma=-3))}{E(\phi(\gamma=1))}$
0.983	0.0205	0.0193	0.0192	0.0193	0.0196	1.045
0.985	0.0263	0.0243	0.0243	0.0243	0.0249	1.055
0.988	0.0351	0.0318	0.0317	0.0318	0.0328	1.070
0.990	0.0492	0.0431	0.0429	0.0431	0.0450	1.094
0.993	0.0723	0.0601	0.0596	0.0601	0.0639	1.132
0.995	0.1150	0.0846	0.0838	0.0848	0.0925	1.190

Table 3: **Preference Shock Persistence and Higher Moments of the SDF in the Demand Side Model.** This table presents the central moments $E[m - E[m]]^n$ for $n = 2, 3, 4$ of the Demand Side Model SDF derived by Albuquerque, Eichenbaum and Rebelo (2012) under their general (extended) model. The moments are computed for different values of the preference shock persistence parameter ρ. Except for ρ, all the other parameters in the model calibration follow exactly Albuquerque, Eichenbaum and Rebelo (2012). Parameters are chosen to match returns on a monthly frequency. The persistence parameter ρ varies between 0.99 and 0.9995.

Persistence Parameter (ρ)	Variance	Skewness	Kurtosis
0.9900	0.0152	0.0007	0.0008
0.9919	0.0192	0.0011	0.0012
0.9938	0.0245	0.0018	0.0020
0.9957	0.0313	0.0030	0.0034
0.9976	0.0399	0.0048	0.0058
0.9995	0.0532	0.0086	0.0110

Table 4: **Diagnosing the Demand Side Model SDF with Different Cressie-Read Discrepancy Functions.** This table presents discrepancy values for the Demand Side Model SDF derived by Albuquerque, Eichenbaum and Rebelo (2012) under their general (extended) model, for different Cressie-Read discrepancies and values of preference shock persistence. The parameter γ that indexes the Cressie Read family assumes four values: -3,-1, 0, and 1. $\gamma = 1$ is the quadratic Hansen and Jagannathan case.$\gamma = 0$ covers Stutzer's KLIC entropic case. $\gamma = -1$ is the Bansal and Lehman (1997) / Backus, Chernov and Martin entropic case. The value $\gamma = -3$ is chosen to represent an example where higher moments (skewness and kurtosis) receive high weights in the discrepancy. Except for ρ, all the other parameters in the model calibration follow exactly Albuquerque, Eichenbaum and Rebelo (2012). Parameters are chosen to match returns on a monthly frequency. The persistence parameter ρ varies between 0.99 and 0.9995.

ρ	$\gamma = -3$	$\gamma = -1$	$\gamma = -0.5$	$\gamma = 0$	$\gamma = 1$	Ratio $\dfrac{E[\phi(\gamma = -3)]}{E[\phi(\gamma = 1)]}$
0.9900	0.0023	0.0023	0.0023	0.0023	0.0023	1.0157
0.9919	0.0029	0.0028	0.0028	0.0028	0.0028	1.0162
0.9938	0.0037	0.0037	0.0037	0.0037	0.0037	1.0171
0.9957	0.0053	0.0052	0.0052	0.0052	0.0052	1.0186
0.9976	0.0080	0.0078	0.0078	0.0078	0.0078	1.0210
0.9995	0.0125	0.0121	0.0120	0.0121	0.0122	1.0244

Table 5: **Disappointment Aversion Intensity and Higher Moments of the SDF in the GDA Model.** This table presents the central moments $E[m - E[m]]^n$ for $n = 2, 3, 4$ of the GDA Model SDF derived by Routledge and Zin (2010). The moments are computed for different values of θ, the disappointment aversion parameter. Except for θ, all the other parameters in the model calibration are set according to Bonomo et al. (2011) and reported in the online appendix. Parameters are chosen to match returns on a monthly frequency.

Disappointment Aversion Parameter (θ)	Variance	Skewness	Kurtosis
0.4000	0.0387	0.0001	0.0016
0.8800	0.1106	0.0012	0.0131
1.3600	0.1851	0.0036	0.0371
1.8400	0.2540	0.0069	0.0704
2.3200	0.3154	0.0107	0.1091
2.8000	0.3696	0.0149	0.1506

Table 6: **Diagnosing the GDA Model SDF with Different Cressie Read Discrepancy Functions.** This table presents discrepancy values for the GDA Model SDF derived by Routledge and Zin (2010) for different Cressie Read discrepancies and values of the disappointment aversion parameter. The parameter γ that indexes the Cressie-Read family assumes four values: -3,-1, -0.5, 0, and 1. $\gamma = 1$ is the quadratic Hansen and Jagannathan case.$\gamma = 0$ covers Stutzer's KLIC entropic case. $\gamma = -1$ is the Bansal and Lehman (1997) / Backus, Chernov and Martin entropic case. The value $\gamma = -3$ is chosen to represent an example where higher moments (skewness and kurtosis) receive high weights in the discrepancy. Except for θ, all the other parameters in the model calibration are set according to Bonomo et al. and reported in the online appendix.

Shock Parameter (θ)	$\gamma = -3$	$\gamma = -1$	$\gamma = -0.5$	$\gamma = 0$	$\gamma = 1$	Ratio $\frac{E[\phi(\gamma = -3)]}{E[\phi(\gamma = 1)]}$
0.4000	0.0217	0.0202	0.0199	0.0197	0.0194	1.1213
0.8800	0.0693	0.0592	0.0578	0.0565	0.0553	1.2526
1.3600	0.1322	0.1024	0.0985	0.0951	0.0926	1.4287
1.8400	0.2078	0.1452	0.1377	0.1314	0.1270	1.6362
2.3200	0.2948	0.1862	0.1742	0.1644	0.1577	1.8697
2.8000	0.3930	0.2250	0.2078	0.1939	0.1848	2.1267

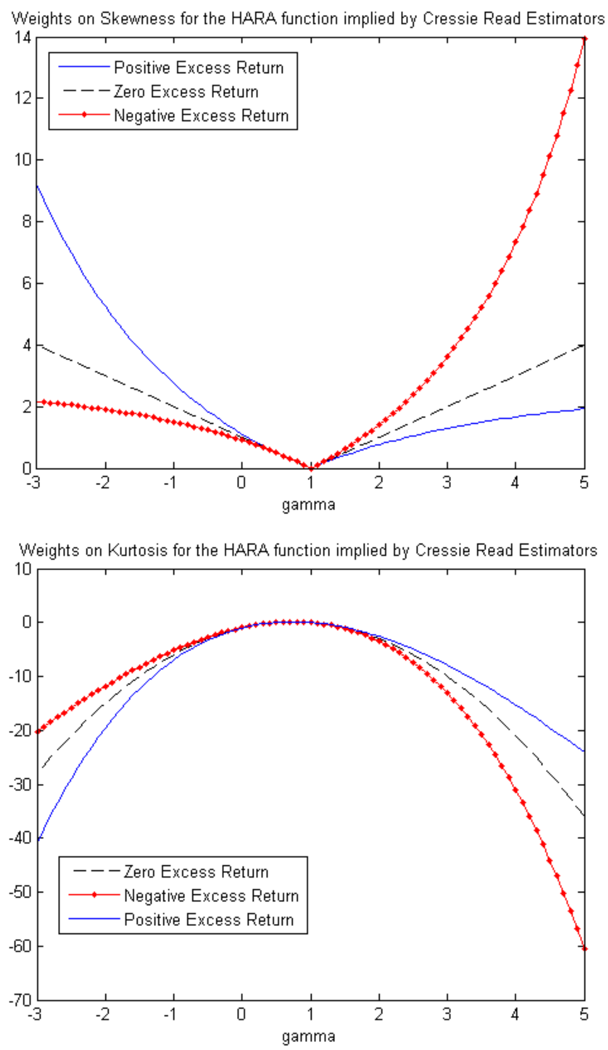

Figure 1: Skewness and Kurtosis Weights on Cressie Read Estimators

This picture presents the third and fourth derivatives of the HARA function $-\frac{1}{\gamma+1}(1+\gamma v)^{\frac{\gamma+1}{\gamma}}$ evaluated at an arbitrary value of v. Based on a Taylor expansion argument, by parameterizing $v = \lambda(R - \frac{1}{a})$ where R is a vector of returns, a is the SDF mean, and λ is a vector of weights, the third and fourth derivatives of the HARA function will indicate the weights given to skewness and kurtosis of the linear combination of returns.

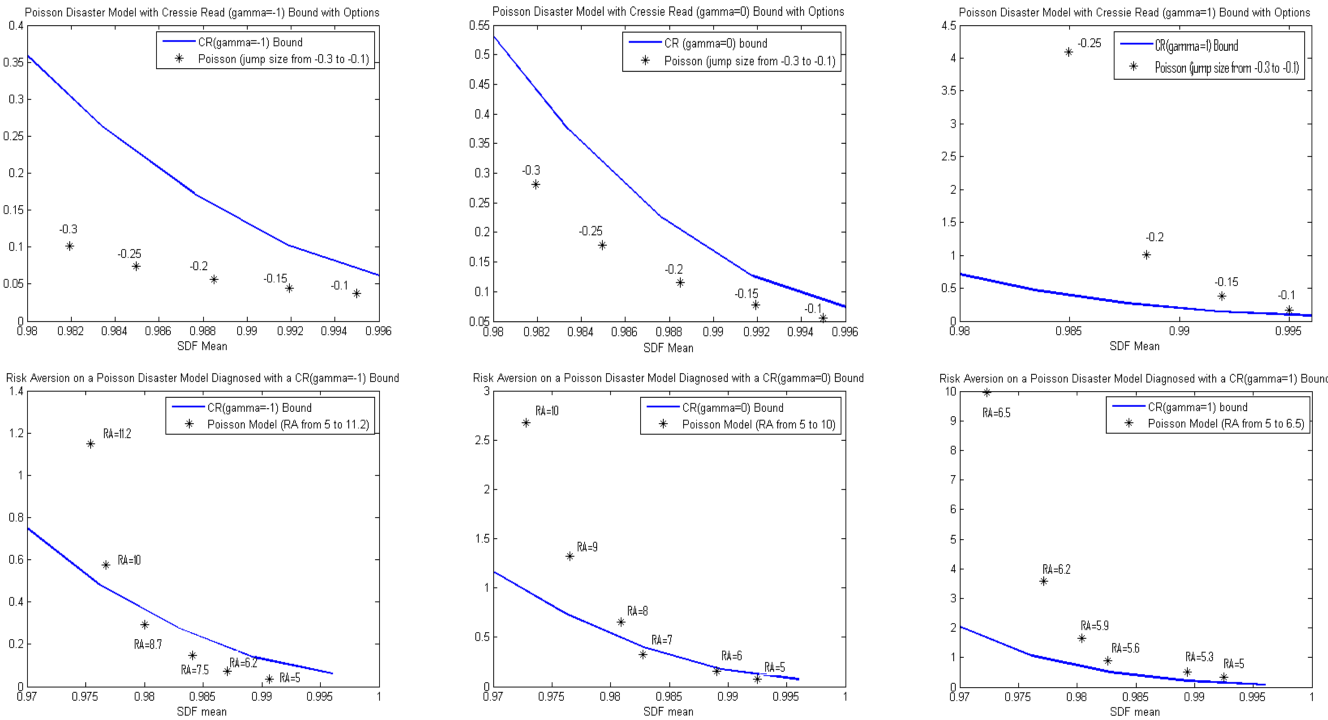

Figure 2: Diagnosing a Poisson Disaster Model Based on Entropic Bounds.
The top panel of graphs presents CR entropic bounds and corresponding discrepancy values for a Poisson disaster model. The Cressie Read bounds were built with monthly returns (January 1994 to September 2009) on a market factor and four option portfolios proposed by Agarwal and Naik (2004) (at-the-money call, out-of-the-money call, at-the-money put, and out-of-the-money put). The Poisson disaster model follows Backus, Chernov and Martin (2011)and is calibrated to match the mean (0.02)and standard deviation (0.0035) of the log consumption growth in the US. The frequency of disasters is equal to 0.01 and the volatility of jump size is equal to 0.15. Each star represents the discrepancy of the Poisson model for different jump sizes that range from -0.3 to -0.1. The bottom panel presents similar CR entropic bounds and corresponding discrepancy values once we vary the risk-aversion coefficient. The mean of jump size is now set to -0.3, while all the other parameters are kept the same. Each star represents the discrepancy of the Poisson model for different risk-aversion coefficients ranging from 5 to 10.

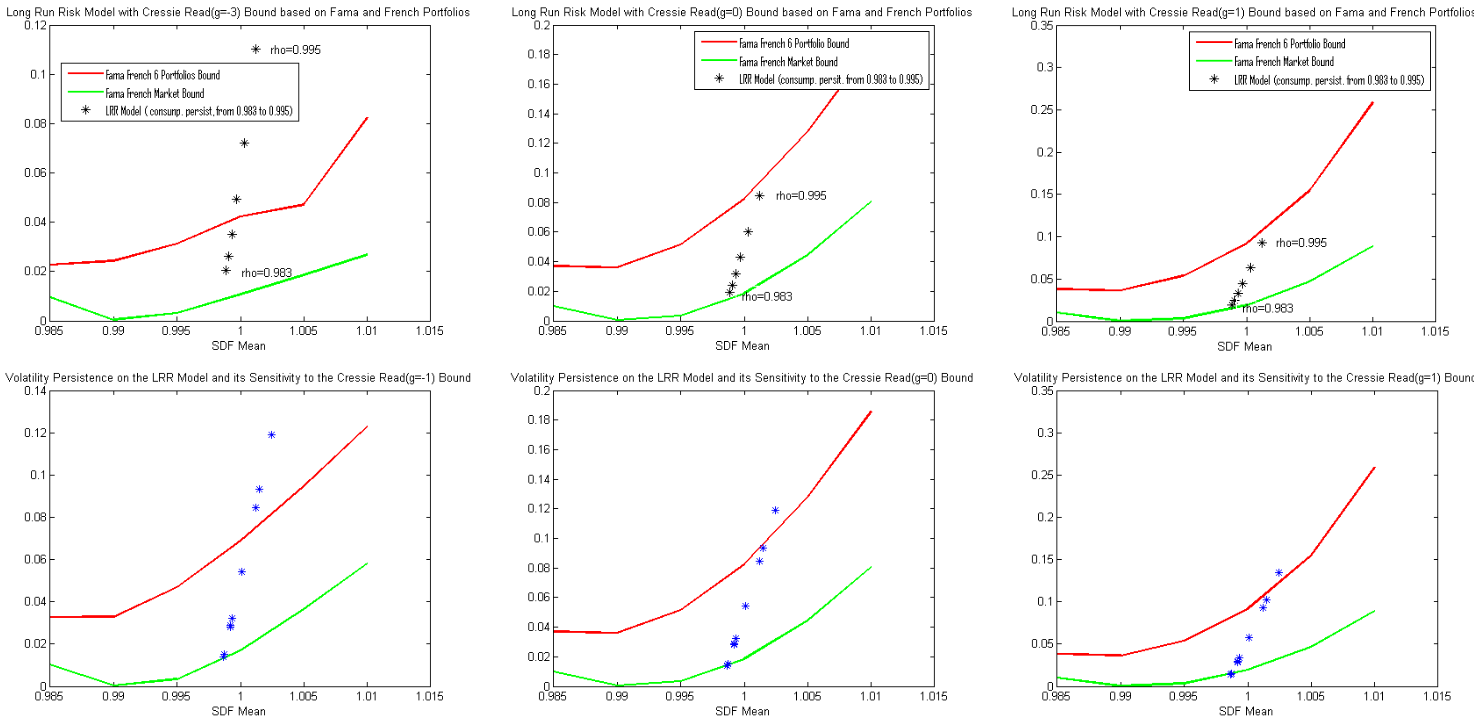

Figure 3: Sensitivity of Entropic Discrepancies to the Persistences of Volatility and Consumption Growth in the Long Run Risk Model.
The top panel of graphs picture presents CR entropic bounds and corresponding discrepancy values for the Long Run Risk model of Bansal and Yaron (2004) once we vary the persistence parameter of consumption growth ρ. The Cressie-Read bounds were built with monthly returns (August 1963 to November 2012) on Fama and French (FF) six size and book-to-market portfolios. The green line bound is based on the sole market return as a primitive security while the red line bound includes the six FF portfolios. Each star represents a different value for ρ from 0.983 to 0.995. In the bottom panel we vary both the volatility persistence parameter ν_1 in $\{0.987, 0.993, 0.999\}$ and the consumption growth persistence ρ in $\{0.979, 0.987, 0.995\}$. Each graph contains nine discrepancy values (stars) that represent all possible combinations of the persistence parameters ν_1 and ρ. Following Bansal and Yaron (2004), the coefficient of risk-aversion is set at 10, the deterministic discount factor at 0.9989, and $\psi = 1.5$. The other parameters are set to match the mean (0.02) and standard deviation (0.0035) of the log consumption growth in the US.

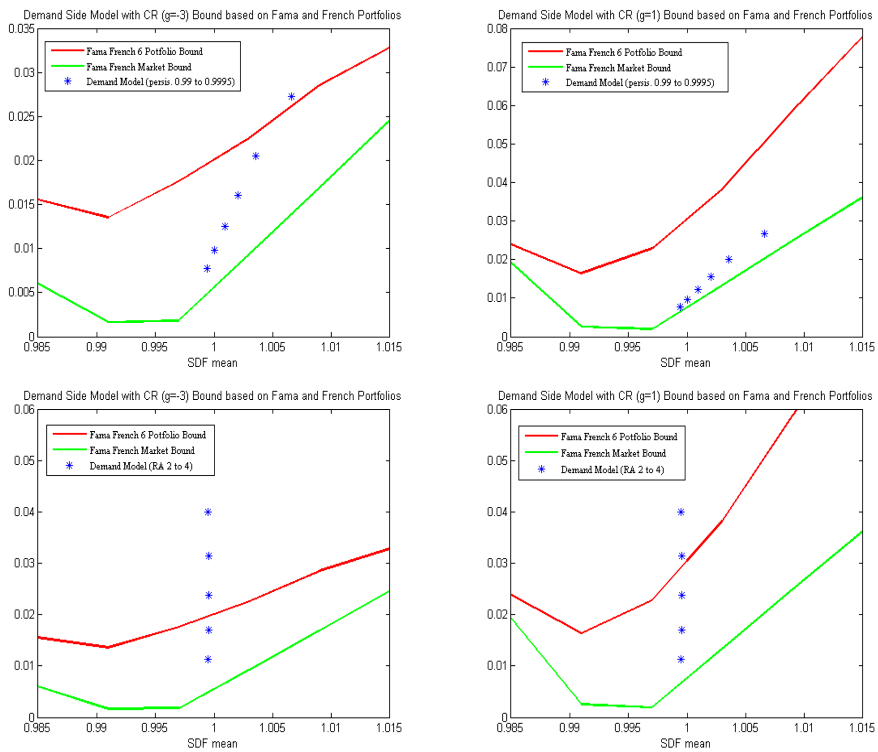

Figure 4: Sensitivity of Entropic Discrepancies to the Persistence of Demand Shock in the Demand Side Model.
This picture presents CR entropic bounds and corresponding discrepancy values for the Demand Side model of Albuquerque, Eichenbaum and Rebelo (2012). The Cressie-Read bounds were built with monthly returns (July 1926 to April 2014) of Fama and French six size and book-to-market portfolios. The coefficient of risk-aversion is set at 1.957, the deterministic discount factor at 0.9981, and the other parameters are set to match the equity premium. In the top panel we vary the preference shock persistence parameter ρ. In the bottom panel we vary the coefficient of risk-aversion from 2 to 4, while keeping the same values for all the other parameters.

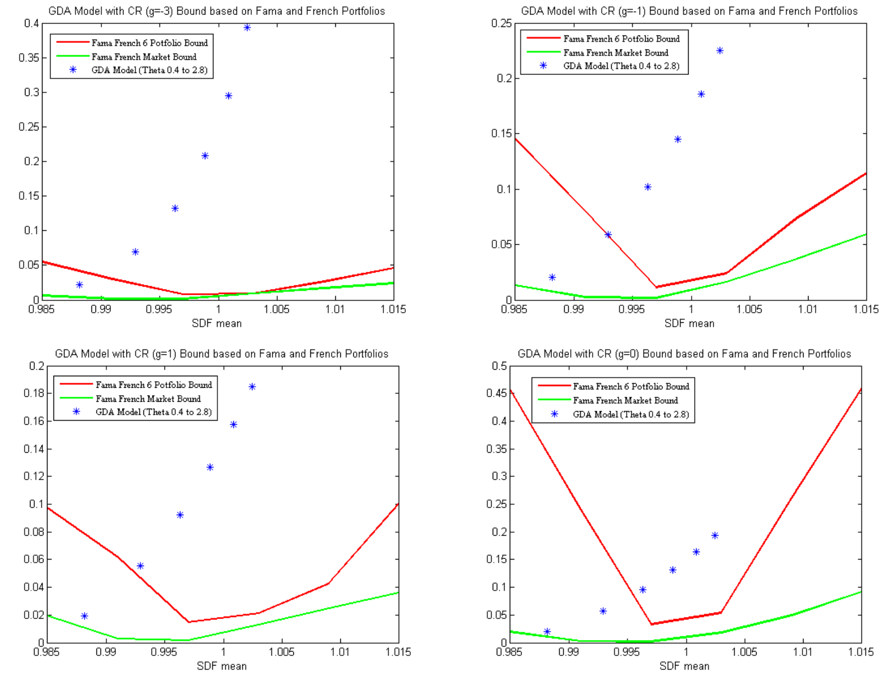

Figure 5: Sensitivity of Entropic Discrepancies to the Disappointment Aversion Parameter θ.
This picture presents CR entropic bounds and corresponding discrepancy values for the GDA model of Routledge and Zin (2010) when we vary the disappointment aversion parameter θ. The Cressie-Read bounds were built with monthly returns (July 1926 to April 2014) on Fama and French six size and book-to-market portfolios. The calibration values for the other parameters of the GDA model are set according to Bonomo et al. (2011) and reported in the online appendix.